INTEGRATING TIMESCALES FROM MOLECULES UP

EDITED BY: Rene A. Nome, Pilar Cossio and Y Z
PUBLISHED IN: Frontiers in Chemistry, Frontiers in Physics and Frontiers in Molecular Biosciences

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-860-1
DOI 10.3389/978-2-88966-860-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

INTEGRATING TIMESCALES FROM MOLECULES UP

Topic Editors:
Rene A. Nome, State University of Campinas, Brazil
Pilar Cossio, University of Antioquia, Colombia
Y Z, University of Illinois at Urbana-Champaign, United States

Citation: Nome, R. A., Cossio, P., Z, Y., eds. (2021). Integrating Timescales From Molecules Up. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-860-1

Table of Contents

04 ***Editorial: Integrating Timescales From Molecules Up***
Rene A. Nome, Pilar Cossio and Y Z

06 ***Energy Landscape for the Membrane Fusion Pathway in Influenza A Hemagglutinin From Discrete Path Sampling***
David F. Burke, Rosemary G. Mantell, Catherine E. Pitt and David J. Wales

17 ***Why Is the Range of Timescale So Wide in Glass-Forming Liquid?***
Takeshi Egami and Chae Woo Ryu

25 ***Evolving Stark Effect During Growth of Perovskite Nanocrystals Measured Using Transient Absorption***
James C. Sadighian, Kelly S. Wilson, Michael L. Crawford and Cathy Y. Wong

33 ***Quasi-Stationary States in Ionic Liquid-Liquid Crystal Mixtures at the Nematic-Isotropic Phase Transition***
Alokmay Datta, Kenichi Yoshikawa, Yukihiro Yoshida and Gunzi Saito

43 ***Dynamics of Photoinduced Energy Transfer in Fully and Partially Conjugated Polymers Bearing π-Extended Donor and Acceptor Monomers***
Youngseo Kim, Na Yeon Kwon, Su Hong Park, Min Ju Cho, Dong Hoon Choi and Sungnam Park

51 ***Slow Transition Path Times Reveal a Complex Folding Barrier in a Designed Protein***
Alexander Mehlich, Jie Fang, Benjamin Pelz, Hongbin Li and Johannes Stigler

62 ***Transcriptional Riboswitches Integrate Timescales for Bacterial Gene Expression Control***
Catherine E. Scull, Shiba S. Dandpat, Rosa A. Romero and Nils G. Walter

72 ***Persistent Homology Metrics Reveal Quantum Fluctuations and Reactive Atoms in Path Integral Dynamics***
Yunfeng Hu, Phonemany Ounkham, Ondrej Marsalek, Thomas E. Markland, Bala Krishmoorthy and Aurora E. Clark

frontiers in Chemistry

EDITORIAL
published: 13 April 2021
doi: 10.3389/fchem.2021.680533

Editorial: Integrating Timescales From Molecules Up

Rene A. Nome[1]*, *Pilar Cossio*[2,3] *and Y Z*[4]

[1] Department of Physical Chemistry, Institute of Chemistry, State University of Campinas - Unicamp, Campinas, Brazil, [2] Biophysics of Tropical Diseases, Max Planck Tandem Group, University of Antioquia UdeA, Medellín, Colombia, [3] Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Frankfurt am Main, Germany, [4] Department of Nuclear, Plasma, and Radiological Engineering, Department of Electrical and Computer Engineering, Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Keywords: rare events, long timescale processes, ultrafast, barrier crossing, single molecule

Editorial on the Research Topic

Integrating Timescales From Molecules Up

Barrier crossing in thermally activated processes is a fundamental and ubiquitous concept in Chemistry, Physics, and Biology. For example, in the study of reaction mechanisms, chemical kinetics investigations enable the determination of rate laws and the elucidation of elementary steps, while femtochemistry has sufficient time resolution to characterize the transition state and reaction intermediates. Despite the large timescale difference between typical reaction rate-constants and transition path lifetimes, both are crucial for understanding reaction mechanisms. Therefore, ideally one would like to integrate, conceptually and experimentally, the methods used at different timescales in order to obtain a more complete picture of how and why reactions occur. Examples where this general strategy can be applied abound, including biomolecular folding, catalysis, glass transition, nucleation and crystal growth, molecular and colloidal self-assembly, and among others.

Although timescale integration is a simple and well-known idea, and despite current efforts, it is still challenging to provide systematic structure-reactivity trends for general chemical and biological systems from this approach. This is not entirely surprising since usually the tools employed to study fast and slow thermally-activated processes are fundamentally different: stroboscopic and asynchronous methods, respectively.

This Frontiers in Chemistry, Frontiers in Physics, and Frontiers in Molecular Biosciences focus topic "*Integrating Timescales from Molecules Up*" tackles the challenges of studying barrier-crossing events from a wide time-scale perspective. We highlight contributions from multiple fields of research employing a growing class of interdisciplinary experimental, computational, and theoretical tools.

Drawing insights from scattering experiments and molecular dynamics simulations of metallic liquids, Egami and Ryu suggest that the noted over 15 orders of magnitude change in viscosity when liquids freeze into glasses can be caused by relatively small changes in the structural coherence in the medium range. Hu et al. use persistent homology to study quantum effects—delocalization, zero-point energy, and tunneling—in glassy dynamics and proton transfer reactions from a computational-topology perspective. The contribution from Sadighian et al. exemplifies how to integrate single-shot femtosecond transient absorption spectroscopy with *in situ* and real-time studies of perovskite nanocrystal nucleation and growth, which occurred on the time-scale from a few minutes to a couple hours. Kim et al. study push-pull chromophore polymers and show how the nature of the linker covalent bond (sigma vs. pi) influences photoinduced energy transfer dynamics on the nanosecond to microsecond timescales. Datta et al. study open system dynamics

OPEN ACCESS

Edited and reviewed by:
Malgorzata Biczysko, Shanghai University, China

***Correspondence:**
Rene A. Nome
nome@unicamp.br

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Chemistry

Received: *14 March 2021*
Accepted: *19 March 2021*
Published: *13 April 2021*

Citation:
Nome RA, Cossio P and Z Y (2021) Editorial: Integrating Timescales From Molecules Up. Front. Chem. 9:680533. doi: 10.3389/fchem.2021.680533

during the laser-induced creation of quasi-stationary states in mixtures of ionic liquids with liquid crystals at the nematic-isotropic phase transition.

On the biological side, Mehlich et al. use single-molecule force-spectroscopy experiments to characterize the slow transition-path time, on the order of milliseconds, of a *de-novo* designed protein. Reaching similar time scales, but from a computational perspective, Burke et al. use discrete path sampling to characterize the metastable and transition states in the membrane fusion pathway of Influenza A Hemagglutinin. Scull et al. highlight the functional importance of timescale integration from microseconds to minutes in riboswitches, encompassing ligand binding, RNA folding and transcription, and bacterial gene expression control.

Overall, this Research Topic highlights the utility of learning ideas from different fields, and suitably adapting them to the study of a system's dynamics. As stated above, recent advances have come from the integration of experimental tools such as ultrafast time-resolved non-linear laser spectroscopy, optical microscopy, single-molecule spectroscopy, neutron, and X-ray scattering, as well as theoretical and computational methods such as non-equilibrium statistical mechanics, stochastic dynamics, and rare event sampling. These advances may impact the study of biophysics, disordered materials, nanomaterials, catalysis, colloid science, and among others.

AUTHOR CONTRIBUTIONS

All authors contributed to manuscript writing and revision, read, and approved the submitted version.

FUNDING

YZ acknowledged the support by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Award Number DE-SC0014084. RAN acknowledged the support by CNPq INCT Catalysis (grant 444061/20185) and FAPESP (grants 2016/23430-9, 2017/11986-5, 2019/00290-5, 2019/23307-0, 2019/27471-0). PC was supported by the University of Antioquia, Minciencias, Ruta N (Colombia), and the Max Planck Society (Germany).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 25 September 2020
doi: 10.3389/fchem.2020.575195

Energy Landscape for the Membrane Fusion Pathway in Influenza A Hemagglutinin From Discrete Path Sampling

David F. Burke[1]*, *Rosemary G. Mantell*[2], *Catherine E. Pitt*[2] and *David J. Wales*[2]*

[1] EMBL-EBI, Wellcome Genome Campus, Hinxton, United Kingdom, [2] Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

The conformational change associated with membrane fusion for Influenza A Hemagglutinin is investigated with a model based upon pre- and post-fusion structures of the HA2 component. We employ computational methods based on the potential energy landscape framework to obtain an initial path connecting these two end points, which provides the starting point for refinement of a kinetic transition network. Here we employ discrete path sampling, which provides access to the experimental time and length scales via geometry optimization techniques to identify local minima and the transition states that connect them. We then analyse the distinct phases of the predicted pathway in terms of structure and energetics, and compare with available experimental data and previous simulations. Our results provide the foundations for future work, which will address the effect of mutations, changes in pH, and incorporation of additional components, especially the HA1 chain and the fusion peptide.

Keywords: energy landscape, rare event algorithms, influenza, membrane fusion, discrete path sampling

OPEN ACCESS

Edited by:
Rene A. Nome,
State University of Campinas, Brazil

Reviewed by:
Paul Whitford,
Northeastern University, United States
Vitor B. P. Leite,
São Paulo State University, Brazil

***Correspondence:**
David J. Wales
dw34@cam.ac.uk
David F. Burke
dburke@ebi.ac.uk

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Chemistry

Received: *22 June 2020*
Accepted: *19 August 2020*
Published: *25 September 2020*

Citation:
Burke DF, Mantell RG, Pitt CE and Wales DJ (2020) Energy Landscape for the Membrane Fusion Pathway in Influenza A Hemagglutinin From Discrete Path Sampling. Front. Chem. 8:575195. doi: 10.3389/fchem.2020.575195

1. INTRODUCTION

The influenza virus is a major cause of morbidity and mortality in humans. Viral infection is initiated by the binding of the trimeric hemagglutinin (HA) surface glycoprotein to glycans, which are terminated by the monosaccharide sialic acid, found on host cells in the upper respiratory tract. The virus is subsequently internalized via endocytosis. A prerequisite of fusion is the proteolytic cleavage of the HA protein into two chains, HA1 and HA2. The lower pH of the endosome triggers a structural rearrangement of HA, causing it to extend (Bullough et al., 1994). This process is thought to be driven through the protonation of amino acids, precipitating the loss of key interactions between the chains, allowing the disassociation of HA1 and HA2. The N-terminal region of HA2 then separates from the helical "stem," leading to the formation of a linear helical structure via a "spring-loaded" mechanism (Carr and Kim, 1993). The latter structural change is thought to be energetically favorable and does not depend on low pH. The largely hydrophobic 20 N-terminal amino acids, often called the fusion peptide, can then insert into the target membrane, initiating fusion.

The conformational changes of HA2 throughout this process are not well-characterized at an atomic level of detail, and our knowledge of the effect of mutations is sparse. Histidine is the only amino acid with a pKa value close to the pH of the endosome (pKa = 6.0), although aspartic acid (pKa = 3.9) and glutamic acid (pKa = 4.3) come close to the pH of the late endosome, and have

been implicated in the fusion process. Amongst others, Asp109 and Asp112 in HA2 have been shown to influence the pH at which fusion occurs (Trost et al., 2019).

In the present contribution we report the initial results from our calculations of the pathway for the conformational changes of a model system based on HA2 pre- and post-fusion structures. We have employed the computational tools of potential energy landscape theory, which exploits geometry optimization procedures to locate transition states and the minima they connect. For this complex conformational change, obtaining an initial connected path of minimum-transition state-minimum triples is itself a significant challenge. The refinements necessary to produce a physically acceptable initial path are described in section 3.

Our initial path contained around 4,500 steps, each one associated with a particular geometrical transition state (Murrell and Laidler, 1968), connecting a chain of local minima between the two selected pre- and post-fusion end points. We then employed the discrete path sampling approach to refine this pathway via creation of a kinetic transition network (Rao and Caflisch, 2004; Nóe and Fischer, 2008; Prada-Gracia et al., 2009; Wales, 2010), as described in section 3. New double-ended connection attempts between selected pairs of local minima were conducted to locate shorter and faster pathways between the two end points. As for creation of the initial pathway, these searches run in parallel, and new transition states and minima are added to the database as they are located.

The results described in section 4 correspond to a database containing 33,715 minima and 41,388 transition states. The phenomenological rate constant for interconversion of the pre- and post-fusion configurations then becomes a sum over all possible pathways through the network, (Wales, 2002, 2004, 2009; Swinburne and Wales, 2020) which can be computed deterministically using the graph transformation procedure (Trygubenko and Wales, 2006; Wales, 2009; Stevenson and Wales, 2014; MacKay and Robinson, 2018; Swinburne and Wales, 2020). The pathway that makes the largest contribution to this sum (the "fastest path" as defined in section 3) has around 3,200 steps. We judged convergence of the database by visualizing the landscape using disconnectivity graphs (Becker and Karplus, 1997; Wales et al., 1998) and from changes in the fastest path. Our description of the mechanism by which the pre-fusion conformation changes to post-fusion in section 4 is based on extracting and visualizing this path. The corresponding database will be made available online, and will be employed as a starting point for future work, where we will investigate the effects of protonation states, mutations, extensions to the model, and further database refinement. Our focus in this report is on our initial mechanistic insights and how they were obtained.

2. FUSION PEPTIDE MODEL

2.1. Definition of Pre- and Post-fusion Structures

In this study, the crystal structure of Influenza A(H3N2)/Aichi/1968(X-31) (PDB:2YPG), solved at neutral pH, was used as the pre-fusion structure (Lin et al., 2012). The extended post-fusion coiled-coil structure was taken from the stable recombinant low pH form of the HA2 subunit of the same strain of virus (PDB:1QUI) (Chen et al., 1999). It has previously been shown that the HA1 subunit, and the transmembrane domains of HA2, are not required for membrane fusion (Kim et al., 2011). Thus, only the portion of the HA2 common between the crystal structures (amino acids 33-172) was used. Here we consider only the monomeric structure of HA2. The protonation state was chosen to correspond to acidic pH, with histidines doubly protonated, but aspartate and glutamates neutral.

2.2. Force Field

We used the `OPTIM`[1] interface to `AMBER` (Cornell et al., 1995) to obtain all the energies and gradients employed in geometry optimization. The force field employed was AMBER ff99sb (Weiner et al., 1986; Pearlman et al., 1995; Hornak et al., 2006; Case et al., 2012) with a generalized Born implicit solvent model, (Onufriev et al., 2000, 2004) corresponding to the AMBER igb2 parameterizations. No cutoffs were used for any of the interactions, and an effective monovalent ionic concentration of 0.1 M was modeled using the Debye-Hückel approximation (Srinivasan et al., 1999). The force field was symmetrized to assure that feasible permutations of atoms of the same element give identical energies and gradients (Małolepsza et al., 2010, 2012). The GPU implementation of the AMBER potential was used (Götz et al., 2012) with the DPDP precision model, where the forces are computed in double precision. This approach is necessary for accurate convergence of transition states.

3. EXPLORING THE ENERGY LANDSCAPE AND FUSION PATHWAY

The initial HA2 pre- and post-fusion structures, described in section 2.1, are illustrated in **Figure 1**. Two numbering schemes are commonly used, namely starting from HA2 after cleavage, or from the start of the HA gene. In some subtypes of influenza viruses the HA1 region as well as the HA1/HA2 cleavage region vary in length, so we prefer to use HA2 numbering. For the virus considered in the present study, the alternative full gene numbering can be obtained by adding 329. We have colored each range of amino acid residues in the sequence that can be identified with local rearrangement events in particular phases of the pathway. These phases of the mechanism are described in section 4, and compared with existing experimental and simulation results as far as possible. New features that emerge from this analysis constitute predictions, which can be tested in future experiments. Specifically, we find that the B-loop achieves a helical conformation in two stages, while the highest barriers are associated with the rearrangement of region F and helices E and G to point in the opposite direction (see **Figure 1**). We return to these predictions in the Conclusions.

[1] *OPTIM: A Program for Geometry Optimisation and Pathway Calculations*. Available online at: http://wwwwales.ch.cam.ac.uk/software.html.

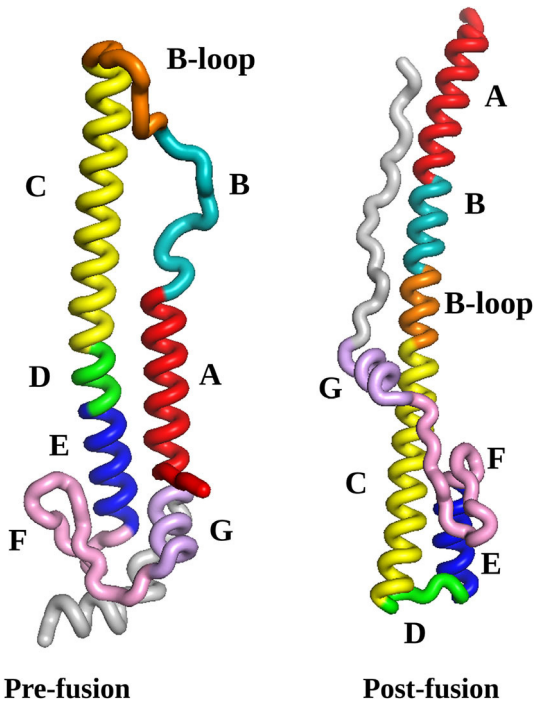

FIGURE 1 | The pre-fusion and post-fusion structures selected as the end points for the pathway search in the present work. The labels A to G refer to the regions identified from the crystal structure (Ni et al., 2014). Here, we distinguish between the N-terminal section of the B-loop, which readily forms an alpha helical conformation (positions 56-66), from the C-termimal section, which converts later in the pathway (positions 67-75). The coloring scheme reflects sections of the protein that can be identified with rearrangements in specific parts of the pathway, as follows: red, helix A (positions 38-55); cyan, helix B (positions 56-66); orange, B-loop (positions 67-75); yellow, helix C (positions 76-105); green, loop D (positions 106-111); blue, helix E (positions 112-125); pink, beta-hairpin F (residues 126-144); mauve, helix G (residues 145-154); gray, C-terminal fragment (residues 155-181); all structures were visualized using VMD (Humphrey et al., 1996). The numbering scheme starts from HA2 after cleavage.

3.1. Identifying Pathways Using Geometry Optimization

Having defined the end points for the pathway of interest our first task is to find a connected path between them, in terms of a sequence of minimum-transition state-minimum-transition state-· · ·-minimum stationary points. Once an initial path is located, we use it as the starting point for construction of a kinetic transition network (Rao and Caflisch, 2004; Nóe and Fischer, 2008; Prada-Gracia et al., 2009; Wales, 2010), with the emphasis on identifying the most kinetically relevant paths. Hence we explore the potential energy landscape, evolving a coarse-grained description in terms of the database of transition states and local minima. These stationary points are all located using geometry optimization techniques, as summarized below, which are well established, and have been reviewed elsewhere (Wales, 2003, 2010, 2018; Joseph et al., 2017). Observable properties are extracted from the database using standard tools of statistical mechanics and unimolecular rate theory is employed to calculate minimum-to-minimum rate constants (Forst, 1973; Laidler, 1987) using consistent approximations.

The standard geometry optimization techniques will only be summarized briefly here. Some additional considerations, necessitated by the complexity of the pathway under investigation, are highlighted below. Our geometrical definition of a transition state as a stationary point (vanishing gradient) with precisely one negative eigenvalue for the Hessian (second derivative) matrix, follows Murrell and Laidler (1968). A local minimum is a stationary point with no negative Hessian eigenvalues, where any infinitesimal displacement of internal coordinates raises the energy. All local minimizations employed the L-BFGS (Nocedal, 1980) (limited memory, Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) implementation of Wetzl and Taubmann[2] (Wetzl et al., 2013) with a modified line search (Asenjo et al., 2013). Double-ended transition state searches between selected pairs of local minima start from a doubly-nudged (Trygubenko and Wales, 2004) elastic band (Henkelman and Jónsson, 2000; Henkelman et al., 2000) (DNEB) calculation, where the images are only optimized sufficiently to distinguish local maxima in the profile. These local maxima are taken as transition state candidates and refined using hybrid eigenvector-following (Henkelman and Jónsson, 1999; Munro and Wales, 1999; Kumeda et al., 2001; Zeng et al., 2014) with custom CUDA kernels (Mantell et al., 2016) and calls to the cuBLAS library[3] to exploit GPU hardware. The connectivity defined by each transition state is established by characterizing the two downhill steepest-descent pathways initiated by small displacements parallel and antiparallel to the Hessian eigenvector corresponding to the unique negative Hessian eigenvalue. The initial displacements were determined using a golden section search to locate the lowest energy along the search direction within a maximum displacement range of 0.4 Å. The convergence condition for all stationary points was defined by a root-mean-square gradient below 10^{-6} kcal/mol/Å.

When seeking connections between local minima with very different structures, which lie far apart in configuration space, an initial straight line interpolation for the DNEB images is likely to produce unphysical structures. This problem is especially prevalent at the beginning of the whole procedure, before we even have a connected discrete path between the target minima. We therefore employed the quasi-continuous interpolation (QCI) scheme (Wales and Carr, 2012; Röder and Wales, 2018) whenever the optimal alignment of end points produced a distance in excess of 5 Å. Below this threshold the standard DNEB procedure was used. The updated version of QCI (Röder and Wales, 2018), where covalent bonds are defined directly from the AMBER topology, and various other improvements are exploited, was actually developed to tackle the present system. The benchmarks for smaller biomolecules, which enabled us to choose efficient parameters for HA2, are described in detail elsewhere (Röder and Wales, 2018). In brief, QCI works by defining an effective

[2] *CudaLBFGS*. Available online at: https://github.com/jwetzl/CudaLBFGS (accessed Oct 1, 2013).

[3] *cuBLAS*. Available online at: https://developer.nvidia.com/cublas (accessed Sep 21, 2016).

potential based on springs and repulsive charges, which enables the system to be interpolated one atom at a time. The quasi-continuous feature corresponds to additional penalty terms when a local minimum in distance is detected for atoms in neighboring images.

Proper alignment of the two minima selected for double-ended searches can have a significant effect on the characterization of pathways, in terms of efficiency, and in terms of locating the most kinetically favorable routes (Bauer et al., 2010; Wales and Carr, 2012). Here we must consider alignment with respect to overall translation and rotation, as well as feasible atomic permutations (Griffiths et al., 2017). Translational alignment is optimal when the centers of coordinates coincide, and orientational alignment can be achieved using quaternions (Kearsley, 1989; Coutsias et al., 2004) or Lagrange multipliers (Kabsch, 1978). For fixed position and orientation the optimal permutational alignment can be obtained using the shortest-augmenting path algorithm (Jonker and Volgenant, 1987). We have recently benchmarked a variety of other procedures, and reported an alternative branch and bound algorithm (Go-PERMDIST), which is competitive with the shortest-augmenting path procedure (Griffiths et al., 2017). In fact, there is one further subtlety, which we have found to be essential in locating physically relevant pathways efficiently for larger biomolecules. The global minimum distance between endpoints for fixed center of coordinates and orientation can produce structures with incorrect local atomic permutations (Wales and Carr, 2012). Such misalignments produce unphysical interpolations. To overcome this issue we instead use a local permutational alignment procedure, where each group of permutable atoms is treated separately, building up a neighborhood by progressively adding atoms within a cutoff distance, with the condition that the optimal distance between the atoms in the local cluster does not exceed a predefined tolerance. The local optimal distance is obtained by aligning the clusters with respect to translation and orientation, and employing the shortest augmenting path algorithm for local permutations (Wales and Carr, 2012). This approach was used for all the double-ended searches in the present work, for a maximum of 11 neighbor atoms within a cutoff distance of 5Å from the center of coordinates of the permuting set, and an alignment tolerance of 0.5Å for inclusion of atoms in the local cluster. Further details of the procedure can be found in the original report (Wales and Carr, 2012).

Two further changes to the DNEB implementation were also introduced. The images were redistributed every 350 DNEB steps to space them at equal distances along the path defined by the straight line segments of the current band. We also adjusted the DNEB spring constant that defines the harmonic potential between images, starting from a value of 100 kcal/mol/Å^2. Every five DNEB steps the value was increased or decreased by a factor of 1.03 if the mean deviation of the image spacing divided by the average was more or less than 6%. Maximum and minimum bounds on the force constant were set to 100 and 5 kcal/mol/Å^2, respectively. These values were found to perform well in systematic benchmarks for smaller systems (Röder and Wales, 2020).

3.2. Refining the Kinetic Transition Network

The local minima and transition states obtained in producing the initial connected pathway constitute the first entries in a kinetic transition network (Rao and Caflisch, 2004; Nóe and Fischer, 2008; Prada-Gracia et al., 2009; Wales, 2010) (KTN). Here, the local minima are nodes in a graph representation, and the transition states define the edges that connect them. Observable thermodynamic and kinetic properties are extracted from the network using standard tools of statistical mechanics and unimolecular rate theory (Forst, 1973; Laidler, 1987). In the present work we employed harmonic vibrational densities of states from normal mode analysis to compute the partition functions, and the corresponding harmonic transition state theory rate constants (Forst, 1973; Laidler, 1987) to estimate the rates for each minimum-to-minimum connection. The transition rates between pre- and post-fusion minima are calculated from a master equation (van Kampen, 1981; Kunz, 1995) representation of the overall kinetics using the graph transformation procedure (Trygubenko and Wales, 2006; Wales, 2009; Stevenson and Wales, 2014), which assumes that the dynamics are Markovian.

The kinetic transition network was refined by alternating strategies for selecting candidate pairs of end point minima in new double-ended searches. To locate shorter pathways, with lower barriers, we employed SHORTCUT procedures (Carr et al., 2005; Strodel et al., 2007; Wales et al., 2009) in the PATHSAMPLE program.[4] The current fastest discrete path between the target pre- and post-fusion minima, including the conditional occupation probability for the starting minimum, was first identified using Dijkstra's shortest path algorithm (Dijkstra, 1959) with edge weights (Carr et al., 2005; Carr and Wales, 2008a) $-\ln P_{\alpha\beta}$, where $P_{\alpha\beta}$ is the branching probability that the next step from minimum β is to adjacent minimum α. $P_{\alpha\beta}$ is calculated as $k_{\alpha\beta}/\sum_{\gamma} k_{\gamma\beta}$, where the sum is over all the pathways out of β and $k_{\alpha\beta}$ is the rate constant for transitions from β to γ. Pairs of minima for shortcutting were then selected in two ways: (1) from either side of the highest barrier within a maximum number of steps, or (2) the closest unconnected minima separated by between 10 and 75 steps. These shortcutting searches were combined with pair selection to remove artificial frustration, where low-lying minima are separated by a high barrier because a lower pathway has not yet been located (Strodel et al., 2007). This untrapping procedure selects pairs based on the downhill barrier height between them, divided by the potential energy difference. The corresponding ratio is similar to the Z-score parameter (Godzik et al., 1993) and the ratio of folding temperature to glass-transition temperature (Bryngelson and Wolynes, 1987), which are used in various energy landscape analysis procedures.

Further details of the discrete path sampling (DPS) approach described above are available in reviews (Joseph et al., 2017; Wales, 2018; Röder et al., 2019). Here we simply note that construction of a master equation framework using geometry optimization procedures is complementary to methods based on

[4] *PATHSAMPLE: A Program for Generating Connected Stationary Point Databases and Extracting Global Kinetics*. Available online at: http://www-wales.ch.cam.ac.uk/software.html

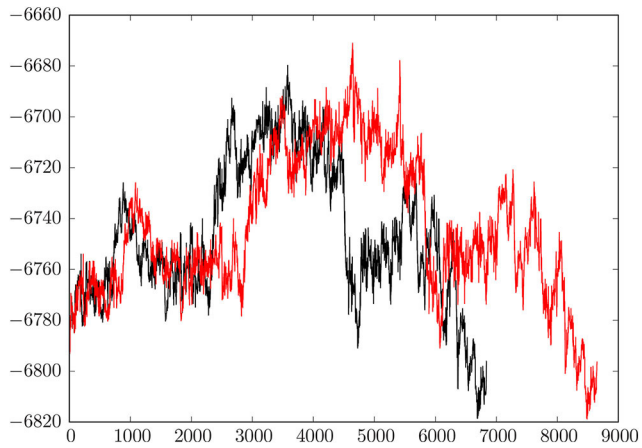

FIGURE 2 | Comparison of the initial connected path and the fastest path after refinement of the database. The relative energy in kcal/mol is plotted against the number of stationary points, which are organized in a connected sequence minimum-transition state-minimum-transition state-minimum, etc. from pre- to post-fusion structures. The initial path has 4,326 transition states and 8,653 stationary points; the refined path has 3,420 transition states and 6,841 stationary points.

explicit dynamics (Schütte C et al., 1999; Shirts and Pande, 2000; Singhal et al., 2004; Swope et al., 2004; Chodera et al., 2007; Pande et al., 2010; Prinz et al., 2011; Husic and Pande, 2018). The main advantage of DPS is that rate-determining steps corresponding to high barriers can be located efficiently using geometry optimization, which addresses the problems caused by trapping and broken ergodicity directly. The principal approximations are the sampling of stationary points and the assumption of Markovian dynamics; the use of harmonic densities of states is a further approximation, but more accurate descriptions of the partition functions and rate constants could be employed. Of course, the underlying force field also places a fundamental limit on the accuracy of our predictions.

4. RESULTS

Figure 2 shows a comparison of the initial discrete path and the fastest path extracted from the much larger database after refinement. Here we plot the energy of the minimum-transition state-minimum-transition state-· · ·-minimum connected sequence as a function of the number of stationary points. The total number of minima in a discrete path is always one more than the total number of transition states. The initial path contains 4,326 transition states (8,653 stationary points) and the fastest path after refinement has reduced to 3,420 transition states (6,841 stationary points). This simplification is typical of previous results for biomolecules, where the number of steps in the initial path usually decreases by 20% or more when discrete path sampling is used to locate more kinetically relevant pathways. We note that the initial pathway obtained for a reasonably complex system, such as this HA2 model, is unlikely to contribute to the true kinetics, and database refinement is essential. This refinement constitutes the majority of the computational effort.

The key result, summarized graphically in **Figure 2**, is that the overall mechanism does not change when the pathway is refined. Instead, the profile basically shortens in terms of the number of steps, which means that the database refinement has succeeded in removing unnecessary local rearrangements. Our analysis of the key steps in the transformation from the pre- to extended post-fusion coiled-coil conformation, presented in section 4, can therefore focus on essential changes in structure.

The disconnectivity graph (Becker and Karplus, 1997; Wales et al., 1998) in **Figure 4** shows how the pathway illustrated in **Figure 3** fits into the overall energy landscape. In this construction a vertical line begins at the potential (or free Krivov and Karplus, 2002; Evans and Wales, 2003) energy corresponding to each local minimum, with energy increasing on the vertical axis. The positions of branches on the horizontal axis are chosen to highlight the organization of the landscape as clearly as possible. At regular intervals of 4 kcal/mol we define an energy threshold, and partition the minima into disjoint sets, whose members can interconvert via pathways that lie below the threshold value. The branches corresponding to individual minima join at the threshold energy where they lie in the same superbasin. Hence we obtain a visualization of the landscape that provides a faithful account of the barriers that separate the minima, avoiding the problems of projections that may group together structures inappropriately (Bolhuis et al., 2002; Krivov and Karplus, 2004, 2006, 2008; Nóe and Fischer, 2008).

The fastest path shown in **Figure 3** reports on the progression of the rearrangement as a function of the number of steps corresponding to each individual transition state. Each one of these steps involves a local barrier between two minima. The disconnectivity graph provides a view of the global structure. Starting from the representative pre-fusion minimum, and proceeding from left to right in **Figure 3**, the next three minima illustrated on the path correspond to higher energy structures in one local subfunnel of the landscape. The first (I) corresponds to the rotation of the N-terminal helix A away from the main HA2 stem. The second step (II) consists of the initial bending of helix D at position 110. After the detachment and reorientation of helix A, the N-terminal end of the B-loop forms an alpha helical conformation, increasing the length of helix A by 2 turns (III). At this stage, the conformation of the C-terminal part of the B-loop is still identical to the initial prefusion state. Within this sub-funnel of structures, a conformation of the B-loop, similar to that seen experimentally, (Xu and Wilson, 2011) is observed in which the backbone and the sidechain of the Phe at position 63 has rotated, making it more exposed relative to the prefusion state. However, it frequently reverts back to the prefusion conformation along the pathway, in combination with other conformational transitions.

The next main structural change requires the system to surmount an energy barrier of about 50 kcal/mol moving into an adjacent local funnel. This change involves further extension of helix A along with unwinding of helix D into a loop separating the two helices C and E (IV). Mutations in this region have been associated with stabilizing the pre-fusion structure (Xu and

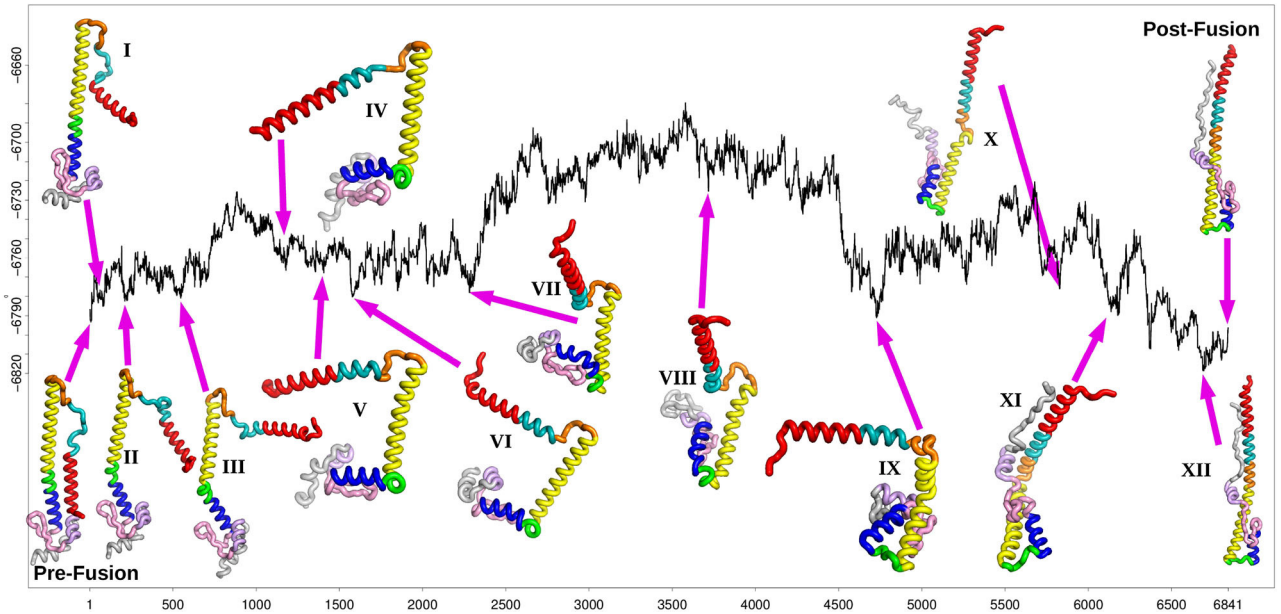

FIGURE 3 | Fastest path after refinement of the database, with selected local minima illustrated for the stationary points at position 81(I), 223(II), 541(III), 1183(IV), 1391(V), 1577(VI), 2283(VII), 3717(VIII), 4737(IX), 5829(X), 6137(XI), and 6693(XII) (the post-fusion end point minimum). The relative energy in kcal/mol is plotted against the number of stationary points, as in **Figure 2**.

Wilson, 2011) and may involve increasing the energy barrier between these local funnels. After further unwinding of loop D, helix E moves at right-angles to helix C (V), followed by helix G and C-terminal fragment (VI). After further structural rearrangements, helix G begins to detach from the beta hairpin F (129-140) and the helix E (VII). These structural rearrangements are similar to motions observed in previous molecular dynamics simulations (Lin et al., 2014).

The next minimum belongs to the highest energy part of the path, which corresponds to a relatively shallow local funnel structure. The structural change involves helices E and G moving to point in the opposite direction from the prefusion state (VIII). As the C-terminal peptide extends to interact with this elongated helix (X), the kink in helix B finally straightens resulting in the extended post-fusion structure (XII).

The structure of the disconnectivity graph is also of interest. At a coarse-grained level the landscape has two principal funnels, associated with structures related mainly either to the pre- or post-fusion structures. The highest barrier between these funnels is expected to constitute the rate-determining step. We have recently discussed how multi-funnel landscapes may be associated with multi-functional systems, for biomolecules and more generally (Chebaro et al., 2015; Joseph et al., 2017; Röder et al., 2019). In particular, a double-funnel landscape can define a molecular switch (Chakrabarti and Wales, 2011; Röder and Wales, 2017; Chakraborty and Wales, 2018). Such features have been investigated in detail for atomic clusters with competing morphologies, and are associated with features in the heat capacity and multiple relaxation time scales (Wales et al., 1998; Doye et al., 1999).

The funnel structure we see for HA2 suggests a model of fusion, consistent with that proposed from an analysis of FRET data (Das et al., 2018), where there is an initial reversible sampling of intermediate structures and atomic contacts are frequently lost. This phase corresponds to the left-hand funnel-structures I through VII in **Figure 4**. Once the high barrier is overcome, postfusion-like intermediates are sampled from the right-hand funnel, corresponding to structures IX through XII. The pre-fusion structure is no longer easily accessible and the structural changes are essentially irreversible. The more detailed sub-funnel structure in **Figure 4** reflects the different stages required to achieve this complex transformation.

Free energies can be estimated for all the stationary points using harmonic vibrational densities of states from normal mode analysis (Wales, 2003). We can then plot free energy disconnectivity graphs (Krivov and Karplus, 2002; Evans and Wales, 2003), where states may be defined using regrouping schemes that lump together free energy minima separated by barriers below a given threshold (Carr and Wales, 2008b). For the present system, we do not see any significant changes in the landscape when the free energy disconnectivity graph is considered. The key mechanistic features that we discuss above are conserved, and the free energy disconnectivity graph is therefore omitted for brevity.

The steps we have described above correspond to the fastest sequential path identified from the database. Overall rate constants can be computed from the infinite sum over discrete paths using the graph transformation approach (Trygubenko and Wales, 2006; Wales, 2009; Stevenson and Wales, 2014), and the k distinct paths algorithm (Sharpe and Wales, 2019)

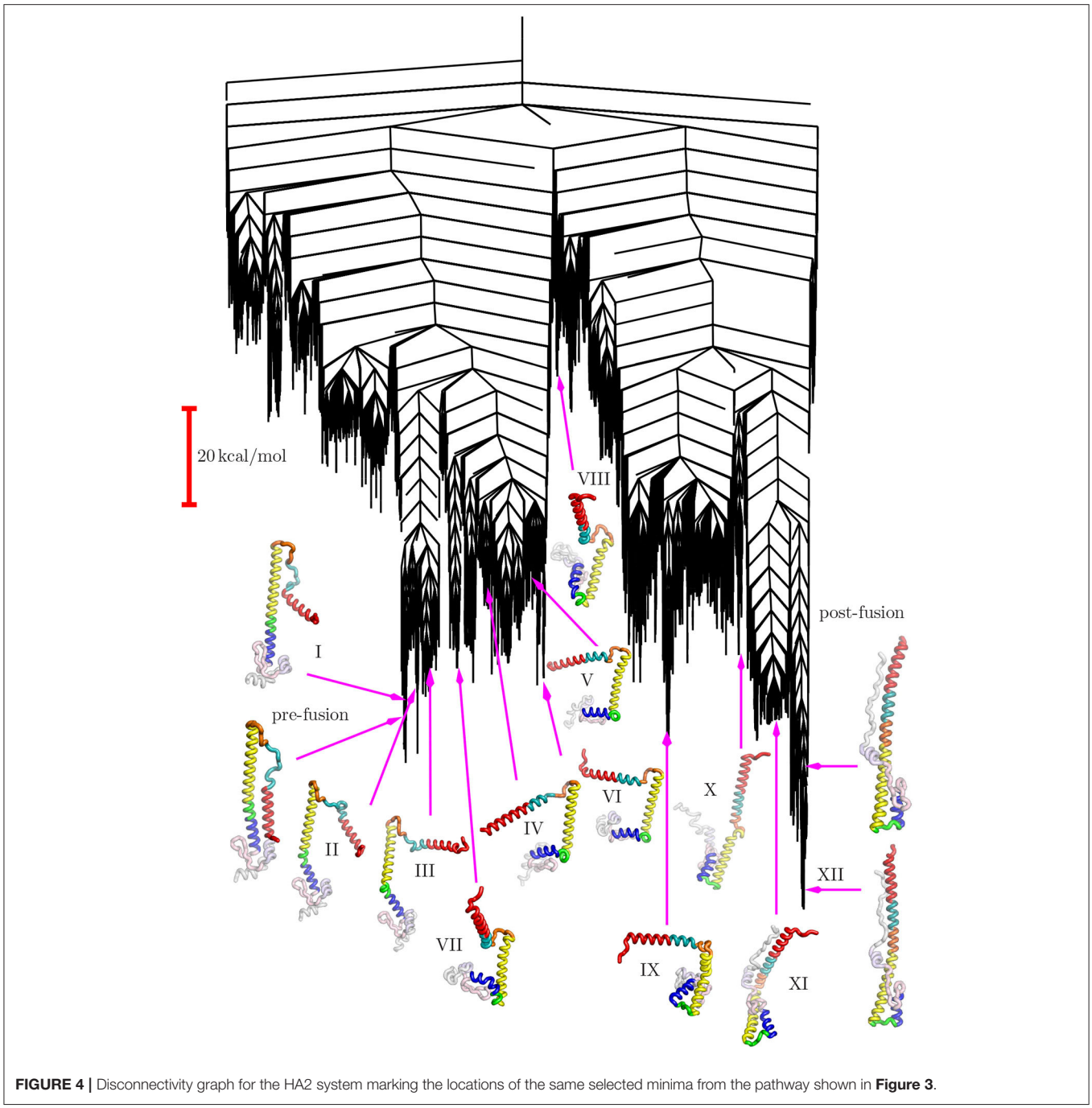

FIGURE 4 | Disconnectivity graph for the HA2 system marking the locations of the same selected minima from the pathway shown in **Figure 3**.

can be employed to distinguish pathways with different rate-determining steps. In this report we simply provide a qualitative account of the essential components we have identified in the fastest pathway. The most likely variations on this theme involve revisits among the low-lying minima that we associate with subfunnels in **Figure 4**. These structures are the most probable candidates for intervening minima that might be identified experimentally. Given the relatively high barrier that separates the regions of the landscape containing structures I through VII and IX to XII, the pathway ensemble is likely to contain contributions with revisits within the first set followed by revisits within the second set. Once the highest barrier has been overcome pre-fusion structures will be relatively inaccessible.

Pathways involving the same essential steps, with additional revisits and returns, constitute the same overall mechanism in this picture, rather than parallel paths with different routes between the low-lying minima. A transformation mediated by a short discrete path also supports a pathway ensemble, with revisits to minima along or adjacent to the fastest path, and a common rate-determining step. In the present case the pathway

ensemble would include revisits to the low-lying regions of the landscape associated with subfunnels. A more quantitative analysis of the kinetics will be conducted once this initial database has been expanded to include additional features, such as the fusion peptide.

The higher-energy regions of the landscape in **Figure 4**, which are not visited in any of the steps on the fastest pathway, would constitute off-pathway intermediates. However, the energies are comparable with the highest-lying stationary points on the fastest path, so we do not expect them to contribute significantly to the observable kinetics. Of course, there could be off-pathway structures and alternative mechanistic possibilities that have simply not been sampled in this initial study. Comparison with available experimental data provides some confidence that important intermediates have not been overlooked, but it is possible that further sampling will reveal additional features of interest.

5. CONCLUSIONS

We have located an initial pathway between pre- and post-fusion conformations of the HA2 component of influenza A hemagglutinin, and refined it to identify kinetically relevant paths using the discrete path sampling framework. This undertaking in itself presented a significant challenge. However, now that we have the HA2 kinetic transition network, it should be possible to build upon it by adding additional components, or introducing mutations and alternative protonation states. The present study therefore lays the foundations for future work, where we plan to extend the current HA2 system to include the fusion peptide and the HA1 chain. It may also be possible to develop a model of the trimeric structure. The structure of the landscape revealed in the present study seems quite well defined, and we expect it will be conserved by more accurate representations of the interatomic potential and solvent. Modifications caused by intermolecular interactions present in the trimer could be interesting, especially if cooperative effects are important. These questions should also be investigated in future work.

The identification of structural intermediates of HA within the fusion process has generally proved difficult. On extracting the fastest path from the kinetic transition network, we have identified well-defined intermediates with large energy barriers between them, even in the absence of HA1. The energy landscape we see for this HA2 model is consistent with a molecular switch. The initial reversible changes involve the detachment and reorientation of helix A, followed by the partial extension of helix B and a 90 degree rotation of helix E relative to the HA2 stem. The largest energy barrier involves the rearrangement of the beta-hairpin F and further rotation of helices E and G to point in the opposite direction from their pre-fusion structure. Once this barrier is overcome, the structural changes are largely irreversible. The C-terminal peptide extends to form a coiled-coil structure and the B helix straightens, resulting in the post-fusion structure.

From this predicted path, we can test mutations that may affect the energy barriers between the different phases of the transformation. Several mutations have been shown to affect the fusion pathway and their effects on the pathway will be considered in additional calculations. In particular, the mutation Thr59Met, situated in the B-helix, has previously been predicted to increase the unfolding temperature (Lin et al., 2018). The mutation R106H in helix D has also been show to stabilize the prefusion structure at normal pH (Xu and Wilson, 2011).

Most previous work has focused on mutations that affect the interactions between HA1 and HA2. We will extend the pathway to include HA1 in future work and analyse the effect on the pathway in the context of HA1/HA2 compared to the HA2-only system.

We will also recalculate the path by reconverging the stationary points for a range of pH, harvesting a new database. Currently, the path corresponds to a low pH system, with doubly protonated histidines, but not glutamate or aspartate. An analysis at neutral pH, as well as even lower pH (protonating glutamate or aspartate), will then be possible. Our hope is that this approach will be more efficient than trying to tackle larger systems from the outset, or performing separate investigations from scratch for alternative mutants and conditions that correspond to different pH.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, on request.

AUTHOR CONTRIBUTIONS

DB and RM set up the HA2 system and simulation parameters. RM wrote the CUDA interface routines. DW is the author and principal maintainer of the OPTIM and PATHSAMPLE programs, and ran the discrete path sampling calculations. DB and DW wrote the report. CP designed and constructed the computer cluster and provided all the technical assistance that was necessary to run the calculations. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

DW gratefully acknowledges financial support from the Engineering and Physical Sciences Research Council. DB wishes to acknowledge support received from the Center for Research on Influenza Pathogenesis (CRIP) funded by the NIAID and BARDA contracts HHSN272201400008C and HHSO100201500033C.

REFERENCES

Asenjo, D., Stevenson, J. D., Wales, D. J., Frenkel, D. (2013). Visualizing basins of attraction for different minimization algorithms. *J. Phys. Chem. B* 117, 12717–12723. doi: 10.1021/jp312457a

Bauer, M. S., Strodel, B., Fejer, S. N., Koslover, E. F., and Wales, D. J. (2010). Interpolation schemes for peptide rearrangements. *J. Chem. Phys*. 132:054101. doi: 10.1063/1.3273617

Becker, O. M., and Karplus, M. (1997). The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics. *J. Chem. Phys*. 106:1495. doi: 10.1063/1.473299

Bolhuis, P. G., Chandler, D., Dellago, C., and Geissler, P. L. (2002). Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem*. 53, 291–318. doi: 10.1146/annurev.physchem.53.082301.113146

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *J. Inst. Math. Appl.* 6, 76–90. doi: 10.1093/imamat/6.1.76

Bryngelson, J. D., and Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 84, 7524-7528. doi: 10.1073/pnas.84.21.7524

Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994). Structure of influenza haemagglutinin at the pH of membrane fusion. *Nature* 371, 37–43. doi: 10.1038/371037a0

Carr, C. M., and Kim, P. S. (1993). A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* 73, 823–832. doi: 10.1016/0092-8674(93)90260-W

Carr, J. M., Trygubenko, S. A., and Wales, D. J. (2005). Finding pathways between distant local minima. *J. Chem. Phys*. 122:234903. doi: 10.1063/1.1931587

Carr, J. M., and Wales, D. J. (2008a). *Latest Advances in Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the Biological Scale*, eds J.-P. Connerade and A. Solov'yov. (London: Imperial College Press), 460.

Carr, J. M., and Wales, D. J. (2008b). Folding Pathways and Rates for the Three-Stranded β-Sheet Peptide Beta3s using Discrete Path Sampling. *J. Phys. Chem. B* 112, 8760–8769. doi: 10.1021/jp801777p

Case, D. A., Darden, T., Cheatham, T. E., Simmerling, C., Wang, J., Duke, R., et al. (2012). *Amber 12 Reference Manual.* 348.

Chakrabarti, D., and Wales, D. J. (2011). Coupled linear and rotary motion in supramolecular helix handedness inversion. *Soft Matter* 7:2325. doi: 10.1039/c0sm01507e

Chakraborty, D., and Wales, D. J. (2018). Energy Landscape and Pathways for Transitions between Watson-Crick and Hoogsteen Base Pairing in DNA. *J. Phys. Chem. Lett*. 9, 229–241. doi: 10.1021/acs.jpclett.7b01933

Chebaro, Y., Ballard, A. J., Chakraborty, D., and Wales, D. J. (2015). Intrinsically disordered energy landscapes. *Sci. Rep.* 5:10386. doi: 10.1038/srep10386

Chen, J., Skehel, J. J., and Wiley, D. C. (1999). N- and C-terminal residues combine in the fusion-pH influenza hemagglutinin HA_2 subunit to form an N cap that terminates the triple-stranded coiled coil. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8967–8972.

Chodera, J. D., Dill, K. A., Singhal, N., Pande, V. S., Swope, W. C., and Pitera, J. W. (2007). Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 126:155101. doi: 10.1063/1.2714538

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197. doi: 10.1021/ja00124a002

Coutsias, E. A., Seok, C., and Dill, K. A. (2004). Using quaternions to calculate RMSD. *J. Comput. Chem*. 25, 1849–1857. doi: 10.1002/jcc.20110

Das, D. K., Govindan, R., Niki-Spiegel, I., Krammer, F., Lemke, E. A., and Munro, B J. (2018). Direct visualization of the conformational dynamics of single influenza hemagglutinin trimers. *Cell* 174, 926–937. doi: 10.1016/j.cell.2018.05.050

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numer. Math*. 1, 269–271. doi: 10.1007/BF01386390

Doye, J. P. K., Miller, M. A., and Wales, D. J. (1999). The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.* 110:6896. doi: 10.1063/1.478595

Evans, D. A., and Wales, D. J. Free energy landscapes of model peptides and proteins. *J. Chem. Phys*. (2003). 118:3891. doi: 10.1063/1.1540099

Fletcher, R. (1970). A new approach to variable metric algorithms. *J Comput*. 13, 317–322. doi: 10.1093/comjnl/13.3.317

Forst, W. (1973). *Theory of Unimolecular Reactions.* New York, NY: Academic Press.

Godzik, A., Kolinski, A., and Skolnick, J. (1993). De novo and inverse folding predictions of protein structure and dynamics. *J. Comput. Aided Mol. Des*. 7, 397–438. doi: 10.1007/BF02337559

Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Math. Comput.* 24, 23–26. doi: 10.1090/S0025-5718-1970-0258249-6

Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born. *J. Chem. Theory Comput*. 8, 1542–1555. doi: 10.1021/ct200909j

Griffiths, M., Niblett, S. P., and Wales, D. J. (2017). Optimal alignment of structures for finite and periodic systems. *J. Chem. Theory Comput.* 13, 4914–4931. doi: 10.1021/acs.jctc.7b00543

Henkelman, G., and Jónsson, H. (1999). A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *J. Chem. Phys*. 111:7010. doi: 10.1063/1.480097

Henkelman, G., and Jónsson, H. (2000). Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys*. 113:9978. doi: 10.1063/1.1323224

Henkelman, G., Uberuaga, B. P., and Jónsson, H. (2000). A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.* 113:9901. doi: 10.1063/1.1329672

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinform*. 65, 712–725. doi: 10.1002/prot.21123

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graphics* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Husic, B. E., and Pande, V. S. (2018). Markov state models: from an art to a science. *J. Am. Chem. Soc*. 140, 2386–2396. doi: 10.1021/jacs.7b12191

Jonker, R., and Volgenant, A. (1987). A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing* 38, 325–340. doi: 10.1007/BF02278710

Joseph, J. A., Röder K, Chakraborty, D., Mantell, R. G., and Wales J. D. (2017). Exploring biomolecular energy landscapes. *Chem. Commun*. 53, 6974–6988. doi: 10.1039/C7CC02413D

Kabsch, W. (1978). A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. Sect*. A 34, 827–828. doi: 10.1107/S0567739478001680

Kearsley, S. K. (1989). On the orthogonal transformation used for structural comparisons. *Acta Cryst*. A 45, 208–210. doi: 10.1107/S0108767388010128

Kim, C. S., Epand, R. F., Leikina, E., and Epand, R. M. (2011). The final conformation of the complete ectodomain of the HA2 subunit of influenza hemagglutinin can by itself drive low pH-dependent fusion. *J. Biol. Chem*. 286, 13226–13234. doi: 10.1074/jbc.M110.181297

Krivov, S. V., Karplus, M. (2004). Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14766–14770. doi: 10.1073/pnas.0406234101

Krivov, S. V., Karplus, M. (2006). One-dimensional free-energy profiles of complex systems: progress variables that preserve the barriers. *J. Phys. Chem. B* 110, 12689–12698. doi: 10.1021/jp060039b

Krivov, S. V., and Karplus, M. (2008). Diffusive reaction dynamics on invariant free energy profiles. *Proc. Natl. Acad. Sci. U.S.A.* 105, 13841–13846. doi: 10.1073/pnas.0800228105

Krivov, S. V., and Karplus, M. Free energy disconnectivity graphs: Application to peptide models. *J. Chem. Phys.* (2002). 117:10894. doi: 10.1063/1.1517606

Kumeda, Y., Munro, L., and Wales, D. J. (2001). Transition states and rearrangement mechanisms from hybrid eigenvector-following and density functional theory.: Application to $C_{10}H_{10}$ and defect migration in crystalline silicon. *J. Chem. Phys.* Lett. 341, 185–194. doi: 10.1016/S0009-2614(01)00334-7

Kunz, R. E. (1995). *Dynamics of First-Order Phase Transitions.* Thun: Deutsch.

Laidler, K. J. (1987). *Chemical Kinetics*. New York, NY: Harper & Row.

Lin, X., Eddy, N. R., Noel, J. K., Whitford, P. C., Wang, Q., Ma, J., et al. (2014). Order and disorder control the functional rearrangement of influenza hemagglutinin. *Proc. Natl Acad. Sci. U.S.A.* 111, 12049–12054. doi: 10.1073/pnas.1412849111

Lin, X., Noel, J. K., Wang, Q., Ma, J., and Onuchic, J. N. (2018). Atomistic simulations indicate the functional loop-to-coiled-coil transition in influenza hemagglutinin is not downhill. *Proc. Natl Acad. Sci. U.S.A.* 115, E7905–E7913. doi: 10.1073/pnas.1805442115

Lin, Y. P., Xiong, X., Wharton, S. A., Martin, S. R., Coombs, P. J., Vachieri, S. G., et al. (2012). Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proc. Natl. Acad. Sci. U.S.A.* 109:21474.

MacKay, R. S., and Robinson, J. D. (2018). Aggregation of Markov flows I: theory. *Philos. Trans. R. Soc. A* 376:20170232. doi: 10.1098/rsta.2017.0232

Małolepsza, E., Strodel, B., Khalili, M., Trygubenko, S., Fejer, S., and Carr, J. M., et al. (2012). Erratum: Symmetrization of the AMBER and CHARMM force fields. *J. Comp. Chem*. 33:2209. doi: 10.1002/jcc.23064

Małolepsza, E., Strodel, B., Khalili, M., Trygubenko, S., Fejer, S., and Wales, D. J. (2010). Symmetrization of the AMBER and CHARMM force fields. *J. Comp. Chem.* 31, 1402–1409. doi: 10.1002/jcc.21425

Mantell, R. G., Pitt, C. E., and Wales, D. J. (2016). GPU-Accelerated exploration of biomolecular energy landscapes. *J. Chem. Theory Comput*. 12, 6182–6191. doi: 10.1021/acs.jctc.6b00934

Munro, L., and Wales, D. J. (1999). Defect migration in crystalline silicon. *J. Phys. Rev*. B 59:3969. doi: 10.1103/PhysRevB.59.3969

Murrell, J. N., and Laidler, K. J. (1968). Symmetries of activated complexes. *Trans. Faraday. Soc*. 64, 371–377. doi: 10.1039/tf9686400371

Ni, F., Chen, X., Shen, J., and Wang, Q. (2014). Structural insights into the membrane fusion mechanism mediated by influenza virus hemagglutinin. *Biochemistry* 53, 846–854. doi: 10.1021/bi401525h

Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage *Math. Comput.* 35, 773–782. doi: 10.1090/S0025-5718-1980-0572855-7

Nóe, F., and Fischer, S. (2008). Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol*. 18, 154–162. doi: 10.1016/j.sbi.2008.01.008

Onufriev, A., Bashford, D., and Case, D. A. (2000). Modification of the generalized born model suitable for macromolecules. *J. Phys. Chem. B* 104, 3712–3720. doi: 10.1021/jp994072s

Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55, 383–394. doi: 10.1002/prot.20033

Pande, V. S., Beauchamp, K., and Bowman, G. R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52, 99–105. doi: 10.1016/j.ymeth.2010.06.002

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E. III, DeBolt, S., et al. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun*. 91, 1–41.

Prada-Gracia, D., Gómez-Gardenes, J., Echenique, P., and Falo, F. (2009). Exploring the free energy landscape: from dynamics to networks and back. *PLoS Comput. Biol*. 5:e1000415. doi: 10.1371/journal.pcbi.1000415

Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., et al. (2011). Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* 134:174105. doi: 10.1063/1.3565032

Rao, F., and Caflisch, A. (2004). The protein folding network. *J. Mol. Biol.* 342, 299–306. doi: 10.1016/j.jmb.2004.06.063

Röder, K, Joseph, J. A., Husic, B. E., and Wales, D. J. (2019). Energy landscapes for proteins: from single funnels to multifunctional systems. *Adv. Theory Simul.* 2:1800175. doi: 10.1002/adts.201800175

Röder, K., and Wales, D. J. (2017). Transforming the energy landscape of a coiled-coil peptide via point mutations. *J. Chem. Theory Comput.* 13, 1468–1477. doi: 10.1021/acs.jctc.7b00024

Röder, K., and Wales, D. J. (2018). Predicting pathways between distant configurations for biomolecules. *J. Chem. Theory Comput*. 14, 4271–4278. doi: 10.1021/acs.jctc.8b00370

Röder, K., and Wales, D. J. (2020). Improving double-ended transition state searches for soft-matter systems. *J. Chem. Phys.* 153:034104. doi: 10.1063/5.0011829

Schütte C, Fischer, A., Huisinga, W., and Deuflhard, P. (1999). A direct approach to conformational dynamics based on hybrid monte carlo. *J. Comput. Phys*. 151, 146–168. doi: 10.1006/jcph.1999.6231

Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* 24, 647–656. doi: 10.1090/S0025-5718-1970-0274029-X

Sharpe, D. J., and Wales, D. J. (2019). Identifying mechanistically distinct pathways in kinetic transition networks. *J. Chem. Phys*. 151:124101. doi: 10.1063/1.5111939

Shirts, M., and Pande, V. S. (2000). Screen savers of the world unite. *Science* 290, 1903–1904. doi: 10.1126/science.290.5498.1903

Singhal, N., Snow, C. D., and Pande, V. S. (2004). Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys*. 121:415. doi: 10.1063/1.1738647

Srinivasan, J., Trevathan, M. W., Beroza, P., and Case, D. A. (1999). Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* 101, 426–434. doi: 10.1007/s002140050460

Stevenson, J. D., and Wales, D. J. (2014). Communication: Analysing kinetic transition networks for rare events. *J. Chem. Phys.* 141:041104. doi: 10.1063/1.4891356

Strodel, B., Whittleston, C. S., and Wales, D. J. (2007). Thermodynamics and kinetics of aggregation for the GNNQQNY peptide. *J. Am. Chem. Soc*. 129, 16005–16014. doi: 10.1021/ja075346p

Swinburne, T. D., and Wales, D. J. (2020). Defining, calculating, and converging observables of a kinetic transition network. *J. Chem. Theory Comput*. 16, 2661–2679. doi: 10.1021/acs.jctc.9b01211

Swope, W. C., Pitera, J. W., and Suits, F. (2004). Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* 108, 6571–6581. doi: 10.1021/jp037421y

Trost, J. F., Wang, W., Liang, B., Galloway, S. E., Agbogu, E., Byrd-Leotis, L., et al. (2019). A conserved histidine in Group-1 influenza subtype hemagglutinin proteins is essential for membrane fusion activity. *Virology* 536, 78–90. doi: 10.1016/j.virol.2019.08.005

Trygubenko, S. A., and Wales, D. J. (2004). A doubly nudged elastic band method for finding transition states. *J. Chem. Phys*. 120:2082. doi: 10.1063/1.1636455

Trygubenko, S. A., and Wales, D. J. (2006). Graph transformation method for calculating waiting times in Markov chains. *J. Chem. Phys*. 124:234110. doi: 10.1063/1.2198806

van Kampen, N. G. (1981). *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland.

Wales, D. J. (2002). Discrete path sampling. *Mol. Phys*. 100, 3285–3305. doi: 10.1080/00268970210162691

Wales, D. J. (2003). *Energy Landscapes*. Cambridge: Cambridge University Press.

Wales, D. J. (2004). Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys*. 102, 891–908. doi: 10.1080/00268970410001703363

Wales, D. J. (2009). Calculating rate constants and committor probabilities for transition networks by graph transformation. *J. Chem. Phys*. 130:204111. doi: 10.1063/1.3133782

Wales, D. J. (2010). Energy landscapes: some new horizons. *Curr. Opin. Struct. Biol*. 20, 3–10. doi: 10.1016/j.sbi.2009.12.011

Wales, D. J. (2018). Exploring energy landscapes. *Ann. Rev. Phys. Chem.* 69, 401–425. doi: 10.1146/annurev-physchem-050317-021219

Wales, D. J., and Carr, J. M. (2012). Quasi-Continuous interpolation scheme for pathways between distant configurations. *J. Chem. Theory Comput.* 8, 5020–5034. doi: 10.1021/ct3004832

Wales, D. J., Carr, J. M., Khalili, de Souza, M. V. K., Strodel, B., and Whittleston, C. S. (2009). "Proteins: Energy, Heat and Signal Flow," in *Computation in Chemistry*, eds D. M. Leitner and J. E. Straub (Boca Raton, FL: CRC Press), 315.

Wales, D. J., Miller, M. A., and Walsh, T. R. (1998). Archetypal energy landscapes. *Nature* 394, 758–760. doi: 10.1038/29487

Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem*. 7, 230–252. doi: 10.1002/jcc.540070216

Wetzl, J., Taubmann, O., Haase, S., Köhler T, Kraus, M., and Hornegger, J. (2013). "Bildverarbeitung für die Medizin 2013: Algorithmen - Systeme – Anwendungen," in *Proceedings des Workshops vom 3. bis 5. März 2013 in Heidelberg,* eds H.-P. Meinzer, M. T. Deserno H. Handels, and T. Tolxdorff (Berlin; Heidelberg: Springer), 21–26.

Xu, R., and Wilson, I. A. (2011). Structural Characterization of an Early Fusion Intermediate of Influenza Virus Hemagglutinin. *J. Virol.* 85:5172. doi: 10.1128/JVI.02430-10

Zeng, Y., Xiao, P., and Henkelman, G. (2014). Unification of algorithms for minimum mode optimization. *J. Chem. Phys.* 140:044115. doi: 10.1063/1.4862410

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 29 September 2020
doi: 10.3389/fchem.2020.579169

Why Is the Range of Timescale So Wide in Glass-Forming Liquid?

Takeshi Egami[1,2,3]* *and Chae Woo Ryu*[1]

[1] Department of Materials Science and Engineering, Shull-Wollan Center – Joint Institute for Neutron Sciences, University of Tennessee, Knoxville, Knoxville, TN, United States, [2] Department of Physics and Astronomy, University of Tennessee, Knoxville, Knoxville, TN, United States, [3] Materials Sciences and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States

The viscosity and the relaxation time of a glass-forming liquid vary over 15 orders of magnitude before the liquid freezes into a glass. The rate of the change with temperature is characterized by liquid fragility. The mechanism of such a spectacular behavior and the origin of fragility have long been discussed, but it remains unresolved because of the difficulty of carrying out experiments and constructing theories that bridge over a wide timescale from atomic (ps) to bulk (minutes). Through the x-ray diffraction measurement and molecular dynamics simulation for metallic liquids we suggest that large changes in viscosity can be caused by relatively small changes in the structural coherence which characterizes the medium-range order. Here the structural coherence does not imply that of atomic-scale structure, but it relates to the coarse-grained density fluctuations represented by the peaks in the pair-distribution function (PDF) beyond the nearest neighbors. The coherence length is related to fragility and increases with decreasing temperature, and it diverges only at a negative temperature. This analysis is compared with several current theories which predict a phase transition near the glass transition temperature.

Keywords: liquid, liquid dynamics, relaxation time, medium-range correlation, fragility

OPEN ACCESS

Edited by:
Rene A. Nome,
State University of Campinas, Brazil

Reviewed by:
Masaru Aniya,
Kumamoto University, Japan
Carmelo Corsaro,
University of Messina, Italy

***Correspondence:**
Takeshi Egami
egami@utk.edu

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Chemistry

Received: *01 July 2020*
Accepted: *27 August 2020*
Published: *29 September 2020*

Citation:
Egami T and Ryu CW (2020) Why Is the Range of Timescale So Wide in Glass-Forming Liquid? Front. Chem. 8:579169. doi: 10.3389/fchem.2020.579169

INTRODUCTION

The viscosity of many liquids, such as water, is of the order of 10^{-2} poise ($= 10^{-3}$ Pa.s). Its timescale, defined by the Maxwell relaxation time, $\tau_M = \eta/G_\infty$, where η is viscosity and G_∞ is the high-frequency shear modulus, is of the order of pico-second (ps). Upon cooling liquid viscosity rises rather quickly, if crystallization can be avoided for instance by fast cooling. At low enough temperatures τ_M becomes so long that the supercooled liquid behaves like a solid. This kinetically frozen liquid is a glass. The transition to the glassy state is defined by the value of η reaching 10^{13} poise ($= 10^{12}$ Pa.s), when τ_M becomes of the order of 10^3 s. Thus, the timescale of liquid dynamics changes by as much as 15 orders of magnitude over a moderate temperature range. Such a rapid change has direct implications on glass-forming ability and other properties of glass-forming liquids, as well as on applications. The origin of this large change has long been debated without wide agreement (Debenedetti and Stillinger, 2001; March and Tosi, 2002; Dyre, 2006; Lubchenko and Wolynes, 2007; Götze, 2009; Donth, 2010; Berthier and Biroli, 2011; Edigar and Harrowell, 2012; Parisi et al., 2020), and remains one of the glass mysteries.

Our recent research results suggest that the medium-range order (MRO) in liquid plays a crucial role in dynamics of metallic and other liquids (Ryu et al., 2019, 2020; Egami, 2020; Ryu and Egami, 2020). In this article we discuss these results and their wider implications in

relation to other theories of liquids. In particular, we point out that our results do not suggest the divergence of viscosity just below the glass transition temperature, T_g, as many other theories do, and provide a resolution to the Kauzmann paradox concerning entropy extrapolating to negative values at low temperatures (Kauzmann, 1948). Our results also challenge the idea that defect-like objects control atomic transport and deformation in liquid and glass and raise questions on some prevailing theories.

VISCOSITY, FRAGILITY AND MRO

The temperature dependence of viscosity can be expressed in terms of the temperature dependent activation energy, $E_a(T)$, as

$$\eta\left(T\right)=\eta_{\infty}exp\left(\frac{E_a\left(T\right)}{k_BT}\right). \tag{1}$$

Above the crossover temperature, T_A, viscosity shows the Arrhenius behavior with a constant value of E_a, and below T_A it becomes strongly super-Arrhenius (Angell, 1995; Kivelson et al., 1995), resulting in rapid increase in viscosity culminating to the glass transition. It has been shown by simulations (Iwashita et al., 2013) and by experiments (Iwashita et al., 2017; Shinohara et al., 2019; Ashcraft et al., 2020) that above T_A viscosity is determined by a bond cutting dynamics, and $\tau_M = \tau_{LC}$, where τ_{LC} is the timescale for an atom to lose just one neighbor. Below T_A, however, the τ_M/τ_{LC} ratio increases rapidly with decreasing temperature, as liquid dynamics becomes more cooperative (Bellissard and Egami, 2018). This increase in cooperativity is the cause of the rapid increase in viscosity with decreasing temperature and eventual glass transition. The rate of increase in viscosity just above T_g is characterized by fragility,

$$m=\left.\frac{dlog\eta\left(T\right)}{d\left(T_g/T\right)}\right|_{T_g}. \tag{2}$$

A liquid with a large value of m is called fragile, whereas the one with a smaller value of m is called strong (Angell, 1995). The origin of the fragility is still in dispute (Angell, 1995; Novikov and Sokolov, 2004).

The structure of liquid and glass is usually described by the atomic pair-distribution function (PDF), $g(r)$, which describes the distribution of distances between atoms by

$$g\left(r\right)=\frac{1}{4\pi r^2N\rho_0}\sum\nolimits_{i,j}\left\langle\delta\left(r-\left|\mathbf{r_i}-\mathbf{r_j}\right|\right)\right\rangle, \tag{3}$$

where $\boldsymbol{r_i}$ is the position of the i-th atom, $i = 1, \ldots., N$, $\delta(r)$ is the δ-function, ρ_0 is the atomic number density, and $<\ldots.>$ denotes thermal average. It is related to the structure function,

$$S\left(Q\right)=\frac{1}{4\pi Q^2N}\sum\nolimits_{i,j}\exp\left(i\mathbf{Q}\cdot\left[\mathbf{r_i}-\mathbf{r_j}\right]\right), \tag{4}$$

which can be determined by x-ray or neutron diffraction, through the Fourier-transformation,

$$g\left(r\right)=1+\frac{1}{2\pi^2\rho_0r}\int_0^{\infty}\left[S\left(Q\right)-1\right]\sin\left(Qr\right)QdQ. \tag{5}$$

According to Ornstein and Zernike (1914) the medium-range PDF beyond the first peak decays with r as

$$G\left(r\right)=4\pi r\rho_0\left[g\left(r\right)-1\right]=G_0\left(r\right)\exp\left(-r/\xi_s\right), \tag{6}$$

where $G_0(r)$ is the $G(r)$ of the ideal glass, and ξ_s is the structural coherence length which characterizes the MRO. The ideal glass state defined by $G_0(r)$ has long-range density correlation without periodicity in the structure (Ryu et al., 2019). Because the medium-range PDF mostly accounts for the first peak of $S(Q)$ (Cargill, 1975; Ryu et al., 2020), the height of the first peak, $S(Q_1)-1$, where Q_1 is the position of the first peak, is proportional to ξ_s (Ryu et al., 2019).

In Ryu et al. (2019) $G(r)$, thus ξ_s, was measured for $Pd_{42.5}Ni_{7.5}Cu_{30}P_{20}$ liquid by high-energy x-ray diffraction using electrostatic levitation, from 420 to 1,100 K through the glass transition (573 K). Just above T_g $E_a(T)$ was found to be directly related to ξ_s by,

$$E_a\left(T\right)=E_0\left(\frac{\xi_s\left(T\right)}{a}\right)^3, \tag{7}$$

where a is the average neighbor distance (Ryu et al., 2019) and E_0 is a scaling parameter. Because

$$n_c\left(T\right)=\rho_0\left(\xi_s(T)\right)^3 \tag{8}$$

is the number of atoms in the coherence volume, $(\xi_s)^3$, it is indicative of the degree of atomic cooperativity of local dynamics in liquid. In other words, $E_a(T)$ is proportional to the number of atoms involved in the activation process for viscous flow;

$$E_a\left(T\right)=n_c\left(T\right)E_B,\qquad E_B=\frac{E_0}{\rho_0a^3}. \tag{9}$$

E_B represents the bond energy per atom, which is of the order of a fraction of eV and is significantly larger than k_BT_g, whereas n_c is relatively small even at T_g, typically below ten. The ratio of E_B/k_BT_g being larger than unity allows small changes in cooperativity $n_c(T)$ resulting in large changes in E_a, and the rapid increase in viscosity below T_A. Moreover, for various liquids examined by experiments as well as by simulations it was found that n_c at T_g is directly linked to fragility by,

$$n_c\left(T_g\right)=\frac{m}{m_0}, \tag{10}$$

where $m_0 = 8.7$ overall, 10.7 for metallic liquids, 7.4 for organic liquids, and 7.3 for network liquids (Ryu and Egami, 2020). Thus, fragility is related to the cooperativity of liquid dynamics and also to the "ideality" of the liquid structure. The liquid ideality is defined by the shape of the first peak of $S(Q)$ being

close to Lorentzian, as implied by the Ornstein-Zernike form, Equation (6), and by long ξ_s (Ryu et al., 2020). The Equation (9) appears similar to that of the classical Adam-Gibbs theory (Adam and Gibbs, 1965) in which the critical size of the cooperatively rearranging region, z^*, determines viscosity. However, the coherence volume defined here refers to the correlation in bulk liquid in equilibrium, whereas the cooperatively rearranging region is a transient defective object. This point will be discussed below.

TEMPERATURE DEPENDENCE OF MRO AND VISCOSITY

In Ryu et al. (2019) we studied the temperature dependence of the structure for various liquids by experiments and simulations, and showed that the height of the first peak of the structure function, $S(Q)$, and the coherence length follow the Curie-Weiss law,

$$\xi_s(T) = \frac{C}{T - T_{IG}}, \tag{11}$$

where T_{IG} is the ideal glass temperature which is negative. The origin of this behavior was briefly discussed in terms of the atomic-level stresses Egami (2011) in Ryu et al. (2019), and will be explained elsewhere (Egami and Ryu, 2020). Then at T_g,

$$\frac{d}{d\left(T_g/T\right)}\left(\frac{\xi_s(T)}{\xi_s\left(T_g\right)}\right)\Bigg|_{T_g} = \frac{1}{1 - T_{IG}/T_g} = m_c, \tag{12}$$

We found that,

$$m_c^2 = \frac{m}{m_1}. \tag{13}$$

with $m_1 = 613$, as shown in **Figure 1** for various metallic alloy liquids. Therefore,

$$\frac{T_{IG}}{T_g} = 1 - \left(\frac{m_1}{m}\right)^{1/2}. \tag{14}$$

Because

$$\rho_0 = \frac{f_p}{V_a} = \frac{6f_p}{\pi a^3}, \tag{15}$$

where V_a is the atomic volume and f_p is the atomic packing fraction,

$$\frac{a}{\xi_s(T)} = \frac{(m)^{\frac{1}{6}}}{(m_1)^{1/2}\left(\frac{\pi}{6m_0 f_p}\right)^{1/3}}\left(\frac{T}{T_g} - \frac{T_{IG}}{T_g}\right), \tag{16}$$

$$\frac{C}{aT_g} = \frac{(m_1)^{1/2}\left(\frac{\pi}{6m_0 f_p}\right)^{1/3}}{m^{1/6}}. \tag{17}$$

Because the value of f_p is similar for all metallic glasses (~0.7), the plots of $a/\xi_s(T)$ against T/T_g should be similar, except for vertical

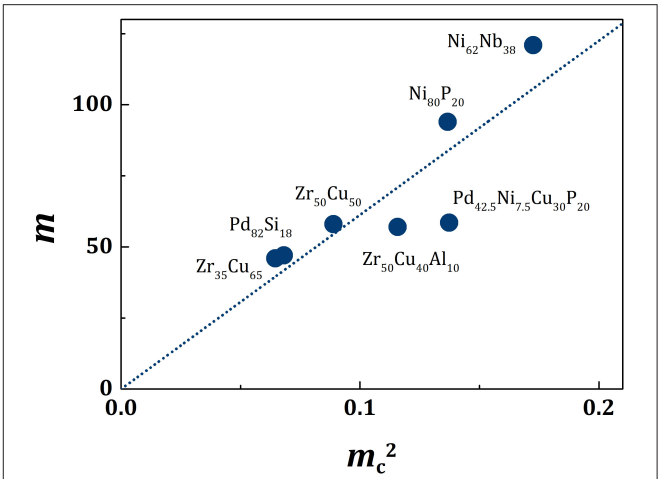

FIGURE 1 | The plot of $m_c{}^2$ vs. m for various metallic liquids. The dotted line is for $m_1 = 613$ in Equation (13).

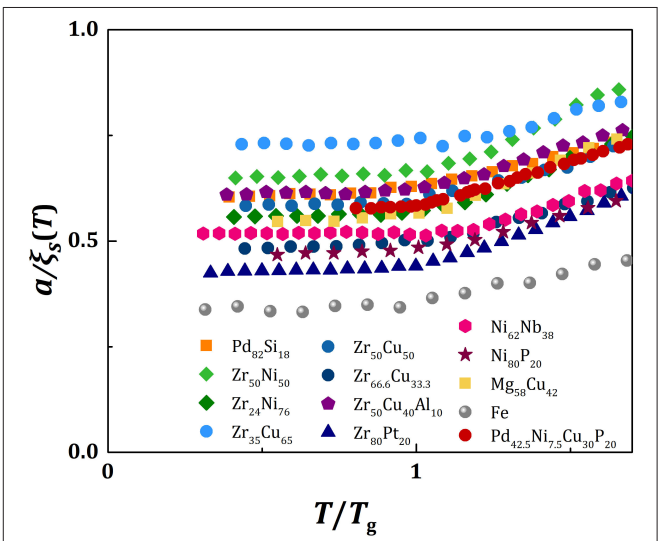

FIGURE 2 | The plots of $a/\xi_s(T)$ against T/T_g for various metallic liquids.

shifts, as shown in **Figure 2**. With vertical shifts they collapse to a near universal curve up, except for weak dependence above T_g on m (**Figure 3**). The value of C calculated by Equation (17), C_{calc}, is compared to the value of C obtained by fit with Equation (11), C_{fit}, for various liquids in **Figure 4**, showing good agreement. This near universality must be the reason for the success of the Kivelson scaling (Kivelson et al., 1995).

From Equations (1, 7, 17) we have

$$\eta(T) = \eta_\infty exp\left(\frac{E_0}{k_B T}\left(\frac{T_1}{T - T_{IG}}\right)^3\right), \tag{18}$$

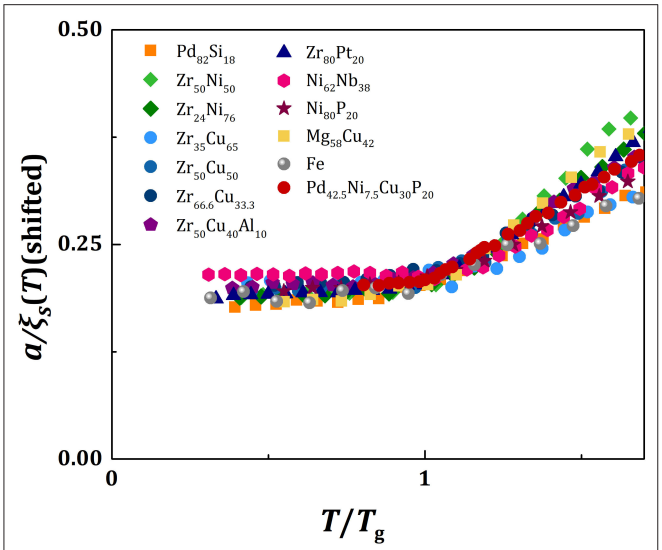

FIGURE 3 | The plots of $a/\xi_s(T)$ against T/T_g for various metallic liquids with vertical shifts to form a near universal curve.

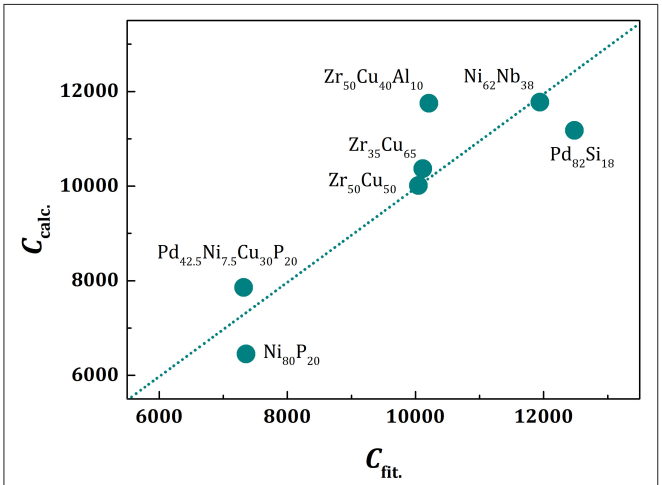

FIGURE 4 | The value of C given by Equation (17), C_{calc}, plotted against the value of C fit by Equation (11), C_{fit}, for various liquids. The dotted line is a linear fit, which is virtually identical to the line for $C_{calc} = C_{fit}$.

where

$$T_1 = \frac{(m_1)^{1/2}\left(\frac{\pi}{6m_0 f_p}\right)^{1/3}}{m^{1/6}} T_g. \tag{19}$$

Thus, the viscosity just above T_g can be described in terms of E_0 and T_{IG}. At temperatures above T_A the value of E_a becomes constant ($= E_\infty$) even though ξ_s keeps decreasing. The crossover is a purely dynamic phenomenon (Iwashita et al., 2013), and the MRO is irrelevant to dynamics above T_A. With a reasonable crossover, for instance,

$$E_a(T) = E_\infty \left(\frac{\xi_s(T)}{b}\right)^{d(T)}, \tag{20}$$

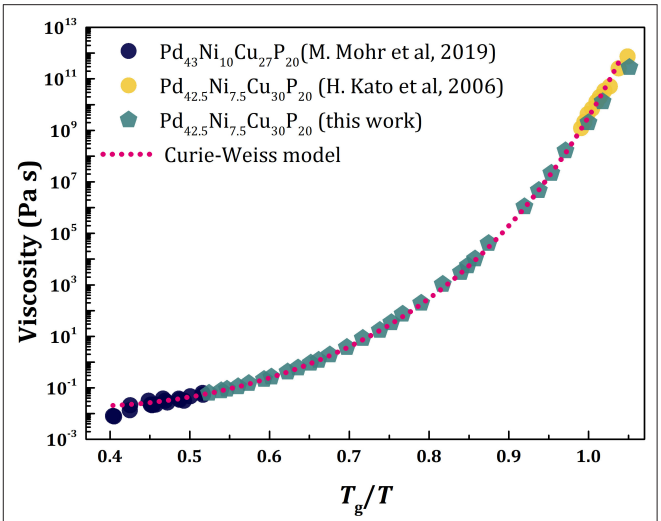

FIGURE 5 | The temperature dependence of viscosity of PdNiCuP liquid: Experimentally determined viscosity data of PdNiCuP liquid (Kato et al., 2006; Mohr et al., 2019) compared to those calculated. The pentagonal symbol denotes the viscosity calculated with the $\xi_s(T)$ determined from the PDF measured by x-ray diffraction (Ryu et al., 2019) using Equation (20), whereas the dashed line was calculated using the Curie-Weiss law, Equation (11).

where $d(T) = 3$ for $T < T_g$, $d(T) = 3\left(\frac{T_g}{T} - \frac{T_g}{T_A}\right)/\left(1 - \frac{T_g}{T_A}\right)$ for $T_g < T < T_A$, and $d = 0$ for $T > T_A$, a realistic temperature dependence of viscosity can be reproduced as shown in **Figure 5**. Here we compare the experimentally determined viscosity data of PdCuNiP liquid (Kato et al., 2006; Mohr et al., 2019) with those calculated with the Equation (20). The pentagonal symbol denotes the viscosity calculated with the $\xi_s(T)$ determined from the PDF measured by x-ray diffraction (Ryu et al., 2019), whereas the dashed line is calculated using the Curie-Weiss law, Equation (11). We assumed $T_A/T_g = 2.0$ (Blodgett et al., 2015), used the high-temperature data (Mohr et al., 2019) to determine the values of E_∞ ($= 0.77$ eV) and η_∞ ($= 1.82 \times 10^{-5}$ Pa.s), and the low-temperature data (Kato et al., 2006) to determine the value of b ($= 3.72$ Å).

COMPARISON WITH OTHER THEORIES AND MODELS

Absence of Divergence

The divergence of viscosity was first predicted by the Vogel-Fulcher-Tamman (VFT) model (Vogel, 1921; Fulcher, 1925; Tammann and Hesse, 1926),

$$\eta(T) = \eta_0 exp\left(\frac{B}{T - T_0}\right). \tag{21}$$

Models based upon structural coherence, such as the icosahedral correlation models (Steinhardt et al., 1981; Tomida and Egami, 1995; Tanaka et al., 2010), predict the divergence of structural coherence, thus the divergence of viscosity, below T_g in the vicinity of the Kauzmann temperature, T_K (Kauzmann, 1948).

The mode-coupling theory (Götze, 2009) predicts the divergence at a temperature, T_c, which is even higher than T_g, and only defect hopping provides mobility below T_c (Biroli et al., 2006). For a long time, it has been difficult to measure the viscosity of simple liquids above T_g because of crystallization. In the absence of serious questioning many of the current theories still assume the divergence of viscosity in the vicinity of T_K.

However, a more recent measurement of viscosity using liquid levitator (Blodgett et al., 2015) suggests that the VFT model actually shows poor fit to the data. It is likely that the prediction of the viscosity divergence is based upon poor extrapolation of viscosity to infinity. In fact, many other models do not predict divergence at $T > 0$ (Cohen and Grest, 1979; Nussinov, 2004; Demetriou et al., 2006; Elmatad et al., 2009; Mauro et al., 2009). According to Equation (7) viscosity diverges when the coherence length $\xi_s(T)$ diverges. For metallic glasses the value of $\xi_s\left(T_g\right)/a$ ranges from 1 to 2.7, with the average around 1.8. Therefore, the structure is quite far from ideal even at T_g. The temperature at which the ideal state is achieved in extrapolation, T_{IG}, is negative. Thus, viscosity never diverges at $T > 0$ and entropy does not become negative, resolving the Kauzmann paradox (Kauzmann, 1948).

Nature of Structural Order

Many theories attribute the origin of increased viscosity to development of some structural order which is frustrated and cannot achieve long-range ordering. The most prominent example of such order is the icosahedral order (Sadoc, 1981; Steinhardt et al., 1981). The idea is that because the icosahedral order is incompatible with periodicity it never grows into long-range order (Nelson, 1983; Sethna, 1983). However, such structural orders depend on chemical composition and local chemistry (Gaskell, 1979). Also, this is just a sophisticated version of the nano-crystalline theory which Frank (1952) tried to disprove by suggesting the possible presence of local icosahedral configuration. Note that liquid is stabilized by configurational entropy: The development of local order of a particular atomic configuration will reduce the entropy and destabilize liquid.

On the other hand, we postulate the ideal liquid/glass state by extrapolating the coherence length ξ_s to infinity (Ryu et al., 2019). This state has very diverse local structures, with widely varying local configurations. For instance, the population of the icosahedral local structure is merely 0.7%. The order parameter, ξ_s, does not describe the structural order, but the MRO of local density fluctuation. Higher-order peaks of the PDF at large distances include many interatomic distances within the peak. The width of the high-order PDF peaks is about 0.1 nm, and this defines the spatial resolution of the structure in the ideal state. Thus, the MRO describes coarse-grained density fluctuations, and not the atomic-level structural correlations, because the spatial resolution needs to be better at least by an order of magnitude to specify the atomic structure. In our view the local icosahedral ordering which occurs in single element liquid is not indicative of glass formation, but it is likely be that of nano-scale crystallization or quasicrystal formation. A single-element metallic liquid is a very poor glass-former and easily crystallizes. For a single-component liquid the second peak of $S(Q)$, which is more sensitive to crystallinity, diverges at a positive temperature below T_g (Ryu et al., 2019). This suggests that the divergence of the local order just below T_g implies nano-scale crystallization.

Idea of Defects

In crystalline solids atomic transport occurs only through the motion of lattice defects, such as vacancies and interstitial defects. Because the structure of liquid and glass is strongly disordered and appears to be full of defect-like structures, it was only natural to assume that more defective parts of the structure allow easier atomic transport. This led to many ideas of defects in liquid and glass, including free-volume (Cohen and Turnbull, 1959), cooperatively rearranging region (Adam and Gibbs, 1965), shear-transformation-zone (Argon, 1979), and others including ours (Egami et al., 1980). However, the results above suggest that the bulk properties, the MRO, control atomic transport, not those of defects. Would the concept of defect be still relevant in elucidating the atomic transport? Our answer is that the concept of defect defined by the specific static structure is not applicable to liquid and glass. We have to consider the "structure" as a dynamic entity.

It has been recognized for a long time that the definition of defect in amorphous system is arbitrary, in the absence of the reference structure. Various attempts have been made to define the defects, by studying the nature of the static structure, including the approaches using machine-learning (Cubuk et al., 2015; Bapst et al., 2020). However, it became apparent recently that what matters is the dynamics, not the static structure before deformation. In crystalline solids the defect retains its structural identity even after motion, because of the translational symmetry of the host lattice. In other words, defects are topologically protected by the lattice. In liquid and glass, however, the topology of atomic connectivity is open, and defects are not topologically protected. The atomic configuration before the motion of a defect is very different from that after the motion. In the picture of the potential energy landscape (PEL), the system moves from one valley to the other through a saddle-point. It was fond that, at the saddle-point, the potential energy of the system is high enough for the system to melt locally for a very short time (∼1 ps) (Ding et al., 2020). Consequently, the system loses the memory of prior thermal history (Fan et al., 2017). The saddle-point is known to be a generator of chaos (Mason and Piiroinen, 2012; Párraga et al., 2018). The simple, usually hand-written, schematic picture of the PEL gives an impression that the pathway from one valley to the next is pre-determined. However, in reality, the kinetic momenta of atoms, which vary rapidly in time, give rise to large uncertainty in the directions toward which the system evolves. The major virtue of the PEL concept is that by removing the kinetic energy the underlying PEL is clearly exposed. However, to describe the dynamics of the system we need to add back the kinetic energy which introduces uncertainty, particularly at the saddle-point.

The local melting at the saddle-point decouples the pathway from a valley of the PEL up to the saddle-point and the pathway down to another valley. Therefore, what determines the nature of the saddle-point, thus the system dynamics, is not the initial state in the prior valley but the nature of the molten state which reflects the bulk property, such as the MRO. The propensity to start the activation process depends on the energy of the initial state, which can be described in terms of the fictive, or effective, temperature in the glassy state (Langer, 2004; Fan et al., 2017). However, once the process of activation over the PEL saddle point starts it does not matter where it started initially. The dynamics of the system at the saddle point is totally controlled by the bulk properties, the coherence volume, to be specific.

Mode-Coupling Theory

The mode-coupling theory (MCT) is one of the most widely used theories of liquid dynamics. It describes the dynamics in terms of continuum hydrodynamic variables, such as density and current auto-correlation functions. It is based on the Boltzmann-type equation of motion initially developed for colloids. In the equation of motion, the dynamics at time t is coupled to the dynamics at a prior time t' through the memory function which represents the frictional force. The input to the theory is the snapshot structure function, $S(Q)$, particularly its first peak. Because the height of the first peak of $S(Q)$ is proportional to ξ_s (Ryu et al., 2019), the MCT focuses on the MRO, similarly to our approach. The dynamic correlations are determined by the equation of motion, and the feedback through the memory function determines dynamics, leading to the glass transition.

In colloids, particles are in touch with solvent, which is in the hydrodynamic steady state, so that the use of the frictional term in the Boltzmann equation is justified. However, in atomic liquids atoms interact each other directly via the potential force, so the application of the MCT becomes more contorted. Viscosity is given in terms of the stress autocorrelation via the Green-Kubo equation. In the MCT this retention of stress correlation is expressed as the memory function which gives rise to the frictional force. Therefore, the feedback from the memory function can produce a runaway leading to the divergence of correlation time.

In our approach the dynamics is governed by discrete local atomic activation processes. The probability of activation is controlled by the activation energy which is directly related to the structural coherence length ξ_s. The ξ_s is an equilibrium property, which depends only on temperature and the elastic constants through the atomic-level stresses, without the feedback loop through the memory function. In colloids local dynamics is closely coupled to local density fluctuations, because density plays the role of temperature in the hard-sphere system. Hard jamming at the critical density leads to divergence of viscosity. In atomic liquids, however, hard jamming never occurs, because atoms are compressible and thermal activation is always possible. Even though the MCT explains the glass transition of colloidal systems, its applicability to atomic liquids should be examined more carefully.

Infinite Dimension Models

The spin-glass theories of Edwards and Anderson (1975) and Sherrington and Kirkpatrick (1975) used the replica method (Aharony, 1975; Emery, 1975) and established the presence of the spin-glass ground state, at least in the infinite dimensions. In many spin-glasses spins interact through the long-range RKKY interaction. The large number of interacting neighboring spins justifies the use of the mean-field approximation. The replica method was applied later to the glass problem (Mézard and Parisi, 2000).

In spin-glasses randomness is quenched, because the spin Hamiltonian does not change with temperature. In contrast in real liquids and glasses the Hamiltonian varies with time and temperature. The number of atoms involved in action, n_c, is small. Therefore, a similar mean-field approximation is more difficult to justify, and atomic discreteness becomes central to the dynamics. For instance, at T_A, $\xi_s\left(T_g\right)/a \approx 1$, so in Equation (7) E_0 represents the bond energy and n_c, ~ 2. The dynamics is totally local, and the action of cutting a bond determines viscosity and diffusivity. Even at T_g, n_c ranges from 2 to 12, whereas the ideal state, where $\xi_s\left(T_g\right)/a \to \infty$, is achieved only at a negative temperature. Thus, the liquid above T_g is very far from the ideal state. The infinite dimension theories may be justified in the ideal state, but they may not be appropriate for the real glass and liquid which are far removed from the ideal state. The glass theories based on exact solutions in infinite dimensions (Parisi et al., 2020) are beautiful, but the success of its application to real liquids and glasses needs to be proven.

CONCLUSIONS

The study of the structural medium-range-order (MRO) in metallic liquids, represented by the coherence length, ξ_s, through diffraction experiment and simulation shows that the MRO is intimately related to local dynamics and viscosity. Namely the activation energy of viscosity is directly related to the number of atoms involved in local atomic rearrangement for structural excitation, $n_c(T)$, which is proportional to $(\xi_s)^3$. The magnitude of $n_c(T)$ is relatively small, $2 \sim 12$ even at T_g, so that the discrete nature of the atomic structure, represented by the topology of atomic connectivity network, is crucial. Conversely, it means that the unit energy for activation per atomic bond, E_B, is relatively large. Therefore, a small increase in $n_c(T)$ would result in large increase in the activation energy and viscosity. In our view this must be the reason why the timescale of liquid dynamics changes so rapidly over a moderate temperature range. At the same time, the system is quite far from the point of viscosity divergence which occurs when $n_c(T)$ diverges to infinity. Actually, we predict $n_c(T)$ to diverge at a negative temperature, by extrapolation with the Curie-Weiss law. Therefore, the Kauzmann catastrophe never

occurs at $T > 0$. Even though the results presented here focus on metallic liquids, the same approach was successful in elucidating the ideality and fragility of network and some organic liquids (Ryu and Egami, 2020; Ryu et al., 2020), suggesting that this approach may be applicable beyond metallic liquids. This view is at odds with some of the theories and ideas. This conflict will be resolved by further theoretical and experimental advances in the future.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article.

AUTHOR CONTRIBUTIONS

This work was conceived and authored by TE with assistance by CR. The data in this work were generated and analyzed by CR. Both authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division.

REFERENCES

Adam, G., and Gibbs, J. H. (1965). On the temperature dependence of cooperative relaxation properties in glass forming liquids. *J. Chem. Phys.* 43, 139–146. doi: 10.1063/1.1696442

Aharony, A. (1975). Tetracritical point in mixed magnetic crystals. *Phys. Rev. Lett.* 34, 590–593. doi: 10.1103/PhysRevLett.34.590

Angell, C. A. (1995). Formation of glasses from liquids and biopolymers. *Science* 267, 1924–1935. doi: 10.1126/science.267.5206.1924

Argon, A. S. (1979). Plastic deformation in metallic glasses. *Acta Metall.* 27, 47–58. doi: 10.1016/0001-6160(79)90055-5

Ashcraft, R., Wang, Z., Abernathy, D. L., Egami, T., and Kelton, K. F. (2020). Experimental determination of the temperature-dependent Van Hove function in a $Zr_{80}Pt_{20}$ liquid-a structural origin for the dynamics at high temperature. *J. Chem. Phys.* 152:074506. doi: 10.1063/1.5144256

Bapst, V., Keck, T., Grabska-Barwińska, A., Donner, C., Cubuk, E. D., Schoenholz, S. S., et al. (2020). Unveiling the predictive power of static structure in glassy systems. *Nat. Phys.* 16, 448–454. doi: 10.1038/s41567-020-0842-8

Bellissard, J., and Egami, T. (2018). Simple theory of viscosity in liquids. *Phys. Rev. E* 98:063005. doi: 10.1103/PhysRevE.98.063005

Berthier, L., and Biroli, G. (2011). Theoretical perspective on the glass transition and amorphous materials. *Rev. Mod. Phys.* 83:587. doi: 10.1103/RevModPhys.83.587

Biroli, G., Bouchaud, J.-P., Miyazaki, K., and Reichman, D. (2006). Inhomogeneous mode-coupling theory and growing dynamic length in supercooled liquids. *Phys. Rev. Lett.* 97:195701. doi: 10.1103/PhysRevLett.97.195701

Blodgett, M. E., Egami, T., Nussinov, Z., and Kelton, K. F. (2015). Proposal for universality in the viscosity of metallic liquids. *Sci. Rep.* 5:13837. doi: 10.1038/srep13837

Cargill, G. S. III. (1975). Structure of metallic alloy glasses. *Solid State Phys.* 30, 227–320. doi: 10.1016/S0081-1947(08)60337-9

Cohen, M. H., and Grest, G. (1979). Liquid-glass transition, a free-volume approach. *Phys. Rev. B* 20, 1077–1098. doi: 10.1103/PhysRevB.20.1077

Cohen, M. H., and Turnbull, D. (1959). Molecular transport in liquids and glasses. *J. Chem. Phys.* 31, 1164–1169.

Cubuk, E. D., Schoenholz, S. S., Rieser, J. M., Malone, B. D., Rottler, J., Durian, D. J., et al. (2015). Identifying structural flow defects in disordered solids using machine-learning methods. *Phys. Rev. Lett.* 114:108001. doi: 10.1103/PhysRevLett.114.108001

Debenedetti, P. G., and Stillinger, F. H. (2001). Supercooled liquids and the glass transition. *Nature* 410, 259–267. doi: 10.1038/35065704

Demetriou, M. D., Harmon, J. S., Tao, M., Duan, G., Samwer, K., and Johnson, W. L. (2006). Cooperative shear model for the rheology of glass-forming metallic liquids. *Phys. Rev. Lett.* 97:065502. doi: 10.1103/PhysRevLett.97.065502

Ding, J., Li, L., Wang, N., Tian, L., Asta, M., Ritchie, R. O., et al. (2020). Nature of the saddle states of structural excitations in Cu64Zr36 metallic glasses.

Donth, E. (2010). *The Glass Transition: Relaxation Dynamics of Liquids and Disordered Materials*. Berlin: Springer.

Dyre, J. C. (2006). The glass transition and elastic models of glass-forming liquids. *Rev. Mod. Phys.* 78, 953–972. doi: 10.1103/RevModPhys.78.953

Edigar, M. D., and Harrowell, P. (2012). Perspective: supercooled liquids and glasses. *J. Chem. Phys.* 137:080901. doi: 10.1063/1.4747326

Edwards, S. F., and Anderson, P. W. (1975). Theory of spin glasses. *J. Phys. F Met. Phys.* 5, 965–974. doi: 10.1088/0305-4608/5/5/017

Egami, T. (2011). Atomic level stresses. *Prog. Mater. Sci.* 56, 637–653. doi: 10.1016/j.pmatsci.2011.01.004

Egami, T. (2020). Local density correlations in liquids. *Front. Phys.* 8:50. doi: 10.3389/fphy.2020.00050

Egami, T., Maeda, K., and Vitek, V. (1980). Structural defects in amorphous solids: a computer simulation study. *Phil. Mag. A* 41, 883–901. doi: 10.1080/01418618008243894

Egami, T., and Ryu, C. W. (2020). Origin of the medium-range order in liquid.

Elmatad, Y. S., Chandler, D., and Garrahan, J. P. (2009). Corresponding states of structural glass formers. *J. Phys. Chem. B* 113, 5563–5567. doi: 10.1021/jp810362g

Emery, V. J. (1975). Critical properties of many-component systems. *Phys. Rev. B* 11, 239–247. doi: 10.1103/PhysRevB.11.239

Fan, Y., Iwashita, T., and Egami, T. (2017). Energy landscape-driven nonequilibrium evolution of inherent structure in disordered material. *Nature Commun.* 8:15417.

Frank, F. C. (1952). Supercooling of liquids. *Proc. Roy. Soc. Lond. A* 215, 43–46. doi: 10.1098/rspa.1952.0194

Fulcher, G. S. (1925). Analysis of recent measurements of the viscosity of glasses. *J. Am. Ceram. Soc.* 8, 339–355. doi: 10.1111/j.1151-2916.1925.tb16731.x

Gaskell, P. H. (1979). A new structural model for amorphous transition metal silicides, borides, phosphides, and carbides. *J. Non-Cryst. Solids* 32, 207–224. doi: 10.1016/0022-3093(79)90073-5

Götze, W. (2009). *Complex Dynamics of Glass-Forming Liquids*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199235346.001.0001

Iwashita, T., Nicholson, D. M., and Egami, T. (2013). Elementary excitations and crossover phenomenon in liquids. *Phys. Rev. Lett.* 110:205504. doi: 10.1103/PhysRevLett.110.205504

Iwashita, T., Wu, B., Chen, W.-R., Tsutsui, S., Baron, A. Q. R., and Egami, T. (2017). Seeing real-space dynamics of liquid water through inelastic x-ray scattering. *Sci. Adv.* 3:1603079. doi: 10.1126/sciadv.1603079

Kato, H., Wada, T., Hasegawa, M., Saida, J., Inoue, A., and Chen, H. S. (2006). Fragility and thermal stability of Pt- and Pd- based bulk glass forming liquids and their correlation with deformability. *Scripta Mat.* 54, 2023–2027. doi: 10.1016/j.scriptamat.2006.03.025

Kauzmann, W. (1948). The nature of the glassy state and the behavior of liquids at low temperatures. *Chem. Rev.* 43, 219–256. doi: 10.1021/cr60135a002

Kivelson, D., Kivelson, S. A., Zhao, X., Nussinov, Z., and Tarjus, G. (1995). A thermodynamic theory of supercooled liquids. *Physica A* 219, 27–38. doi: 10.1016/0378-4371(95)00140-3

Langer, J. S. (2004). Dynamics of shear-transformation zones in amorphous plasticity: formulation in terms of an effective disorder temperature. *Phys. Rev. E* 70:041502.

Lubchenko, V., and Wolynes, P. G. (2007). Theory of structural glasses and supercooled liquids. *Ann. Rev. Phys. Chem.* 58, 235–266. doi: 10.1146/annurev.physchem.58.032806.104653

March, N. H., and Tosi, M. P. (2002). *Introduction to Liquid State Physics.* Singapore: World Scientific. doi: 10.1142/4717

Mason, J. F., and Piiroinen, P. T. (2012). Saddle-point solutions and grazing bifurcations in an impacting system. *Chaos* 22:013106. doi: 10.1063/1.3673786

Mauro, J. C., Yue, Y., Ellison, A. J., Gupta, P. K., and Allan, D. C. (2009). Viscosity of glass-forming liquids. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19780–19784. doi: 10.1073/pnas.0911705106

Mézard, M., and Parisi, G. (2000). Statistical physics of structural glass. *J. Phys. Cond. Mat.* 12, 6655–6673. doi: 10.1088/0953-8984/12/29/336

Mohr, M., Wunderlich, R. K., Zweiacker, K., Prades-Rödel, S., Sauget, R., Blatter, A., et al. (2019). Surface tension and viscosity of liquid $Pd_{43}Cu_{27}Ni_{10}P_{20}$ measured in a levitation device under microgravity. *npj Microgravity* 5:4. doi: 10.1038/s41526-019-0065-4

Nelson, D. R. (1983). Order, frustration, and defects in liquids and glasses. *Phys. Rev. B* 28, 5515–5535.

Novikov, V. N., and Sokolov, A. P. (2004). Poisson's ratio and the fragility of glass-forming liquids. *Nature* 431, 961–963. doi: 10.1038/nature02947

Nussinov, Z. (2004). Avoided phase transitions and glassy dynamics in geometrically frustrated systems and non-Abelian theories. *Phys Rev B* 69:014208. doi: 10.1103/PhysRevB.69.014208

Ornstein, L. S., and Zernike, F. (1914). Accidental deviations of density and opalescence at the critical point of a single substance. *R. Netherlands Acad. Arts Sci.* 17, 793–806.

Parisi, G., Urbani, P., and Zamponi, F. (2020). *Theory of Simple Glasses.* Cambridge: Cambridge University Press. doi: 10.1017/9781108120494

Párraga, H., Arranz, F. J., Benito, R. M., and Borondo, F. (2018). Above saddle-point regions of order in a sea of chaos in the vibrational dynamics of KCN. *J. Phys. Chem. A* 122, 3433–3441. doi: 10.1021/acs.jpca.8b00113

Ryu, C. W., Dmowski, W., and Egami, T. (2020). Ideality of liquid structure: a case study for metallic alloy liquids. *Phys. Rev. E* 101:030601. doi: 10.1103/PhysRevE.101.030601

Ryu, C. W., Dmowski, W., Kelton, K. F., Lee, G. W., Park, E. S., Morris, J. R., et al. (2019). Curie-weiss behavior of liquid structure and ideal glass state. *Sci. Rep.* 9:18579. doi: 10.1038/s41598-019-54758-y

Ryu, C. W., and Egami, T. (2020). Origin of Liquid Fragility.

Sadoc, J. F. (1981). Use of regular polytopes for the mathematical description of the order in amorphous structures. *J. Non-Cryst. Solids* 44, 1–16.

Sethna, J. P. (1983). Frustration and curvature: glasses and the cholesteric blue phase. *Phys. Rev. Lett.* 51:2198.

Sherrington, D., and Kirkpatrick, S. (1975). Solvable model of a spin glass. *Phys. Rev. Lett.* 35, 1792–1796. doi: 10.1103/PhysRevLett.35.1792

Shinohara, Y., Dmowski, W., Iwashita, T., Wu, B., Ishikawa, D., Baron, A. Q. R., et al. (2019). Erratum: Viscosity and real space molecular motion of water: observation with inelastic X-ray scattering [Phys. Rev. E 98, 022604 (2018)]. *Phys. Rev. E* 100:039904. doi: 10.1103/PhysRevE.98.022604

Steinhardt, P. J., Nelson, D. R., and Ronchetti, M. (1981). Icosahedral bond order in supercooled liquids. *Phys. Rev. Lett.* 47, 1297–1300. doi: 10.1103/PhysRevLett.47.1297

Tammann, G., and Hesse, G. (1926). The dependency of viscosity on temperature in hypothermic liquids. *Z. Anorg. Allg. Chem.* 156, 245–257. doi: 10.1002/zaac.19261560121

Tanaka, H., Kawasaki, T., Shintani, H., and Watanabe, K. (2010). Critical-like behavior of glass-forming liquids. *Nat Mater.* 9, 324–331. doi: 10.1038/nmat2634

Tomida, T., and Egami, T. (1995). Molecular-dynamics study of orientational order in liquids and glasses and its relation to the glass transition. *Phys. Rev. B* 52, 3290–3308. doi: 10.1103/PhysRevB.52.3290

Vogel, H. (1921). The temperature dependence law of the viscosity of fluids. *Z Phys.* 22, 645–646.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 15 October 2020
doi: 10.3389/fchem.2020.585853

Evolving Stark Effect During Growth of Perovskite Nanocrystals Measured Using Transient Absorption

James C. Sadighian[1], *Kelly S. Wilson*[1], *Michael L. Crawford*[1] *and Cathy Y. Wong*[1,2,3]*

[1] *Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, United States,* [2] *Oregon Center for Optical, Molecular, and Quantum Science, University of Oregon, Eugene, OR, United States,* [3] *Materials Science Institute, University of Oregon, Eugene, OR, United States*

Methylammonium lead triiodide ($MAPbI_3$) nanocrystals (NCs) are emerging materials for a range of optoelectronic applications. Photophysical characterization is typically limited to structurally stable NCs owing to the long timescales required for many spectroscopies, preventing the accurate measurement of NCs during growth. This is a particular challenge for non-linear spectroscopies such as transient absorption. Here we report on the use of a novel single-shot transient absorption (SSTA) spectrometer to study $MAPbI_3$ NCs as they grow. Comparing the transient spectra to derivatives of the linear absorbance reveals that photogenerated charge carriers become localized at surface trap states during NC growth, inducing a TA lineshape characteristic of the Stark effect. Observation of this Stark signal shows that the contribution of trapped carriers to the TA signal declines as growth continues, supporting a growth mechanism with increased surface ligation toward the end of NC growth. This work opens the door to the application of time-resolved spectroscopies to NCs *in situ*, during their synthesis, to provide greater insight into their growth mechanisms and the evolution of their photophysical properties.

Keywords: perovskite, nanocrystals, surface, ultrafast, transient absorption, spectroscopy, Stark effect

OPEN ACCESS

Edited by:
Rene A. Nome,
State University of Campinas, Brazil

Reviewed by:
Arkaprabha Konar,
Kent State University, United States
Minsu Jung,
Dong-Eui University, South Korea

***Correspondence:**
Cathy Y. Wong
cwong3@uoregon.edu

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Chemistry

Received: *21 July 2020*
Accepted: *28 August 2020*
Published: *15 October 2020*

Citation:
Sadighian JC, Wilson KS, Crawford ML and Wong CY (2020) Evolving Stark Effect During Growth of Perovskite Nanocrystals Measured Using Transient Absorption. Front. Chem. 8:585853. doi: 10.3389/fchem.2020.585853

1. INTRODUCTION

Hybrid organic-inorganic perovskite nanocrystals (NCs) are currently the focus of significant interest owing to their potential applications in optoelectronic devices. Their large absorption coefficients (Fu et al., 2015), high defect tolerance (Dirin et al., 2016), excellent photoluminescence quantum yield (PLQY) (Hassan et al., 2019), and potential for low-cost, facile production (Protesescu et al., 2015) coupled with a narrow, tuneable emission spectrum (Hassan et al., 2016) has driven a boom of research in the synthesis and characterization of these materials. These NCs are ordinarily grown through either a hot injection or reprecipitation style synthesis. In these solution-based syntheses the reaction is initiated when dissolved precursor reaches a critical threshold to cause LaMer nucleation (LaMer and Dinegar, 1950). Following this, NCs are allowed to grow until the desired size and morphology is reached. The morphology (Pan et al., 2016; Sun et al., 2016), stability (Huang et al., 2017), and photophysics (Peterson et al., 2014; Teunis et al., 2017) of NCs are strongly dependent on the surface owing to their large surface-to-volume ratios. Surface atoms lacking bonds to capping ligands exhibit localized electronic states with energies that can lie within the band gap. These mid-gap states act as traps for excited electrons or holes, suppressing radiative recombination and hampering performance in light emitting devices (Boles et al., 2016).

The quality of the NC surface during growth is still poorly understood and the timescales of nucleation and growth are prohibitively short for investigation using typical surface

characterization techniques, such as X-ray photoelectron spectroscopy (Katari et al., 1994), electron energy loss spectroscopy (Wang et al., 1998), small-angle X-ray scattering (Mattoussi et al., 1998), and 2D nuclear magnetic resonance techniques (De Roo et al., 2016). Recently, use of a solvation-mediated synthesis (Wang et al., 2017), coupled to a rapid sampling technique (Sadighian et al., 2019), permitted the measurement of linear absorbance and fluorescence during growth. This study revealed that NCs initially grow in size while their surfaces remain poorly-capped by passivating ligands, and do not become well-capped until they are almost fully grown (**Figure 1**) (Sadighian et al., 2020). Visible absorbance and fluorescence measurements report on transitions from the ground and emissive band-edge states, respectively. The peak positions and lineshapes can provide insight into the NC size distribution, and fluorescence intensity is often used to infer the degree of NC surface passivation. However, these spectroscopies are insensitive to other important transitions, such as carrier trapping and non-radiative recombination, and the dynamics of the excited carriers. A comprehensive understanding of how NC photophysics evolves during synthesis may provide deeper insights into NC growth mechanisms, the nature of the NC surface, and how a synthesis can be tuned to achieve desired morphologies and optoelectronic properties.

In this paper, we demonstrate a technique that can provide further insight into the evolving NC surface by probing the electric field generated by carriers localized at surface traps. Photogenerated electron-hole pairs become spatially separated when a carrier is trapped at these surface sites, creating an electric field inside the NC. The presence of an electric field can modulate the optical transitions of an NC *via* the Stark effect (Colvin and Alivisatos, 1992; Colvin et al., 1994; Klimov, 2000; Sharma et al., 2019). Analysis of the modulated absorbance spectrum lineshape can provide insight into the electric fields in the NCs (Bublitz and Boxer, 1997). The quantum-confined Stark effect (QCSE) changes the bandgap transition energy by shifting the electron and hole energy levels (Walters et al., 2018). This typically redshifts the bandedge absorption and causes the differential absorbance spectrum to exhibit the lineshape of the first derivative of the linear absorbance. In systems that lack any specific orientation, such as randomly distributed surface traps on NCs, the internal electric field generated by spatially separated, trapped carriers results in a population of randomly oriented dipoles in the sample. This would act to inhomogenously broaden the overall transition, and as a result the differential absorbance spectrum would resemble the second derivative of the linear absorbance (Tanaka and Kondo, 2003; Queloz et al., 2020).

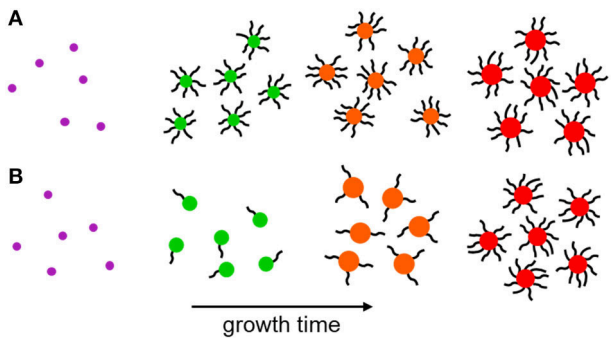

FIGURE 1 | Schematic of NC growth. **(A)** Following LaMer nucleation the immature NCs are well-capped by surface ligands throughout most of their subsequent growth. **(B)** Nascent NCs nucleate following the same LaMer-type mechanism, but surface ligation occurs primarily after NC growth. A terminal ligation stage is proposed in literature for CdSe and $MAPbI_3$ NCs and is supported by the measurements in this work.

Electroabsorbance measurements of 2D hybrid perovskites have exhibited lineshapes that could be fit to a weighted sum of first and second derivatives of the absorbance spectrum (Queloz et al., 2020). These two components were assigned to a spectral redshift arising from a QCSE and broadening due to loosely-bound, screened electron-hole pairs, respectively. This same lineshape was observed upon photogeneration of charge carriers in these perovskites during transient absorption (TA) measurements. This indicates that the presence of spatially separated electrons and holes in surface traps can cause internal electric fields that yield lineshapes characteristic of the Stark effect. Thus, the Stark lineshape measured by TA can report on the surface quality of NCs.

TA is a powerful time-resolved spectroscopy that has been used to understand excited state processes such as Auger recombination (Klimov and McBranch, 1998; Guyot-Sionnest et al., 1999), energy transfer to phonons (Urayama et al., 2001) or ligands (Guyot-Sionnest et al., 2005; Li et al., 2019), and carrier trapping (Mondal and Samanta, 2017) in NCs. Typically, this pump-probe technique is performed by varying the path length of one pulse relative to the other by use of a retroreflector on a motorized translation stage. The transmission of many successive laser pulses is recorded at each pump-probe time delay in series. This technique typically requires measurement timescales on the order of tens of minutes to several hours, depending on factors such as sample response and laser noise. As a result, in its typical implementations TA fails to accurately report on excited state dynamics in non-equilibrium systems that are chemically changing on timescales shorter than a few hours, such as growing NCs.

TA measurements can be conducted more rapidly by using a single-shot transient absorption (SSTA) spectrometer that enables an entire transient to be recorded from a single pump-probe pulse pair. This can be achieved by tilting the wavefront of the pump pulse relative to the probe (Fourkas et al., 1995; Makishima et al., 2006). In this case, the time delay range is determined by Equation (1):

$$t_{range} = \frac{dsin(\theta)}{c} \tag{1}$$

where d is the length of overlap between pump and probe pulses, θ is the angle between the tilted pump pulse and the probe pulse, and c is the speed of light (**Figure 2**). Here, we use a recently developed broadband SSTA spectrometer (Wilson

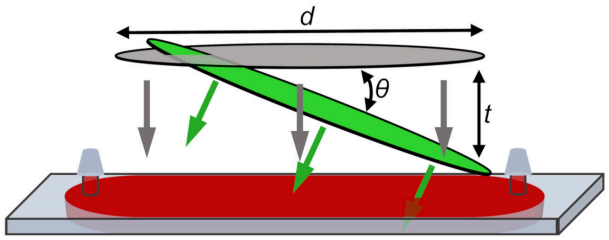

FIGURE 2 | Sample plane in SSTA instrument showing pump (green) and probe (gray) focused to lines. Probe pulse is incident normal to the cuvette. Spatially encoded time delay, t, is generated by the angle of the pump pulse relative to the probe pulse, θ, and the length of overlap between the two pulses, d. Sample is injected into the flow-cell cuvette through septa to prevent solvent evaporation during the measurement.

and Wong, 2018; Wilson et al., 2020) to track the evolution of exciton dynamics in methylammonium lead triiodide ($MAPbI_3$) perovskite NCs as they nucleate and grow and as their surfaces are passivated with ligands. A complete TA spectrum with excellent signal to noise can be collected using this instrument in less than 1 min, allowing us to accurately measure immature NCs before they degrade (Sadighian et al., 2019). As a result, we are able to observe the surface of NCs being capped in real time by monitoring the evolving Stark lineshape. A carrier that has been photoexcited by the pump may localize on a surface trap state, creating an electric field within the NC. Using differential measurement, the probe then reports the effect of an ensemble of these electric fields on the absorption of the NC sample. These findings agree with previous reports of the growth mechanism of CdSe (Teunis et al., 2017) and $MAPbI_3$ (Sadighian et al., 2020) NCs, and open up a new avenue for studying the surface of these materials during growth.

2. METHODS

2.1. Materials

All reagents were used as received: lead iodide (99.999%, trace metals basis, Sigma-Aldrich), methylammonium iodide (MAI, ≥99%, anhydrous, Sigma-Aldrich), octylamine (99%, Sigma-Aldrich), oleic acid (90%, technical grade, Sigma-Aldrich), and hexane (≥95%, laboratory reagent grade, Sigma-Aldrich). Cresyl violet (62%, J.T. Baker) in methanol (99.8%, Certified ACS, Fisher) was used to calibrate the beam profile and spatially encoded time delay of the SSTA spectrometer.

2.2. Nanocrystal Synthesis

$MAPbI_3$ NCs were synthesized using a previously reported solvation-limited synthesis (Sadighian et al., 2019, 2020). 460 mg of PbI_2 and 127 mg of MAI were combined with 40 mL of hexane in a glass test tube and suspended in an ultrasonication bath (VWR, 97043-992) to provide constant mixing. The reaction was initiated with the simultaneous introduction of 150 μL octylamine and 300 μL oleic acid, and the recorded reaction time is in reference to this addition. These organic ligands act to solubilize PbI_2 and MAI, which are otherwise insoluble in hexane (Wang et al., 2017). A recirculating chiller (VWR, 1165) in a closed-loop configuration with an aluminum block was used to maintain a temperature of 22°C in the ultrasonication bath. An HDPE syringe was used to withdraw aliquots of the reaction mixture at selected time points. Each aliquot was filtered through a syringe filter (VWR) with a 0.45 μm pore polytetrafluoroethylene (PTFE) membrane and into a 0.2 mm path length quartz flow cell cuvette (Starna Cells, 48-Q-0.2). Following the 15 min mark a 1.0 μm PTFE pre-filter (Whatman Rezist) was used in conjunction with the 0.45 μm filter to compensate for increased suspended particulate. An additional 5.0 μm filter (Whatman Rezist) was added after 60 min. The flow cell was emptied, rinsed with acetone and hexane, and dried with a stream of nitrogen before each successive measurement.

2.3. Absorbance and Fluoresence

Absorbance and fluorescence of the filtered NC aliquots were simultaneously recorded on a homebuilt spectrometer (**Supplementary Figure 1**) using the same cuvette and sample described above. To measure absorbance, light from a tungsten-halogen lamp (Thorlabs, SLS201) was directed into the sample using a fiber optic cable (Thorlabs, M28L01) and the resulting transmission collected using a second fiber. A 405 nm laser (Thorlabs, CPS405) was used as the fluorescence excitation source. Emitted light was collected using a fiber optic cable (Thorlabs, M95L01) directed to the spot upon which the laser was incident on the cuvette and angled to avoid the specular reflection of the excitation source. Absorbance and fluorescence spectra were recorded using an Ocean Optics Flame-T-VIS-NIR and Flame-T-UV-VIS spectrometer, respectively. The spectrometers were operated using a homebuilt Python software package. Absorbance and fluorescence were recorded immediately before and after collecting SSTA measurements of each aliquot to make sure the spectra did not change significantly during the measurement. The pairs of spectra were then averaged together for analysis.

2.4. Single-Shot Transient Absorption

SSTA measurements were performed using a previously described homebuilt instrument (Wilson and Wong, 2018; Wilson et al., 2020). A 1 kHz Ti:sapphire laser (Coherent, Astrella) with an 800 nm output was used to pump an optical parametric amplifier (Light Conversion, Topas Prime Plus) to generate 520 nm pump pulses that were compressed to 50 fs using a prism pair. A 2 m focal length mirror focused part of the 800 nm fundamental in a 1.6 m homebuilt gas cell with 1.5 mm quartz windows and containing 0.55 bar differential pressure of argon (PurityPlus, 99.999%) to generate broadband probe pulses. The spectral profiles of both pulses are shown in **Supplementary Figure 2**. The pump and probe pulses were optically chopped at 250 and 125 Hz, respectively. The addition of a chopper in the probe line enabled the subtraction of background signals arising from pump induced fluorescence, scatter, stray light, and dark current from the camera (Wilson et al., 2019). A spatial light modulator (Meadowlark, 1920 × 1152 XY Phase Series SLM) placed after the choppers was used to reshape both beams to a flat-top intensity profile to provide

a uniform excitation density across the entire spatially encoded time delay range.

The pump pulse energy at the sample was set to 410 nJ to prevent non-linear interactions. The pump and probe beams were focused to lines using cylindrical lenses with focal lengths of 200 mm and 150 mm, respectively, and overlapped on a 20 μm × 22 mm area of the cuvette. While the probe was incident normal to the sample, the pump beam was tilted at an angle of 55.5° to achieve a spatially encoded time delay of 60 ps. The probe beam at the sample plane was imaged onto the slit of a grating spectrograph (Princeton Instruments, Isoplane 160), where it was measured to be 10 nJ. The probe beam was slightly defocused at the sample plane such that the entire measured wavelength range overlaps well onto the slit of the spectrograph with sufficient intensity. The spectrograph was coupled to a CMOS camera (Andor, Zyla 5.5) with a 1.3 ms exposure time which acquires 180 x 2560 pixel (1.17 × 16.6 mm) images, with the signal at each pixel corresponding to a pump-probe time delay of 24 fs. One axis of the pixel array recorded wavelength resolution of the probe and the other captured the spatially encoded time delay. Each SSTA spectrum was recorded for 60 s to maximize signal-to-noise ratio while still avoiding sample degradation. The SSTA spectrometer was operated using homebuilt Python software. Spectral calibration was performed using a HgAr calibration source, which accounts for spherical aberrations in the imaging setup through the spectrometer. Calibration of the spatially encoded pump-probe time delay was performed using SSTA measurements of cresyl violet in methanol in the same cuvette used for the NC measurements. This process corrects for chirp in the broadband probe pulse. Both the wavelength and time delay calibrations are discussed in detail elsewhere (Wilson et al., 2020).

3. RESULTS AND DISCUSSION

Absorbance and fluorescence spectra at various reaction timepoints show the evolving physical and electronic structure of PTFE-filtered NCs over 120 min of growth (**Figure 3**). Nucleation occurred within the first 5 min, evidenced by the appearance of a broad, low intensity emission centered around 595 nm and weak absorbance near 525 nm. The fluorescence of the reaction mixture significantly increased in intensity by the 10 min mark and began to exhibit two distinct peaks. A feature emerged at 575 nm in the absorbance spectrum, which we ascribe to nascent NCs. These absorbance and fluorescence features continued to grow in intensity, reaching a maximum at the 30 min mark. Following this, the sharp absorbance peak at 575 nm began to disappear and gave rise to a broad shoulder at 610 nm, indicative of the small, nascent NCs growing larger. Likewise, the fluorescence spectrum began to lose intensity at shorter wavelengths while the peak at 635 nm continued to grow and redshift until the final measured timepoint. The evolution of these spectra are in agreement with previously reported experiments performed under similar conditions (Sadighian et al., 2019, 2020).

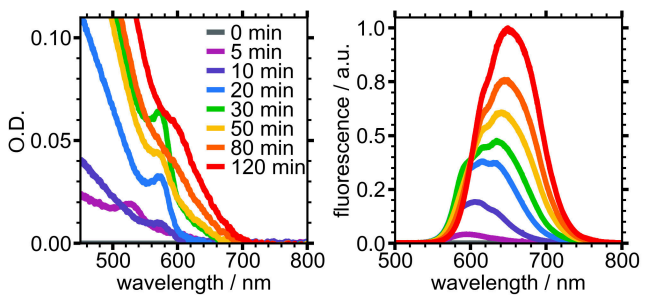

FIGURE 3 | Absorbance **(Left)** and fluorescence **(Right)** of PTFE-filtered reaction mixture sampled at different times during growth.

Select SSTA spectra for NCs at various stages of growth are shown in **Figure 4**. For each sample, the transient spectrum redshifts approximately 10 nm during the first 500 fs as a result of carrier cooling (Righetto et al., 2020). The spectra are quite stable for the remainder of the measured 60 ps time window. The spectrum of NCs grown for 120 min (**Figure 4D**) is typical of $MAPbI_3$ perovskite NCs (Wang et al., 2017). The negative TA at wavelengths longer than 600 nm overlaps with the band-edge absorbance and the emission spectrum. This feature is typically ascribed to a combination of stimulated emission (SE) and ground-state bleach (GSB). The signal at shorter wavelengths is broad and positive, indicating a photoinduced absorption (PIA) to higher electronic states. The SSTA spectra of NCs grown for 20, 30, and 50 min (**Figures 4A–C**) show two distinct features not present in the 120 min spectrum; a strong, narrow, negative TA signal centered at 582 nm and a region of low signal intensity near 600 nm. This signal reached its maximum in the 30 min sample and had all but disappeared 50 min into the reaction. The negative signal at 582 nm does not coincide with the absorbance peak (575 nm) and the fluorescence spectrum has a shoulder at 595 nm, suggesting neither GSB nor SE can explain this signal.

First and second derivatives of the absorbance spectra for the 20, 30, and 50 min NC samples are shown in **Figure 5**. The lineshape of the derivatives is similar across the three selected timepoints, with the magnitude of the derivative traces reaching their maximum in the 30 min sample when the sharp absorbance peak at 575 nm is most intense. This peak is less intense and broader in width in the 50 min sample, resulting in smaller derivatives for this timepoint.

In order to elucidate the origin of the TA lineshapes and gain additional insight into the electronic structure of growing NCs, a slice of the TA spectrum, reported in differential optical density (ΔO.D.) and averaged between 5 ps and 10 ps for each growth time, t, was fit using Equation (2).

$$\Delta\text{O.D.}(t,\lambda) = A\frac{d\text{O.D.}(t,\lambda)}{d\lambda} + B\frac{d^2\text{O.D.}(t,\lambda)}{d\lambda^2} + C\Delta\text{O.D.}(120\text{ min},\lambda) \quad (2)$$

The first two terms represent the first and second derivatives of the absorbance spectrum at the selected growth time and the third term is the analogous TA spectrum of the NC sample after 120 min of growth. This term accounts for the contribution of well-passivated NCs to the overall TA spectrum

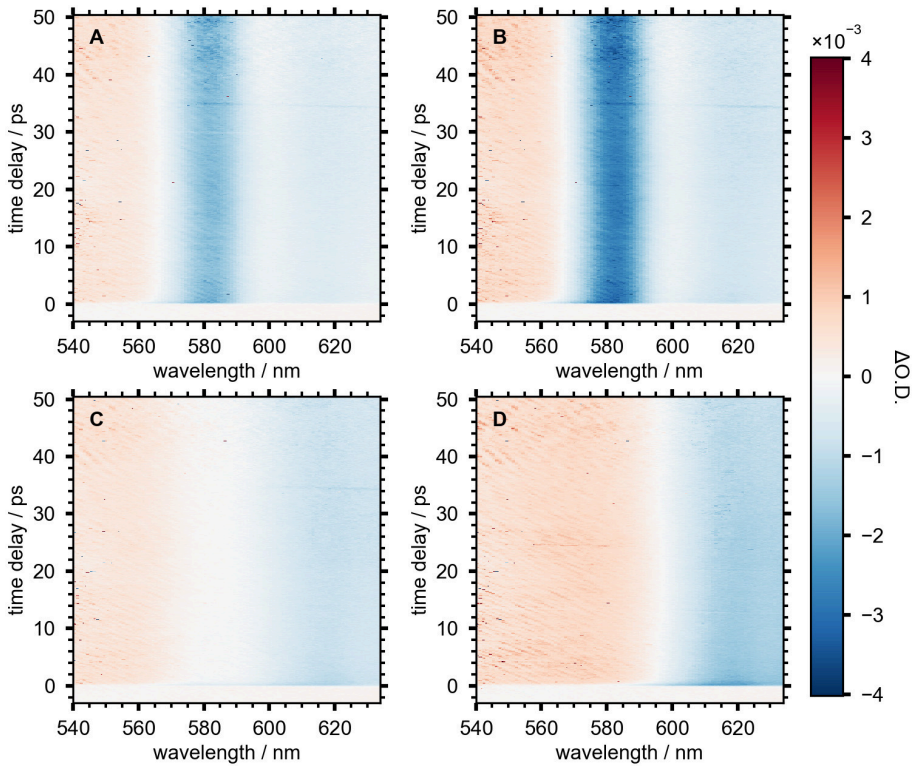

FIGURE 4 | SSTA spectra of NC aliquots measured after **(A)** 20 min, **(B)** 30 min, **(C)** 50 min, and **(D)** 120 min after starting the reaction.

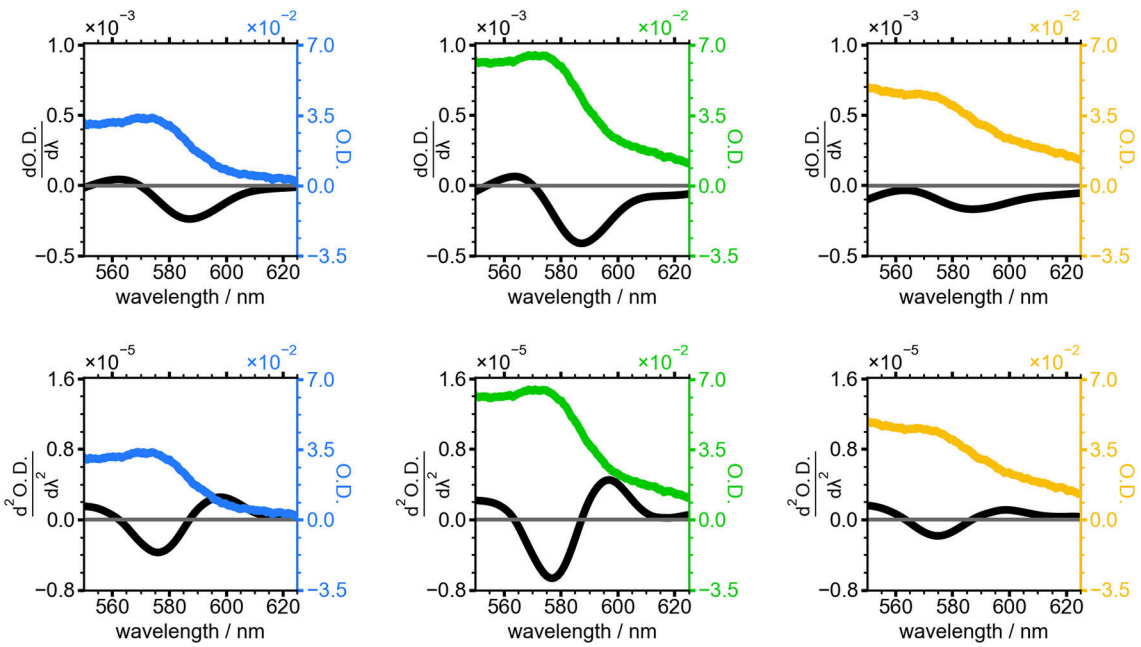

FIGURE 5 | Absorbance spectra of NCs grown for 20 min (blue), 30 min (green), and 50 min (yellow). First (top) and second (bottom) derivative of the absorbance spectrum for each time point is shown in black.

at each timepoint. The resulting fits are overlaid with TA slices in **Figure 6**. The colored, dashed lines are TA slices for the three time points from **Figures 4**, **5**, and the fits (solid black lines) show good agreement. These slices reveal the evolution of the electric field induced by electron-hole pairs generated by the pump pulse in the nascent NCs. The negative TA signal at 580 nm was clearly visible after 20 min of NC growth and reached a maximum after 30 min, indicating the presence of growing, poorly-capped NCs. During the remainder of the reaction this feature lost intensity and by 50 min was barely discernible.

The values of the three coefficients from Equation (2) are displayed with fit errors in **Table 1**. Tracking their values during the reaction quantifies the evolving contributions to the TA lineshapes (**Figure 7**). The first term, *A*, relates the observed signal to the first derivative of the absorbance, which occurs when the field causes a shift in the transition energy for the NCs. Here, the presence of spatially separated electrons and holes at surface traps would induce a dipole that could stabilize the excited electronic states, potentially redshifting the optical transition. The second derivative term, *B*, has the largest contribution to the signal throughout nearly the entire measured range. This term arises from an overall broadening of the absorbance spectrum, suggesting the presence of many randomly oriented dipoles in the sample arising from surface-trapped carriers.

The electric field produced by a trapped carrier should become smaller as a NC grows larger, so the decreasing contribution of the derivative lineshapes during NC growth could be the result of both increasing NC size and improved surface capping, resulting in fewer NCs with internal electric fields. While the contributions from both derivatives decline to zero over the course of the reaction, *B* shows a brief period of growth between 30 and 70 min into the reaction. The electric field strength at any particular time point during NC growth could be estimated from these results if the NC size were known, assuming that one carrier is surface-trapped while the other is delocalized (i.e., on average centrally located within the NC). Future work will focus on concurrent measurements of NC size during the reaction, which will enable the magnitude of the electric field caused by a surface-trapped carrier to be modeled during NC growth. This will aid in the interpretation of the rise in *B* while *A* continually decreases. *C*, the contribution of well-passivated NCs, shows a fairly linear growth throughout the entire synthesis. By the end of the reaction the NCs are well-capped with ligands, and surface traps no longer contribute to the TA signal. Thus, our

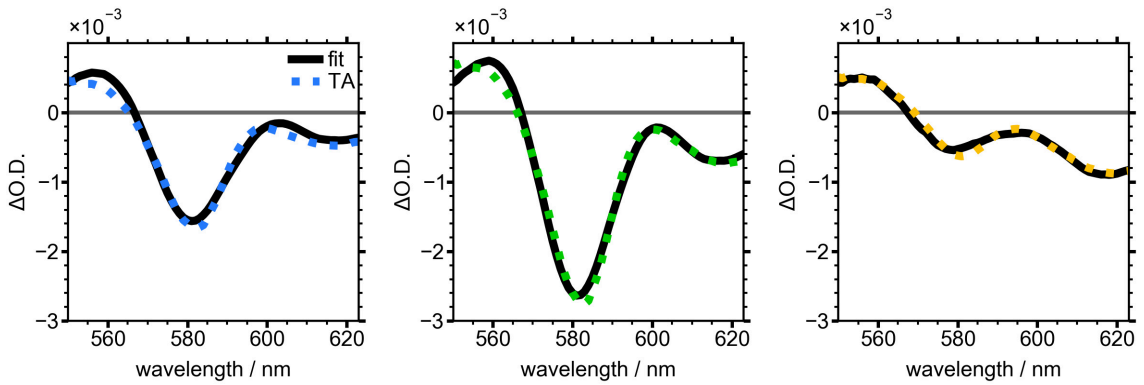

FIGURE 6 | Averaged TA spectra from 5 to 10 ps for NCs grown for 20 min (blue), 30 min (green), and 50 min (yellow). Black line shows fit to Equation (2).

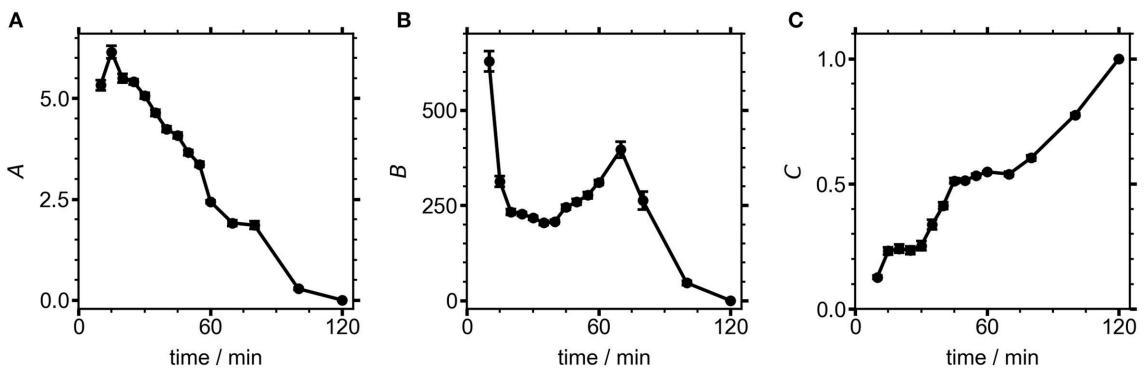

FIGURE 7 | Fit coefficients from Equation (2) for NCs measured at different growth times. Contributions from the **(A)** first and **(B)** second derivatives, as well as the **(C)** 120 min NC component to the overall fit.

TABLE 1 | Best-fit values for parameters *A*, *B*, and *C* with one standard deviation error of the fitted variables (σ_X).

Reaction timepoint (min)	$A \pm \sigma_A$	$B \pm \sigma_B$	$C \pm \sigma_C$
10	5.33 ± 0.12	628 ± 27	0.127 ± 0.007
15	6.15 ± 0.15	313 ± 14	0.232 ± 0.015
20	5.49 ± 0.11	232 ± 8	0.241 ± 0.017
25	5.41 ± 0.07	227 ± 5	0.235 ± 0.015
30	5.07 ± 0.07	217 ± 5	0.253 ± 0.018
35	4.64 ± 0.08	204 ± 5	0.337 ± 0.019
40	4.24 ± 0.07	208 ± 5	0.413 ± 0.013
45	4.08 ± 0.07	245 ± 6	0.513 ± 0.011
50	3.66 ± 0.07	259 ± 8	0.514 ± 0.009
55	3.36 ± 0.06	277 ± 8	0.533 ± 0.008
60	2.44 ± 0.04	310 ± 6	0.548 ± 0.004
70	1.91 ± 0.07	396 ± 21	0.539 ± 0.009
80	1.86 ± 0.09	262 ± 23	0.604 ± 0.011
100	0.281 ± 0.03	46.6 ± 5.6	0.776 ± 0.006
120	0	0	1

measurements indicate that poorly-capped NCs are dominant during the growth of perovskite NCs, becoming progressively better capped as the growth process continues, similar to the case shown in **Figure 1B**. Future studies using different polarities of filter media to separate well- and poorly-capped NCs (Sadighian et al., 2020) will seek to test this assumption and further isolate the evolving lineshapes of these sub-populations within the reaction mixture. As demonstrated here by the intriguing trends in the weights of the two derivative features, the ability of SSTA to measure lineshapes during a NC synthesis provides a new avenue to deeper insights into how NCs grow. Further analyses of both the lineshapes and the exciton dynamics hold promise for understanding the evolving nature of carrier traps in nascent NCs.

4. CONCLUSION

A novel, broadband, tilted-pulse SSTA spectrometer with a 60 ps time delay was used to investigate evolving excited state dynamics in NCs grown *via* a solvation-limited synthesis. Growing NCs were found to exhibit a unique TA lineshape indicative of the Stark effect. Fits of these data to a weighted sum of linear absorbance spectrum derivatives show that this lineshape is likely caused by spatially seperated charge carriers in surface trap states. This adds to the growing body of evidence that these NCs are poorly capped during most of their growth (Teunis et al., 2017; Sadighian et al., 2020). This work proves the applicability of this technique to the study of non-equilibrium systems such as growing NCs that were previously inaccessible with non-linear spectroscopies. The development of SSTA and this sampling technique provide powerful tools for understanding how the electronic structure and excited state dynamics of NCs change during their synthesis. These types of experiments may offer new insight into NC growth mechanisms and how reaction parameters can be changed to target desired photophysics.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

JS optimized and executed the synthesis. KW optimized and operated the instrument. JS, KW, and MC executed the experiments. JS analyzed the data. JS and CW designed the research. All authors contributed to manuscript revision.

FUNDING

This material is based upon work supported by the National Science Foundation under grant CHE-1752129. MC acknowledges financial support from ARCS Oregon Chapter.

ACKNOWLEDGMENTS

The authors acknowledge Martin D. Gruber and Madelyn N. Scott for assistance running experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2020.585853/full#supplementary-material

REFERENCES

Boles, M. A., Ling, D., Hyeon, T., and Talapin, D. V. (2016). The surface science of nanocrystals. *Nat. Mater.* 15, 141–153. doi: 10.1038/nmat4526

Bublitz, G. U., and Boxer, S. G. (1997). Stark spectroscopy: applications in chemistry, biology, and materials science. *Annu.Rev. Phys. Chem.* 48, 213–242.

Colvin, V. L., and Alivisatos, A. P. (1992). Cdse nanocrystals with a dipole moment in the first excited state. *J. Chem. Phys.* 97, 730–733. doi: 10.1063/1.463573

Colvin, V. L., Cunningham, K. L., and Alivisatos, A. P. (1994). Electric field modulation studies of optical absorption in CdSe nanocrystals: Dipolar character of the excited state. *J. Chem. Phys.* 101, 7122–7138. doi: 10.1063/1.468338

De Roo, J., Ibez, M., Geiregat, P., Nedelcu, G., Walravens, W., Maes, J., et al. (2016). Highly dynamic ligand binding and light absorption coefficient of cesium lead bromide perovskite nanocrystals. *ACS Nano* 10, 2071–2081. doi: 10.1021/acsnano.5b06295

Dirin, D. N., Protesescu, L., Trummer, D., Kochetygov, I. V., Yakunin, S., Krumeich, F., et al. (2016). Harnessing defect-tolerance at the nanoscale: highly luminescent lead halide perovskite nanocrystals in mesoporous silica matrixes. *Nano Lett.* 16, 5866–5874. doi: 10.1021/acs.nanolett.6b02688

Fourkas, J. T., Dhar, L., Nelson, K. A., and Trebino, R. (1995). Spatially encoded, single-shot ultrafast spectroscopies. *J. Opt. Soc. Am. B* 12, 155–165. doi: 10.1364/JOSAB.12.000155

Fu, Y., Meng, F., Rowley, M. B., Thompson, B. J., Shearer, M. J., Ma, D., et al. (2015). Solution growth of single crystal methylammonium lead halide perovskite nanostructures for optoelectronic and photovoltaic applications. *J. Am. Chem. Soc.* 137, 5810–5818. doi: 10.1021/jacs.5b02651

Guyot-Sionnest, P., Shim, M., Matranga, C., and Hines, M. (1999). Intraband relaxation in CdSe quantum dots. *Phys. Rev. B* 60, R2181–R2184. doi: 10.1103/PhysRevB.60.R2181

Guyot-Sionnest, P., Wehrenberg, B., and Yu, D. (2005). Intraband relaxation in CdSe nanocrystals and the strong influence of the surface ligands. *J. Chem. Phys.* 123:074709. doi: 10.1063/1.2004818

Hassan, Y., Ashton, O. J., Park, J. H., Li, G., Sakai, N., Wenger, B., et al. (2019). Facile synthesis of stable and highly luminescent methylammonium lead halide nanocrystals for efficient light emitting devices. *J. Am. Chem. Soc.* 141, 1269–1279. doi: 10.1021/jacs.8b09706

Hassan, Y., Song, Y., Pensack, R. D., Abdelrahman, A. I., Kobayashi, Y., Winnik, M. A., et al. (2016). Structure-tuned lead halide perovskite nanocrystals. *Adv. Mater.* 28, 566–573. doi: 10.1002/adma.201503461

Huang, H., Xue, Q., Chen, B., Xiong, Y., Schneider, J., Zhi, C., et al. (2017). Top-down fabrication of stable methylammonium lead halide perovskite nanocrystals by employing a mixture of ligands as coordinating solvents. *Angew. Chem. Int. Ed.* 56, 9571–9576. doi: 10.1002/anie.201705595

Katari, J. E. B., Colvin, V. L., and Alivisatos, A. P. (1994). X-ray photoelectron spectroscopy of CdSe nanocrystals with applications to studies of the nanocrystal surface. *J. Phys. Chem.* 98, 4109–4117. doi: 10.1021/j100066a034

Klimov, V. I. (2000). Optical nonlinearities and ultrafast carrier dynamics in semiconductor nanocrystals. *J. Phys. Chem. B* 104, 6112–6123. doi: 10.1021/jp9944132

Klimov, V. I., and McBranch, D. W. (1998). Femtosecond 1*P*-to- 1*S* electron relaxation in strongly confined semiconductor nanocrystals. *Phys. Rev. Lett.* 80, 4028–4031. doi: 10.1103/PhysRevLett.80.4028

LaMer, V. K., and Dinegar, R. H. (1950). Theory, production and mechanism of formation of monodispersed hydrosols. *J. Am. Chem. Soc.* 72, 4847–4854.

Li, Y., Lai, R., Luo, X., Liu, X., Ding, T., Lu, X., et al. (2019). On the absence of a phonon bottleneck in strongly confined $CsPbBr_3$ perovskite nanocrystals. *Chem. Sci.* 10, 5983–5989. doi: 10.1039/C9SC01339C

Makishima, Y., Furukawa, N., Ishida, A., and Takeda, J. (2006). Femtosecond real-time pump-probe imaging spectroscopy implemented on a single shot basis. *Jpn. J. Appl. Phys.* 45:5986. doi: 10.1143/JJAP.45.5986

Mattoussi, H., Cumming, A. W., Murray, C. B., Bawendi, M. G., and Ober, R. (1998). Properties of CdSe nanocrystal dispersions in the dilute regime: Structure and interparticle interactions. *Phys. Rev. B* 58, 7850–7863. doi: 10.1103/PhysRevB.58.7850

Mondal, N., and Samanta, A. (2017). Complete ultrafast charge carrier dynamics in photo-excited all-inorganic perovskite nanocrystals ($CsPbX_3$). *Nanoscale* 9, 1878–1885. doi: 10.1039/C6NR09422H

Pan, A., He, B., Fan, X., Liu, Z., Urban, J. J., Alivisatos, A. P., et al. (2016). Insight into the ligand-mediated synthesis of colloidal $CsPbBr_3$ perovskite nanocrystals: the role of organic acid, base, and cesium precursors. *ACS Nano* 10, 7943–7954. doi: 10.1021/acsnano.6b03863

Peterson, M. D., Cass, L. C., Harris, R. D., Edme, K., Sung, K., and Weiss, E. A. (2014). The role of ligands in determining the exciton relaxation dynamics in semiconductor quantum dots. *Annu. Rev. Phys. Chem.* 65, 317–339. doi: 10.1146/annurev-physchem-040513-103649

Protesescu, L., Yakunin, S., Bodnarchuk, M. I., Krieg, F., Caputo, R., Hendon, C. H., et al. (2015). Nanocrystals of cesium lead halide perovskites ($CsPbX_3$, X = Cl, Br, and I): novel optoelectronic materials showing bright emission with wide color Gamut. *Nano Lett.* 15, 3692–3696. doi: 10.1021/nl5048779

Queloz, V. I. E., Bouduban, M. E. F., Garca-Benito, I., Fedorovskiy, A., Orlandi, S., Cavazzini, M., et al. (2020). Spatial charge separation as the origin of anomalous stark effect in fluorous 2d hybrid perovskites. *Adv. Funct. Mater.* 30:2000228. doi: 10.1002/adfm.202000228

Righetto, M., Lim, S. S., Giovanni, D., Lim, J. W. M., Zhang, Q., Ramesh, S., et al. (2020). Hot carriers perspective on the nature of traps in perovskites. *Nat. Commun.* 11:2712. doi: 10.1038/s41467-020-16463-7

Sadighian, J. C., Crawford, M. L., Suder, T. W., and Wong, C. Y. (2020). Surface ligation stage revealed through polarity-dependent fluorescence during perovskite nanocrystal growth. *J. Mater. Chem. C* 8, 7041–7050. doi: 10.1039/C9TC06545H

Sadighian, J. C., Crawford, M. L., and Wong, C. Y. (2019). Rapid sampling during synthesis of lead halide perovskite nanocrystals for spectroscopic measurement. *MRS Adv.* 4, 1957–1964. doi: 10.1557/adv.2019.263

Sharma, D. K., Hirata, S., Biju, V., and Vacha, M. (2019). Stark effect and environment-induced modulation of emission in single halide perovskite nanocrystals. *ACS Nano* 13, 624–632. doi: 10.1021/acsnano.8b07677

Sun, S., Yuan, D., Xu, Y., Wang, A., and Deng, Z. (2016). Ligand-mediated synthesis of shape-controlled cesium lead halide perovskite nanocrystals via reprecipitation process at room temperature. *ACS Nano* 10, 3648–3657. doi: 10.1021/acsnano.5b08193

Tanaka, K., and Kondo, T. (2003). Bandgap and exciton binding energies in lead-iodide-based natural quantum-well crystals. *Sci. Technol. Adv. Mater.* 4, 599–604. doi: 10.1016/j.stam.2003.09.019

Teunis, M. B., Liyanage, T., Dolai, S., Muhoberac, B. B., Sardar, R., and Agarwal, M. (2017). Unraveling the mechanism underlying surface ligand passivation of colloidal semiconductor nanocrystals: a route for preparing advanced hybrid nanomaterials. *Chem. Mater.* 29, 8838–8849. doi: 10.1021/acs.chemmater.7b03240

Urayama, J., Norris, T. B., Singh, J., and Bhattacharya, P. (2001). Observation of phonon bottleneck in quantum dot electronic relaxation. *Phys. Rev. Lett.* 86, 4930–4933. doi: 10.1103/PhysRevLett.86.4930

Walters, G., Wei, M., Voznyy, O., Quintero-Bermudez, R., Kiani, A., Smilgies, D.-M., et al. (2018). The quantum-confined Stark effect in layered hybrid perovskites mediated by orientational polarizability of confined dipoles. *Nat. Commun.* 9:4214. doi: 10.1038/s41467-018-06746-5

Wang, L., Williams, N. E., Malachosky, E. W., Otto, J. P., Hayes, D., Wood, R. E., et al. (2017). Scalable ligand-mediated transport synthesis of organic inorganic hybrid perovskite nanocrystals with resolved electronic structure and ultrafast dynamics. *ACS Nano* 11, 2689–2696. doi: 10.1021/acsnano.6b07574

Wang, Z. L., Harfenist, S. A., Whetten, R. L., Bentley, J., and Evans, N. D. (1998). Bundling and interdigitation of adsorbed thiolate groups in self-assembled nanocrystal superlattices. *J. Phys. Chem. B* 102, 3068–3072.

Wilson, K. S., Mapile, A. N., Wong, C. Y., Wong, C. Y., and Wong, C. Y. (2020). Broadband single-shot transient absorption spectroscopy. *Opt. Express* 28, 11339–11355. doi: 10.1364/OE.390938

Wilson, K. S., Scott, M. N., and Wong, C. Y. (2019). Excited state dynamics of organic semiconductors measured with shot-to-shot correction of scatter and photoluminescence. *Synth. Metals* 250, 115–120. doi: 10.1016/j.synthmet.2019.03.007

Wilson, K. S., and Wong, C. Y. (2018). Single-shot transient absorption spectroscopy with a 45 ps pump-probe time delay range. *Opt. Lett.* 43, 371–374. doi: 10.1364/OL.43.000371

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Physics

ORIGINAL RESEARCH
published: 20 October 2020
doi: 10.3389/fphy.2020.583173

Quasi-Stationary States in Ionic Liquid-Liquid Crystal Mixtures at the Nematic-Isotropic Phase Transition

Alokmay Datta[1]*, *Kenichi Yoshikawa*[2,3], *Yukihiro Yoshida*[4] and *Gunzi Saito*[4]

[1] XRD and SEM Units, Materials Characterization and Instrumentation Division, Council for Scientific and Industrial Research-Central Glass and Ceramic Research Institute, Kolkata, India, [2] Laboratory of Life Physics, Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe, Japan, [3] Institute for Advanced Study, Kyoto University, Kyoto, Japan, [4] Department of Chemistry, Kyoto University, Kyoto, Japan

An open system is a system driven away from equilibrium by a source that supplies an inflow of energy and a sink to maintain an outflow. A typical example of an open system is a system close to its phase transition temperature under irradiation by a laser. This provides a steady flow of energy through a photon flux. The sink in that case is the environment to which energy is lost in the form of heat dissipation. Creation of such a thermodynamically open state suggests that we can expect generation of exotic spatio-temporal structures length-scale independent correlation maintained under the global dissipative forces provided by the surroundings. Internal long-range forces can bring in additional spatio-temporal correlations, giving rise to states with a very long lifetime, the "quasistationary states" (QSS). In this communication, we report evolution of quasistationary states, in a mixture of the well-known liquid crystal (*N*-(4-methoxybenzylidene)-4-butylaniline, MBBA) and an iron-based room temperature ionic liquid (RTIL), namely, 1-ethyl-3-methylimidazolium tetrachloroferrate (EMIF) at the *Nematic-Isotropic* phase transition, when focused radiation with 532 nm wavelength from a Nd:YAG laser (200–300 mW optical power) is incident on the sample. We explain the QSS by invoking a sharp negative thermal gradient due to the laser photon flux and dipolar interactions. In our model, the dipoles are the charge transfer complexes (CTCs) formed in the RTIL by resonant laser pumping, which create an orientational ordering and balance the fluctuating force of the thermal gradient to create the QSS. In the absence of such CTCs in a mixture of MBBA and a Gallium-based RTIL (1-ethyl-3-methylimidazolium tetrachlorogallate, EMIG), the QSS was not observed.

Keywords: open system, non-equilibrium thermodynamics, room temperature ionic liquid (RTIL), liquid crystal (LC), nematic-isotropic interface, optical pumping, quasi-stationary state

OPEN ACCESS

Edited by:
Rene A. Nome,
State University of Campinas, Brazil

Reviewed by:
Arkaprabha Konar,
Kent State University, United States
Naved I. Malek,
Sardar Vallabhbhai National Institute of Technology Surat, India

***Correspondence:**
Alokmay Datta
fellow1@cgcri.res.in;
alokmaydatta@gmail.com

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Physics

Received: *14 July 2020*
Accepted: *31 August 2020*
Published: *20 October 2020*

Citation:
Datta A, Yoshikawa K, Yoshida Y and Saito G (2020) Quasi-Stationary States in Ionic Liquid-Liquid Crystal Mixtures at the Nematic-Isotropic Phase Transition. Front. Phys. 8:583173. doi: 10.3389/fphy.2020.583173

INTRODUCTION

When in a phase transition the system transforms from one equilibrium structure to another with a change in symmetry, there is exchange of a fixed amount of energy with the surrounding at a fixed temperature. Both the phases co-exist during this conversion. These are the hallmarks of a first order transition following Arrhenius law of linear dependence of logarithm of the rate of energy transfer on the inverse of the temperature. The slope of this line is the fixed enthalpy of transition or the activation barrier of the transition.

In complex systems, there are departures from these characteristics. An example of this is the Nematic-Isotropic (*NI*) transition in liquid crystals (LCs), classified as "weakly first order" by de Gennes [1]. Here, while there is a definite change in symmetry in transition from the orientationally ordered nematic to the totally disordered isotropic phase, there is no finite co-existence region of these phases.

This transition in the well-known LC material *N*-(4-methoxybenzylidene)-4-butylaniline (MBBA) does not follow the Arrhenius law [2]. More interestingly, it is not strictly an equilibrium phase transition [3] in that both the transition enthalpy and temperature are dependent on the rate of energy supply. However, mixing MBBA with polar molecules [4], and in particular with an iron-based room temperature ionic liquid (RTIL), 1-ethyl-3-methylimidazolium tetrachloroferrate ($[EMI]^+[FeCl4]^-$, EMIF), converts this *NI* transition to a first order equilibrium transition between an Isotropic phase and a phase where the conformational entropy of the MBBA molecule is drastically reduced [5]. These studies were carried out on systems through near-equilibrium processes. Here the interaction between MBBA and EMIF molecules is carried out at the molecular, i.e., sub nm length scale and a time scale covering 100 ps−10 ns, while the orientational ordering giving rise to the phase transition takes place at a macroscopic length scale covering the bulk (mm to cm) and over at least 100 s. Hence, the complex nature of the transition, working at these two spatiotemporal hierarchies, is apparent.

In thermodynamically open systems, on the other hand, an external source supplies an inflow and an external sink provides an outflow of energy and/or material to the system [6]. This "sink" may be the surrounding or the environment itself and these inflow-outflow processes are, in general, irreversible. An open system is thus in a non-equilibrium state, characterized by an inhomogeneous distribution of the variables and non-vanishing values of the fluxes. The inhomogeneities and fluxes give rise to structures with spatiotemporal hierarchies, which are not observed in equilibrium systems. These *dissipative structures* are dependent on time by definition, implicitly if not explicitly, and they can be changed by changing the inflow-outflow rates [7, 8]. They are also coupled by a length-scale-independent correlation imposed by the fluxes [9]. Internal long-range forces can bring in additional spatiotemporal correlations to such open systems, giving rise to states with a very long lifetime, the so-called "quasistationary states" (QSS) [10–12]. Quasistationary states involving spin chains [13], lattices with infinity of absorbing states [14], and in hydrodynamics on a torus [15] have proved the general character of these states.

Treating the EMIF-MBBA system at around the *NI* transition under open, non-equilibrium conditions is expected to generate dissipative structures that will highlight this spatiotemporal hierarchy through their dynamics. RTILs are composed of bulky cations (such as imidazolium) and considerably large anions and remain liquid and ionized at ambient temperatures due to the large amount of configurational entropy contained in these ions, as well as the reduced inter-ionic Coulomb forces owing to the charge delocalization within the ions [16]. Their large dipole moments can couple them to LC molecules and transfer this large entropy to the LC system, affecting its phase transitions, viz., taking the nematic-isotropic transition to lower temperatures. RTILs also have the attractive physical properties of having extremely low vapor pressure, thermal stability and a large liquid window (∼200°C). LC systems, due to their orientational order, are amenable to studies with non-invasive polarization microscopy and the dipolar coupling between RTILs such as EMIF and EMIG on one hand, and LCs like MBBA on the other, is expected to have a strong effect on the polarization of the latter and cause a commensurate variation in the microscopic images.

A radial pattern of polarization has been observed in EMIF on irradiation by a focused laser [17]. On switching the laser off, the pattern showed a fast (<1 s) change followed by a quasistationary structure for more than 100 s. That this dissipative structure is indeed caused by a resonant pumping by the laser is confirmed from the fact that it is not observed with 1-ethyl-3-methylimidazolium tetrachlorogallate ($[EMI]^+[GaCl4]^-$, EMIG), a Ga-based RTIL with the same laser power.

Thus, we see that an EMIF-MBBA mixture, resonantly pumped by a focused laser may give rise to an open system, where the coupling between the EMIF and MBBA molecules can compete with the thermal dissipation provided by laser heating. Creation of such a thermodynamically open state suggests that we can expect generation of exotic spatiotemporal structures different from the static structures under equilibrium. In particular, the orientational ordering imposed on MBBA by EMIF, as evident from ref 5, can give rise to a QSS consisting the nematic and isotropic phases in co-existence. These considerations and the expected strong variation of images in polarization microscopy due to the change in orientation of MBBA as evidence of presence or otherwise of RTIL-LC coupling motivated us to carry out these studies. To the best of our knowledge, such studies have not been carried out with other LC and RTIL mixtures.

In this communication, we report evolution of QSS under the above conditions. We explain the QSS by invoking a sharp negative thermal gradient due to the laser photon flux and dipolar interactions. In our model, the dipoles are formed in the RTIL by resonant laser pumping, which create an orientational ordering and balance the fluctuating force of the thermal gradient to create the QSS. This allows probing of the transition through polarization microscopy. Hence, we have carried out polarization microscopy experiments to arrive at our results.

EQUILIBRIUM PROPERTIES OF THE MIXTURES

In all cases, the sample is a dip-coated film on a glass slide. As samples, we used mixtures of MBBA and EMIF, and MBBA and EMIG. The structures of MBBA, EMIF, and EMIG are shown in **Figures 1A–C**, respectively. MBBA (99% pure) was purchased from Nacalai Tesque, Japan and used without further purification. EMIF and EMIG were synthesized and purified in the lab [18]. The LC was found to be strongly hydrophobic.

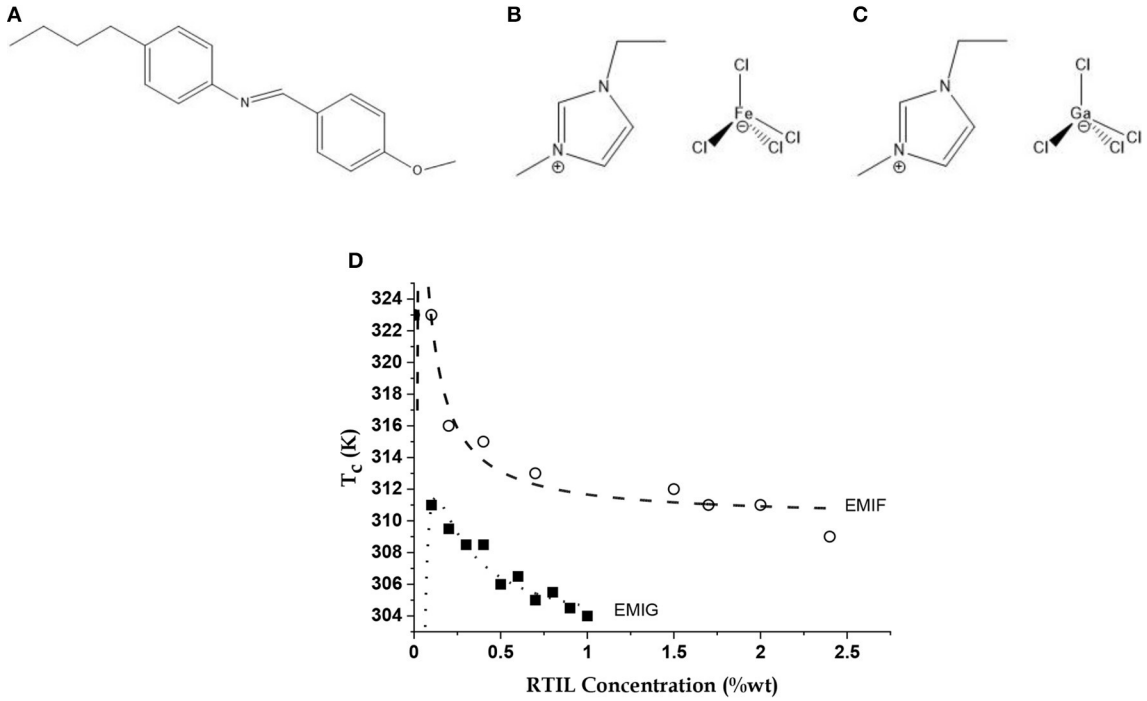

FIGURE 1 | The molecules and their equilibrium property. Molecular structures of the liquid crystal (LC) **(A)** *n*-(4-methoxybenzylidene)-4-butylaniline (MBBA), and the room temperature ionic liquids (RTILs) **(B)** 1-ethyl-3-methylimidazolium tetrachloroferrate (EMIF) and **(C)** 1-ethyl-3-methylimidazolium tetrachlorogallate (EMIG); **(D)** Variations of Nematic-Isotropic transition temperature (T_c) of LC-RTIL mixtures with EMIF and EMIG concentrations shown as open circles and filled squares, respectively. Data simulated (dashed and dotted lines, respectively) with a model of the MBBA dipoles screened by RTIL ions taking into account the self-screening of the ions [5] using Equation (1). See text for details.

The LC and the RTILs were mixed to vary the concentration of EMIF at different values from 0.00 and 2.42 wt%. The mixture became turbid on keeping and had to be periodically stirred and heated with warm air to regain its transparency. The NI transition temperature (T_c) was determined for each of these mixtures by varying temperature between 298 and 328 K in 1 K steps using a hotplate stage (accuracy 0.1 K). The phases were detected by a polarization microscope (IX70, Olympus, Japan) equipped with a 20× objective lens and a CCD camera (640 × 480 pixels). The data is the average over three measurements in each case with insignificant errors.

The EMIF-MBBA and EMIG-MBBA mixtures showed lowering of *N-I* transition temperature (T_c) with *C*, the weight percentage of the RTIL in the mixture, until the components segregated (**Figure 1D**). The data are calculated from a model of interacting MBBA dipoles screened by the free charges of the RTILs, accounting also for the self-screening of the charges. Here we assume that the ions of EMIF are active in both screening the MBBA molecular quadrupoles and in self-screening while other forces (such as, dispersion forces) play no role in the transition. Thus, there is an effective charge q located at each MBBA molecule screened by the counterions of EMIF surrounding charges on the molecular quadrupole. If it is further assumed that addition of a small amount of ionic liquid only changes the value of q keeping the intermolecular separations, unchanged, then increase in C will lead to more screening and hence to a lowering of q ($q = k_1/C$). However, this process will generate its own self-screening and the final relation between C and q, with all other factors such as intermolecular separation being lumped into two constant terms k1 and k2, is given by $C = k_1/q\,(1 - k_2/q)$.

The dependence of T_c on *C*, the wt% concentration of RTIL, is then given by this model as

$$T_c = T_c^0 + \frac{A_T}{C^2} - \frac{B_T}{C} \quad (1)$$

where T_c^0 is the transition temperature of pristine MBBA (323 K), and A_T and B_T are the self-screening and screening parameters, respectively [5]. The values of A_T and B_T for EMIF come out to be −0.03 and −1.5 K, respectively, and for EMIG they are −0.14 and −2.25 K, respectively. Thus, EMIG has a considerably stronger screening and self-screening effect, possibly due to the smaller size of Ga and thus the tetrachlorogallate anion.

THE OPEN SYSTEMS AND THEIR DYNAMICS

We have investigated the microscopic *N-I* transition induced by a continuous wave (cw) at 532 nm, 200 mW optical power focused

(spot diameter $\sim$ 1 μm) on samples taken in a 3 cm $\times$ 3 cm $\times$ 25 μm cell. The linearly polarized laser was assembled from a randomly polarized laser source (JOL-D8PK-Y, JENOPTIK, Germany), and a Ti: Sapphire polarization beam splitter. The laser was made incident on the sample by a dichroic mirror. We have performed polarization microscopy (IX70, Olympus, Japan) to study the *NI* transition induced by the laser. We recorded the microscopic phase transitions in EMIF-(EMIG-) MBBA through a CCD camera (30 frames/s, 640 $\times$ 480 pixels) attached to the eyepiece of the polarization microscope. We took the terminal value of C as the maximum achievable value [C_{max}, $\sim$ 2.42 ($\sim$1.32) wt% for EMIF- (EMIG-) MBBA, from **Figure 1D**] and carried out the laser-induced transition on these "C_{max}" mixtures at 33°C, close to but lower than the corresponding T_c.

Upon irradiation, both mixtures generate an isochromic (or black) droplet of diameter $\sim$ 5 μm, centered about the laser beam, of the *I*-phase to form on the nematic bulk. These are shown for 200 mW of laser power in **Supplementary Movies 1**, **2**, respectively. **Supplementary Movie 1** has been truncated in the final portion after the isotropic phase has been achieved, i.e., without any loss of important information, due to the large filesize. We have detected the macroscopically defined orientation of the *N*-phase by rotating the sample in plane.

We extracted the spatiotemporal maps of the entire events from M1 and M2 using the freeware ImageJ (National Institute of Health, USA) and presented them in the respective **Figures 2A,B**, where we also show typical frames corresponding to the main stages of the evolution at the bottom of maps. We indicate the macroscopic orientation of the nematic phase by an arrow in the relevant frames. In **Figure 2C**, we present the evolution of MBBA only, under irradiation by the same laser power. As shown, we could observe no transition to the isotropic phase (isochromic or black droplet) other than an increase in the local orientational disorder. These coarsen into increasingly bigger regions shown in **Figure 2C**ii and finally cover the field of view, with few isolated domains. This occurred even on increasing optical power to 500 mW without RTIL or with the mixtures at all proportions but with the laser radiation at 1,064 nm.

In EMIF-MBBA, the isotropic droplet first appears with a clean and sharp boundary (region 1, **Figure 2A**). However, within $\sim$100 ms, a number of clusters of different colors (but with no color variation within each cluster), on the average going from red to green outwards, appear in the region located between the isotropic droplet and the nematic bulk (region 2, **Figure 2A**), producing an "opal-like" texture [19]. Since a particular color corresponds to a particular polarization it is clear that these clusters are individual *domains* of the nematic or, better, "nematic-like" phase with the size $\sim$ 500 nm whose directors are randomly oriented and M1 shows that, unlike natural or synthetic opals [20], these clusters are in incessant motion. The major types of motion involved are (i) convective motion of domains, and (ii) predominantly two-body coalescence of domains ending in merging of these coalesced domains into the nematic bulk. From region 2 of **Figure 2A** we see that throughout these motions, for $\sim$50 s, the diameter of the isotropic droplet remains near 5 μm and the width of the interfacial zone that contains the nematic domains remains at $\sim$1.5 μm, which indicates the presence of a QSS.

With EMIG-MBBA, the above interfacial region with its internal dynamics is entirely absent (**Figure 2B**). Though the system shows slow dynamics at 200 mW, the speed increases monotonically with laser power, and the in data at 500 mW presented in **Figure 2B**, there is no trace of a QSS. This contrasts sharply with the Fe-based RTIL system where increase in laser power within this range does not affect the QSS formation and stability, nor are there any other changes in the dynamics with laser power from 200 to 500 mW, and similarly the change in the film thickness of the sample made no noticeable difference in behavior. This is consistent with the long range correlations in QSS [10–12].

In the next stage of EMIF-MBBA *N-I* transition dynamics (region 3, **Figure 2A**), roughly triangular areas with their bases on the nematic-isotropic boundary open up in the nematic bulk. These patches consist of randomly oriented, nanometer-sized domains similar to those in the interfacial zone. The former domains move inwards while the latter move outwards, whereby they meet and coalesce. The coalesced domains, unlike in the second stage, now shrink and disappear into the isotropic phase. Thus, this process disrupts the quasistationary structure and finally the isotropic region expands to cover the whole field of view, as can be seen in region 4 of the spatiotemporal image. The last stage, region 5, is after switching the laser off, when the reverse transition to the nematic phase takes place without any intermediate QSS but only through coalescence of large nematic domains.

We have investigated the kinematics of the QSS at two levels: for the overall isotropic-interfacial region system and for the clusters or domains within the interfacial region. To quantify the overall kinematics, we have presented the log-log plots of the diameter (D) of isotropic droplet (open circles) and width (w) of the interfacial region (filled circles), vs. elapsed time (t), in **Figure 3**. Due to the inherent large dynamic range of the data, this average over 10 measurements has to be presented in a log-log scale and error bars are less meaningful. In both plots, the QSS (region 2) and the next state (region 3) in system dynamics show excellent linear fits, which gave reliable values of p and q in the relations: $D \sim t^p$ and $w \sim t^q$. For the QSS, $p = 0.31 \pm 0.04$, $q = -0.32 \pm 0.07$, showing the characteristic slow dynamics in both the isotropic phase and the interfacial region in this stage, whereas for the next state $p = 2.16 \pm 0.14$, $q = 4.54 \pm 0.55$. Hence, expansion of the interfacial region clearly dominates the expansion in this dynamic state. After 76 s, the interfacial region suddenly vanishes and the isotropic phase covers the field of view in 84 s. Therefore, there is an initial fast growth phase of the interfacial zone covering the first 5 s after the laser is switched on, followed by the QSS that stabilizes this zone for the next 70 s, and finally the 8–9 s of fast decay of the interfacial zone.

The interfacial zone in QSS has its own dynamics. We find that there are two co-existing groups of domains separated by size, "small" ($D < 500$ nm) and "big" ($D > 500$ nm), and they have quite distinct dynamics. We show these two types of domains in the magnified image of the interfacial region in **Figures 4A,B**, respectively. The small domains have uniform

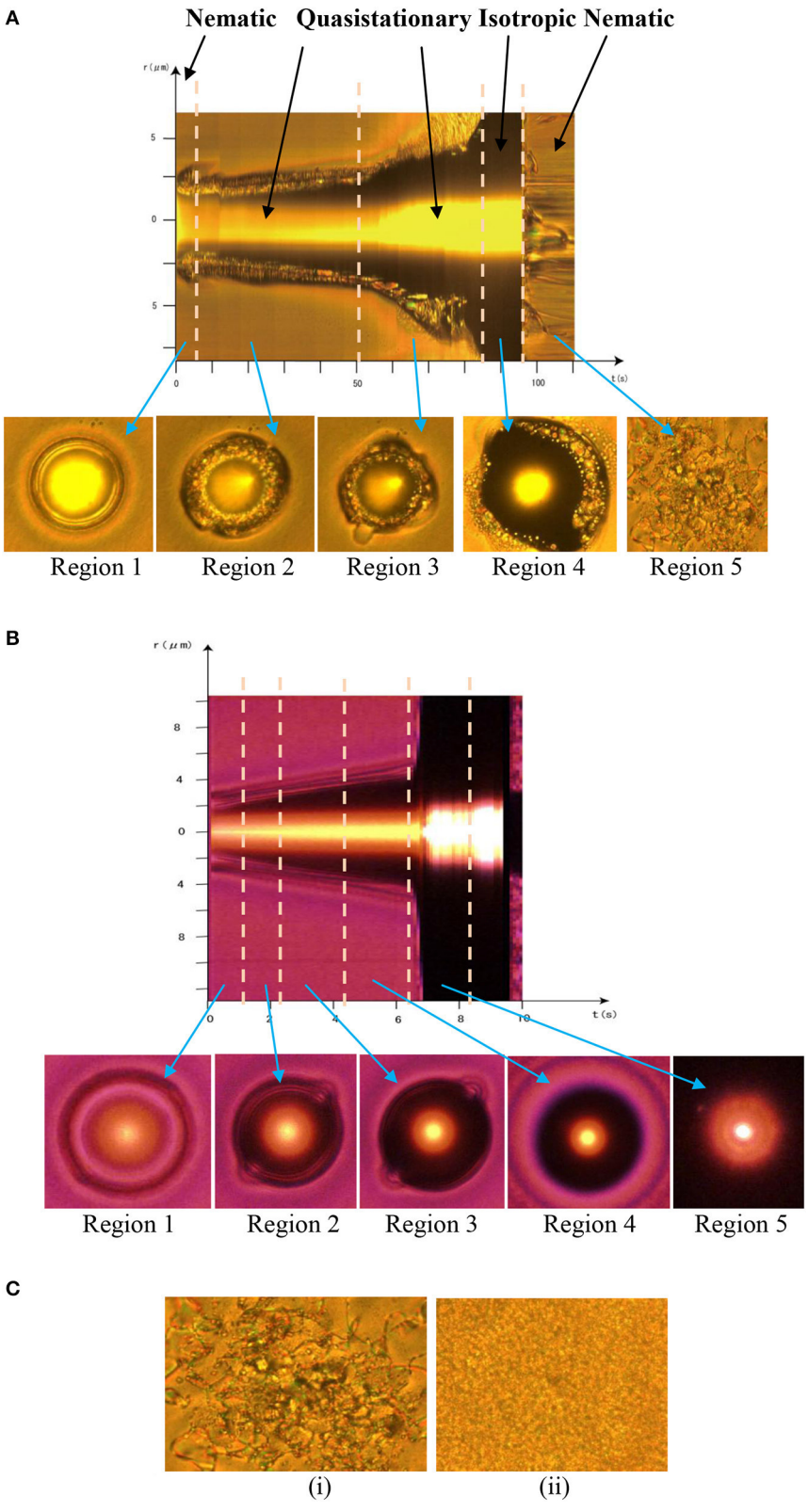

FIGURE 2 | Spatiotemporal evolution of the mixtures. Spatiotemporal maps of **(A)** EMIF-MBBA (with phases marked) and **(B)** EMIG-MBBA mixtures (C_{max} concentration) over the entire evolution at the corresponding T_c, extracted from **Supplementary Movies 1, 2**, respectively. The different regions in each map are separated by dashed lines and the typical frame of each region is shown below the map, designated by arrows. **(C)** Frames at (i) 100 μs and (ii) 300 μs of irradiation of MBBA only, by the laser. All data were taken at 200 mW of laser power.

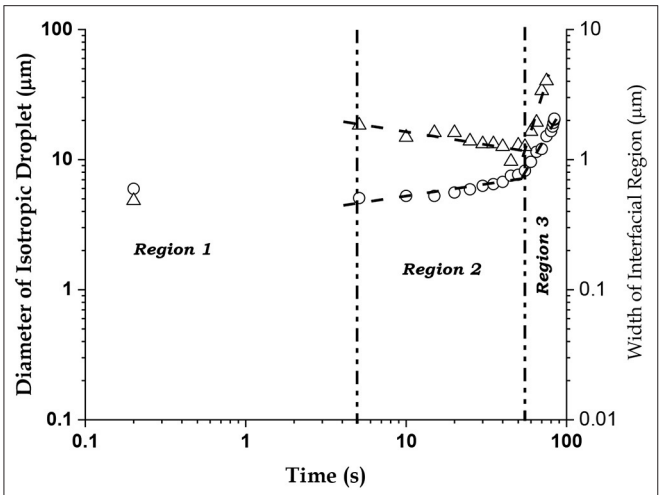

FIGURE 3 | The Quasistationary State (QSS). Log-log plots of the diameter (D, in μm, open circles) of the isotropic phase (dark droplet at the center of the laser beam) and the width (w, in μm, open triangles) of the interfacial region containing nematic clusters, vs. time (in s). The QSS occurs in Region 2.

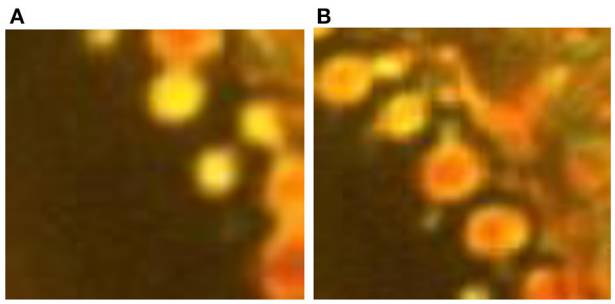

FIGURE 4 | Domains in the interfacial zone. The interfacial zone consists of **(A)** "small" ($D < 500$ nm) and **(B)** "big" ($D > 500$ nm) domains.

yellow color that does not change noticeably with time and they show some movement, while big domains show non-uniformity of color and change color from red to yellow to green (sometimes almost black) randomly but stay almost stationary. Thus, small domains change position but not orientation while big domains do not change position but change orientation randomly. What is perhaps more important is that big domains very seldom acquire the color, and hence the molecular orientation, of the small domains. The incident of two adjacent large domains acquiring the same color is also rare. These observations suggest that the small domains may contain a higher fraction of the EMIF that holds the LCs to a state with fixed orientation uniformly within the domain, against thermal fluctuations caused by the laser. Similarly, the big domains may have lower fractions of the RTIL. The LCs within them are fluctuating in orientation randomly with the heat and seldom reaching the highly oriented state of the small domains, or stable enough to induce the orientation in adjacent large domains.

We present log-log graphs of growth in domain size vs. time of two randomly chosen members from each group in **Figures 5A,B**, respectively, the data being shown by open and filled circles for the two members. From linear fits, in solid and dashed lines for the two members in each group, the growth exponents of the small clusters come out to be 1.71 ± 0.26 and 0.74 ± 0.24, whereas those of the big clusters are 0.11 ± 0.09 and 0.20 ± 0.09. Thus, though both groups show fluctuations in their growth dynamics the small clusters exhibit a general growth trend by coalescence in contrast to the big clusters whose size fluctuates about a mean value. This becomes clearer from **Figures 5C,D**, showing the size distributions of the clusters. The interesting thing, however, is that the primary maxima of both members in each group are very close: 275 and 285 nm for the small clusters, and 562 and 565 nm for the big clusters. This shows the quasistationary nature of dynamics even at the scale of the clusters or domains.

ORIGIN OF THE QSS: A TENTATIVE EXPLANATION

From the equilibrium and non-equilibrium properties of the mixtures described in sections Equilibrium Properties of the Mixtures and The Open Systems and Their Dynamics, respectively, we find that (1) EMIG has a stronger effect on the *NI* transition than EMIF under equilibrium conditions while (2) it has essentially no effect on the transition under non-equilibrium conditions. This sharp difference between the two conditions is based on the way the RTIL molecules interact with the MBBA in absence and presence of the laser.

In the absence of the laser radiation, RTILs are characterized by free or uncoordinated ions. In an EMIF-MBBA mixture these free charges have been shown to screen the MBBA molecular dipoles, with some degree of self-screening among these charges [5]. This interaction also converts the *NI* transition from a non-Arrhenius to an Arrhenius transition by suppressing the in-plane and out-of-plane motions of the benzene rings of MBBA. This freezes the *intra*-molecular motion of the molecular axis which converts the entropic *NI* activation barrier to an enthalpic or constant *inter*-molecular barrier. As **Figure 1D** shows, the same model can very well explain the EMIG-MBBA interaction. Indeed, this interaction is stronger in this case.

With pure EMIF, there is a vibronic, dipole forbidden d–d transition from the $^6A_1$ ground state to the $^4A_{1,}^4E$ (degenerate) excited state of Fe^{3+} in the $FeCl_4^-$ molecular anion, at around 536 nm (**Figure 6**) [21]. It is also seen in this figure, that while the spectrum of MBBA has no band at or near that wavelength, this transition in the EMIF is quenched in the EMIF+MBBA mixture, indicating a strong coupling between the molecules in which this state plays the most important role. While the ground state of the $FeCl_4^-$ molecular anion has a tetrahedral structure [22], the excited state has a doubly degenerate vibration mode and is most probably a square planar structure [23]. We propose that, since the imidazolium cation is also a predominantly planar structure, the coordination between the counterions is much enhanced in this state leading to the formation of EMIF dipoles.

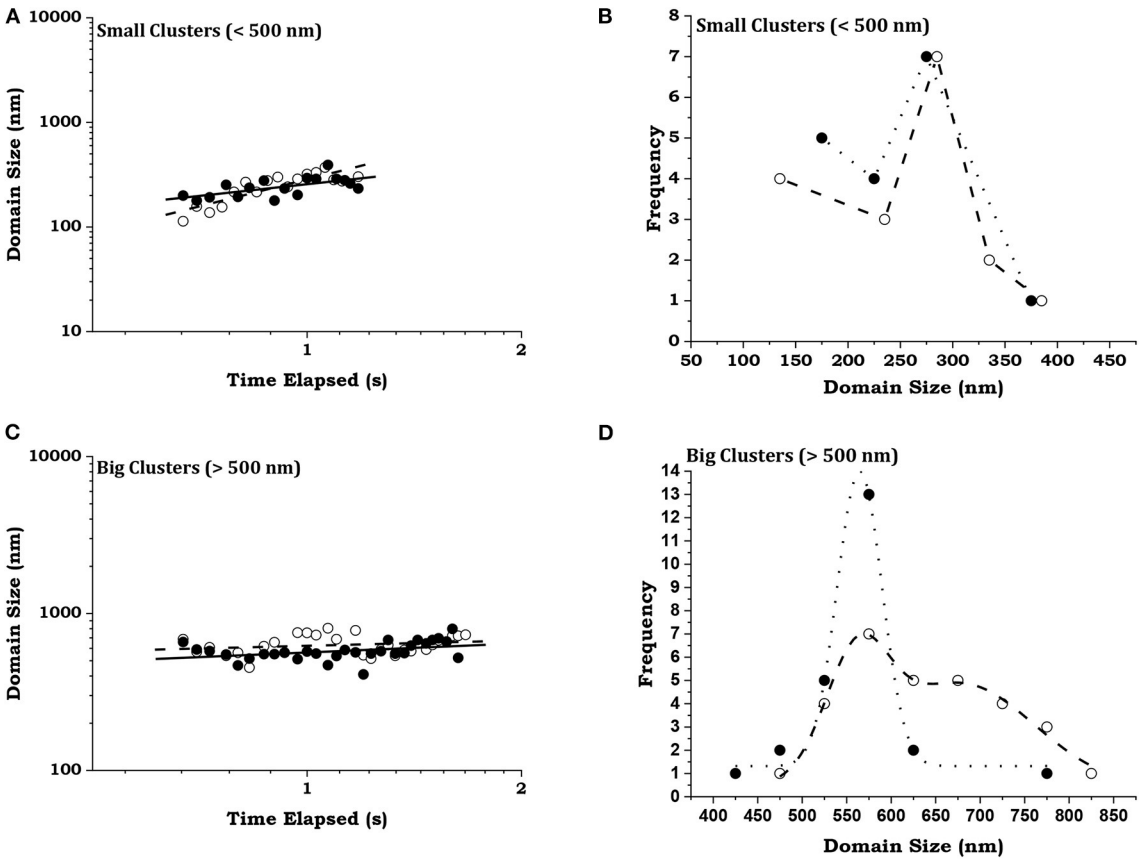

FIGURE 5 | Distribution of domains. Two randomly chosen members are represented by open and filled circles from **(A)**, **(C)** small (diameter < 500 nm) and **(B)**, **(D)** big (diameter > 500 nm) domains. The variations of domain size with time are shown as log-log plots in **(A,B)**, with dotted and dashed lines as fits to the data. The distributions of the domain sizes are shown in **(C,D)** with Gaussian simulations of the data.

Due to the weakness of the transition, strong resonant optical pumping by the laser is required to produce a sufficient number of such dipoles to form long range orientational correlation, since the dipole-dipole interaction is orientation dependent, unlike the screened Coulomb interaction obtaining in absence of the laser radiation. On the other hand, the high intensity of the laser also produces a thermal gradient across the laser intensity profile, falling off from the center outwards, which causes a stochastic, dissipative force to disrupt this orientational order. The continuous competition between these ordering and disordering forces stabilizes EMIF into a QSS [17]. The dissipative structure produced lasts after the laser is switched off as the laser induced heating dies down. Pure EMIG cannot form such a QSS since it does not have such a transition in this spectral region

We then explain the dynamics of the EMIF-MBBA open system by proposing that in presence of the laser, the EMIF-MBBA interaction is a dipole-dipole attraction, where the EMIF dipole, generated by the laser through this vibronic transition in a time scale of about 100 ps, interacts with the MBBA molecular dipole. This orientation dependent force creates a long range orientational correlation that competes with the thermal dissipative force and takes the correlated structure to a time scale of seconds. The latter drives the nematic phase to the isotropic phase while the former has a reverse drive and the competition between them stabilizes the interfacial zone into a quasistationary state, creating a "*dynamic* coexistence region" of the two phases. Again, as EMIG does not have this band resonant with the laser this QSS is not produced.

Here we put forward a qualitative model to explain the formation of the QSS. The major assumptions in our model are: (1) Since RTILs have mostly un-coordinated ions, they behave predominantly as collection of unscreened charges; (2) Optical pumping by the laser at 532 nm combines these charges to form dipoles resonantly for EMIF and not for EMIG. Let dipoles of magnitude μ_R be generated by the laser uniformly within the sample at a steady rate and stochastically, i.e., the dipole moment is either μ_R or 0. The number of these dipoles remains constant within the laser beam diameter of $d_0 = 2r_0$ ($d_0 = 1\mu$m) and the orientation of the dipoles, given by angle θ of the dipole with the nematic or optic axis, remains completely random within this diameter, due to the large thermal fluctuation set up by the laser.

For the EMIF-MBBA mixture, we assume the number of dipoles to fall off exponentially outside the direct influence of the laser beam ($r \geq r_0$, where r is the distance from the beam center) while they become exponentially oriented to the optic axis. We

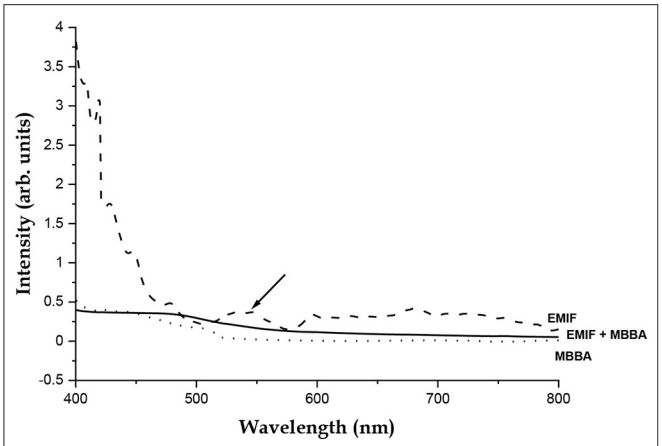

FIGURE 6 | Ultraviolet-Visible Spectrum of EMIF (dashed line), MBBA (dotted line) and the MBBA-EMIF mixture used (solid line). The $^6A_1 \rightarrow {}^4A_1^4$E transition band at 536.68 nm is marked with an arrow. This transition is seen to be quenched in the mixture.

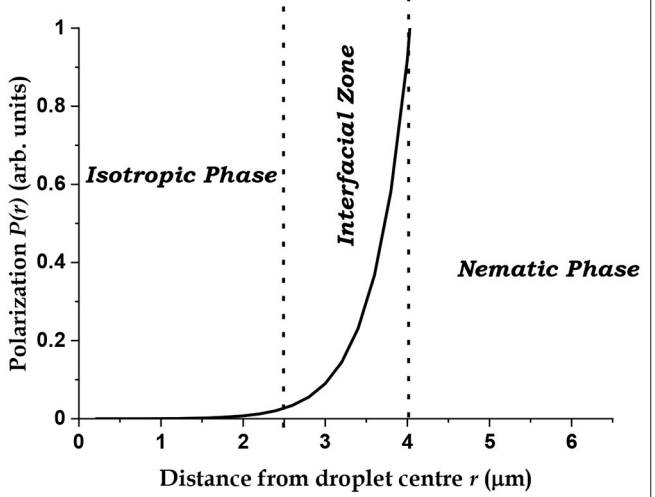

FIGURE 7 | Calculated average polarization profile of the EMIF-MBBA open system based on Equation (3).

base this last assumption on the presence of the orientating force of the molecules in the nematic bulk. We also assume that the orientation of the EMIF dipoles decides the orientation of the LC dipoles and thereby the polarization of the sample. Thus, our model points to a balance between the long-range dipolar orientating force and the thermal fluctuating force, since both are produced by the same source, i.e., the laser. Following our previous discussion, we propose that this balance is vital for the formation of the quasi-stationary state.

This model can be mapped to the Hamiltonian mean-field (HMF) model of globally coupled rotators [11] with the important modification that N, the number of coupled rotators is now a function of r, the distance from the laser center. Here we present a very simple and phenomenological version of the model to explain the profile of the polarization $P(r)$ across the phase boundary, assuming that the laser produces dipoles with a time-averaged number density $<n(r)>$ from the RTIL molecules, where each molecule has the dipole moment of magnitude μ_R and time-averaged orientation with the optic axis given by $<cos\theta(r)>$.

For the LC molecule, the magnitude of the dipole moment is μ_L and orientation is given by the dipoles of RTIL molecules. There is of course a fraction of LC molecules that are not acted on by these dipoles but we can consider that this fraction is basically unaffected by the laser and can be neglected. The total effective polarization of the sample along the optic axis, caused by the laser, is then

$$P(r) = \langle n(r) \rangle (\mu_R + \mu_L) \langle \cos\theta(r) \rangle \quad (2)$$

According to our model, for $r < r_0$, $<n(r)> = n_0$, $<cos\theta(r)> \sim 0$, hence $P(r) \sim 0$, while, for $r \geq r_0$, $<n(r)> = n_0 \exp[-k_1 (r_0 - r)]$, $<cos\theta(r)> = 1 - \exp[k_2 (r_0 - r)]$

Hence,

$$P(r) = n_0 (\mu_R + \mu_L) \exp\left[-k_1 (r_0 - r)\right] \left(1 - exp\left[k_2 (r_0 - r)\right]\right) \quad (3)$$

A plot of $P(r)$ calculated from Equation (3) is shown in **Figure 7**. Though the value of the polarization is completely arbitrary, the functional form captures the essential physical aspects of the QSS, with nearly correct values for D and w, if k_1 and k_2 are set to the values of 2.0 and 0.025 μm^{-1}, respectively. These correspond to a "decay length" $\sim$ 500 nm for the dipoles, which corresponds to the average size of the nematic domains in the interfacial zone and an "effective range" $\sim$40 μm for the force responsible for orientational ordering. This relatively long-range force is responsible for the QSS.

The above model in itself, however, is inadequate since it cannot explain the slow dynamics at the nanocluster level. Here, we have to invoke the concept, apparent from the results, that, clusters can coalesce only when they are in the same orientational state. Since the orientational states of the large clusters fluctuate with time, coalescence to the bulk nematic phase has a fractional growth exponent, typical of random events. This mechanism is responsible for the QSS-like situation at the nanometer length scales. For the EMIG-MBBA mixture, our model indicates that the laser cannot produce dipoles but can only create the thermal fluctuation in the un-coordinated charges. Hence, we see a monotonically increasing isotropic domain, where the growth rate of that domain increases with laser power.

This qualitative and tentative model of the RTIL system, besides explaining the quasi-stationary states formed in this particular situation provides the framework for deciding the requirements to generate QSS in this class of materials that are emerging as highly promising for a wide range of application.

CONCLUSION

We have shown that a quasistationary state is generated at the nematic-isotropic interfacial region of a mixture of a liquid

crystalline material and an iron-based room temperature ionic liquid on focused irradiation by a laser. The quasistationary state consists of a coexistence of incessantly moving domains of the nematic phase at different orientational states and the isotropic phase. A mixture of the liquid crystal with a gallium-based analog of the ionic liquid does not produce the quasistationary state, though in absence of the laser both ionic liquids affect the nematic-isotropic phase transition in a similar way by lowering the transition temperature. The iron-based ionic liquid has a vibronic band that resonates with the laser wavelength and it is proposed that the resulting optical pumping produces molecular dipoles in the ionic liquid that interact with the liquid crystal molecular dipoles. This leads to a long range orientational ordering that competes with the dissipative forces of laser heating to generate the quasistationary state.

The iron-based ionic liquid in question has been found to exhibit long range antiferromagnetic order among spin orientations of the iron-containing anion [24] under equilibrium conditions. The long range molecular orientational order in the dissipative structures of the ionic liquid molecules themselves, as found in [17], or of the ionic liquid-liquid crystal mixtures, as found in this work, make these structures fascinating candidates for study of magnetic properties under non-equilibrium in general, and of the coupling between spin orientation and molecular orientation in particular.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AD carried out the experimental study, analyzed the data, and wrote the manuscript. KY defined the problem, provided the infrastructure, and took part in analysis and manuscript preparation. YY synthesized the samples in collaboration with GS. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP20H02708.

ACKNOWLEDGMENTS

AD would like to thank the Japan Society for Promotion of Science for a Visiting Professorship and the Department of Atomic Energy, Government of India, for a Raja Ramanna Fellowship, as well as Swapnasopan Datta of Jawaharlal Nehru Center for Advanced Scientific Research for **Figures 1A–C** and for useful discussions regarding the vibronic transition and structural changes of the anions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2020.583173/full#supplementary-material

REFERENCES

1. de Gennes PG, Prost J. *The Physics of Liquid Crystals*. Oxford: Oxford University Press (1995).
2. Dan K, Roy M, Datta A. Convex Arrhenius behaviour in a nematic-isotropic phase transition. *Europhys Lett*. (2014) **108**:36007. doi: 10.1209/0295-5075/108/36007
3. Dan K, Roy M, Datta A. Non-equilibrium phase transitions in a liquid crystal. *J Chem Phys*. (2015) **143**:094501. doi: 10.1063/1.4929607
4. Dan K, Roy M, Datta A. Entropic screening preserves non-equilibrium nature of nematic phase while enthalpic screening destroys it. *J Chem Phys*. (2016) **144**:064901. doi: 10.1063/1.4941365
5. Dan K, Datta A, Yoshida Y, Saito G, Yoshikawa K, Roy M. Screening out the non-arrhenius behaviour of nematic-isotropic transition by room temperature ionic liquid. *J Chem Phys*. (2016) **144**:084904. doi: 10.1063/1.4942521
6. de Hemptinne X. *Non-Equilibrium Statistical Thermodynamics*. Singapore: World Scientific (1992).
7. Mukai SA, Magome NH, Kitahata H, Yoshikawa K. Liquid/liquid dynamic phase separation induced by a focused laser. *Appl Phys Lett*. (2003) **83**:2557. doi: 10.1063/1.1613795
8. Sadakane K, Kitahata H, Seto H, Yoshikawa K. Rhythmic oscillation dynamic instability of micrometer-size phase separation under continuous photon flux by a laser. *Phys Rev E Stat Nonlin Soft Matter Phys*. (2008) **78**:046214. doi: 10.1103/PhysRevE.78.046214
9. Toyama H, Yoshikawa K, Kitahata H. Homogenization of a phase-separated droplet in a polymer mixture caused by the dielectric effect of a laser. *Phys Rev*. (2009) **78**:060801. doi: 10.1103/PhysRevE.78.060801
10. Latora V, Rapisarda A, Ruffo S. Lyapunov instability finite size effects in a system with long-range forces. *Phys Rev Lett*. (1998) **80**:692. doi: 10.1103/PhysRevLett.80.692
11. Campa A, Giansanti A, Morelli G. Long-time behavior of quasistationary states of the Hamiltonian mean-field model. *Phys Rev E*. (2007) **76**:041117. doi: 10.1103/PhysRevE.76.041117
12. Barré J, Bouchet F, Dauxois T, Ruffo S. Large deviation techniques applied to systems with long-range interactions. *J Stat Phys*. (2005) **119**:677–713. doi: 10.1007/s10955-005-3768-8
13. Fagotti M. Locally quasi-stationary states in noninteracting spin chains. *Sci Post Phys*. (2020) **8**:048. doi: 10.21468/SciPostPhys.8.3.048
14. Jara DAC, Alcaraz FC. Quasi-stationary states in nonlocal stochastic growth models with infinitely many absorbing states. *J Stat Mech*. (2017) **4**:043205. doi: 10.1088/1742-5468/aa668c
15. Beck M, Cooper E, Lord GJ, Spiliopoulos K. Selection of quasi-stationary states in the stochastically forced navier–stokes equation on the torus. *J Nonlin Sci*. (2020) **30**:1677–702. doi: 10.1007/s00332-020-09621-0
16. Welton T. Room-temperature ionic liquids. solvents for synthesis and catalysis. *Chem Rev*. (1999) **99**:2071–84. doi: 10.1021/cr980032t
17. Iguchi N, Datta A, Yoshikawa K, Yoshida Y, Saito G. A non-equilibrium quasistationary state in an ionic liquid caused by a focused laser. *Chem Phys Lett*. (2010) **485**:110–13. doi: 10.1016/j.cplett.2009.12.035
18. Yoshida Y, Otsuka A, Saito G, Natsume S, Nishibori E, Takata M, et al. Conducting and magnetic properties of 1-ethyl-3-methylimidazolium (EMI) salts containing paramagnetic irons: liquids $[EMI][M^{III}Cl_4]$ (M = Fe and $Fe_{0.5}Ga_{0.5}$) and solid $[EMI]_2[Fe^{II}Cl_4]$. *Bull Chem Soc Jpn*. (2005) **78**:1921–8. doi: 10.1246/bcsj.78.1921

19. O'Donoghue M. Characterization of crystals with gem applications. *Prog Cryst Growth Charact*. (1981) **3**:193–209. doi: 10.1016/0146-3535(80)90019-2
20. van der Beek D, Radstake PB, Petukhov AV, Lekkerkerker HNW. Fast formation of opal-like columnar colloidal crystals. *Langmuir*. (2007) **23**:11343–6. doi: 10.1021/la7012914
21. Yoshida Y, Saito G. Influence of structural variations in 1-alkyl-3-methylimidazolium cation and tetrahalogenoferrate(III) anion on the physical properties of the paramagnetic ionic liquids. *J Mat Chem*. (2006) **16**:1254–63. doi: 10.1039/b515391c
22. Murata K, Irish DE. Raman studies of the hydrated melt of $FeCl_3 6H_2O$. *Spectrochim Acta*. (1988) **44**:739–43. doi: 10.1016/0584-8539(88)80136-3
23. Krzystek K, Telser J. Insight into electronic and magnetic properties through HFEPR studies. In: Holynska M, editor. *Single-Molecule Magnets: Molecular Architectures and Building Blocks for Spintronics*. John Wiley and Sons (2019). p. 135–72.
24. de Pedro I, Rojas DP, Albo J, Luis P, Irabien A, Blanco JA, et al. Long-range magnetic ordering in magnetic ionic liquid: emim[$FeCl_4$]. *J Phys Condens Matter*. (2010) **22**:296006. doi: 10.1088/0953-8984/22/29/296006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 05 November 2020
doi: 10.3389/fchem.2020.605403

Dynamics of Photoinduced Energy Transfer in Fully and Partially Conjugated Polymers Bearing π-Extended Donor and Acceptor Monomers

Youngseo Kim†, Na Yeon Kwon†, Su Hong Park, Min Ju Cho, Dong Hoon Choi and Sungnam Park**

Department of Chemistry, Research Institute for Natural Sciences, Korea University, Seoul, South Korea

OPEN ACCESS

Edited by:
Rene A. Nome,
State University of Campinas, Brazil

Reviewed by:
Wenkai Zhang,
Beijing Normal University, China
Dibyendu Mondal,
Jain University, India

****Correspondence:***
Dong Hoon Choi
dhchoi8803@korea.ac.kr
Sungnam Park
spark8@korea.ac.kr

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Chemistry

Received: *12 September 2020*
Accepted: *19 October 2020*
Published: *05 November 2020*

Citation:
Kim Y, Kwon NY, Park SH, Cho MJ, Choi DH and Park S (2020) Dynamics of Photoinduced Energy Transfer in Fully and Partially Conjugated Polymers Bearing π-Extended Donor and Acceptor Monomers. Front. Chem. 8:605403. doi: 10.3389/fchem.2020.605403

The photophysical properties of donor (**D**)-acceptor (**A**) polymers were studied by designing two types of polymers, **(D-σ-A)$_n$** and **(D-π-A)$_n$**, with non-conjugated alkyl (sp^3) and π-conjugated (sp^2) linkers using π-extended donor and acceptor monomers that exhibit planar A-D-A structures. The non-conjugated alkyl linker provides structural flexibility to the **(D-σ-A)$_n$** polymers, while the π-conjugated linker retains the rigid structure of the **(D-π-A)$_n$** polymers. Photoinduced energy transfer occurs from the large donor to acceptor units in both polymers. However, the photoinduced energy transfer dynamics are found to be dependent on the conformation of the polymers, where the difference is dictated by the types of linkers between the donor and acceptor units. In solution, intramolecular energy transfer is relatively favorable for the **(D-σ-A)$_n$** polymers with flexible linkers that allow the donor and acceptor units to be proximally located in the polymers. On the other hand, intermolecular (or interchain) energy transfer is dominant in the two polymer films because the π-extended donor and acceptor units in polymers are closely packed. The structural flexibility of the linkers between the donor and acceptor repeating units in the polymers affects the efficiency of energy transfer between the donor and acceptor units and the overall photophysical properties of the polymers.

Keywords: π-extended donor monomer, π-extended acceptor monomer, fully conjugated polymer, partially conjugated polymer, photophysical property, time-resolved fluorescence, energy transfer

INTRODUCTION

Photoinduced energy transfer or charge transfer has been extensively studied using dyad- or triad-type small molecules composed of donor and acceptor moieties (Kuss-Petermann et al., 2012; Wiebeler et al., 2017; Wang et al., 2019). Most studies on the photophysical properties of such molecules have been carried out using boron dipyrromethenes or porphyrin as the donor moiety (Duvanel et al., 2013; Villamaina et al., 2013; Badgurjar et al., 2016). In dyad systems, the donor and acceptor moieties are connected by different types of linkers. The photophysical properties of dyad systems are largely influenced by the linkers between the donor and acceptor moieties; that is, π-conjugated (sp^2-type) and non-conjugated (or aliphatic, sp^3-type) linkers. In the past, various donor (**D**)- acceptor (**A**)-type dyad systems have been designed and synthesized as electronic

and optoelectronic materials, including organic photovoltaic cells, organic field effect transistors, and organic light-emitting diodes, and their photophysical properties have been studied (Wu and Brand, 1994; Scholes, 2003; Schwartz, 2003; Murphy et al., 2004; Jones and Bradshaw, 2019). Although the focus has been on small-molecule donors and fullerene receptor systems (Jose et al., 2009; Caprasecca and Mennucci, 2014), a few studies have been performed on the photophysical properties of conjugated polymer systems. For example, Feng et al. introduced a single active polymer by binding a small molecule that exhibits acceptor properties as a side chain moiety to the main chain conjugated polymer with donor properties (Feng et al., 2019; Li et al., 2019). They observed new optoelectronic properties in this polymer by enabling energy transfer or electron transfer in the electronically excited state.

Our group reported polymer solar cells with two different copolymers synthesized by connecting conjugated donor- and acceptor-based macromolecular units (Lee et al., 2017). In our previous study, donor- and acceptor-based macromolecular units were connected using sp^3 and sp^2 linkers, respectively, to investigate the performance of polymer solar cells. The performance of polymer solar cells was found to dependent on the linkers between donor- and acceptor-based macromolecular units in the polymers. However, photoinduced energy transfer and photoinduced electron transfer in the two polymers were not able to be studied. Specifically, the donor- and acceptor-based macromolecular units were characterized by a large molecular weight distribution, making it difficult to obtain well-defined spectroscopic features in the UV-visible absorption and steady-state emission spectra. Generally, the photoinduced energy transfer and photoinduced electron transfer process in the polymers are important at the molecular level for optoelectronic devices and need to be comprehensively studied. The photoinduced energy transfer and photoinduced electron transfer in the polymers can be reliably studied by using the polymers with donor- and acceptor-based macromolecular units having a uniform molecular weight.

In this study, we designed and synthesized large A-D-A type donor monomers (**B3TP**, **D**) and A-D1-D2-D1-A type acceptor monomers (**B2IC**, **A**), and utilized them to synthesize two different polymers with non-conjugated alkyl linkers, **(D-σ-A)$_n$**, and π-conjugated linkers, **(D-π-A)$_n$**. The structural flexibility of **(D-σ-A)$_n$** and **(D-π-A)$_n$** is dictated by the linkers. Overall, **(D-σ-A)$_n$** is the partially conjugated polymer in which the π-extended donor and acceptor monomers are connected through flexible non-conjugated sp^3 linkers. In contrast, **(D-π-A)$_n$** is the fully conjugated polymer in which the large donor and acceptor monomers are connected *via* conjugated sp^2 linkers. It is very noteworthy that such polymer structures contain highly π-extended donor and acceptor monomers with a uniform molecular weight, and the large donor and acceptor monomers are alternatively arranged in the polymers. The photophysical properties of **(D-σ-A)$_n$** and **(D-π-A)$_n$** in toluene and films were investigated using UV-visible absorption, steady-state emission, and time-resolved fluorescence (TRF) spectroscopy, in conjunction with quantum chemical calculations. The photophysical properties of **(D-σ-A)$_n$** and **(D-π-A)$_n$** depend on the type of linker between the large donor and acceptor units in the polymers. The energy profiles of **(D-σ-A)$_n$** and **(D-π-A)$_n$** were found to be similar, but the photoinduced energy transfer from the donor to acceptor units occurred via different pathways. Photoinduced energy transfer in **(D-σ-A)$_n$** and **(D-π-A)$_n$** in toluene is mainly due to the intramolecular energy transfer, whereas photoinduced energy transfer in the **(D-σ-A)$_n$** and **(D-π-A)$_n$** films proceeds via both intramolecular and intermolecular (or interchain) energy transfer. This is the first report on the photophysical properties of partially and fully conjugated polymers containing well-defined π-extended donor and acceptor monomers. And our current results are expected to be applied to optoelectronic devices using π-conjugated donor-acceptor type polymers.

EXPERIMENTAL

Synthesis

Compound (D-σ-A)$_n$

In a Schlenk tube, **M1** (85.3 mg, 28.8 μmol), **M2** (75 mg, 28.8 μmol), $Pd_2(dba)_3$ (1.3 mg, 5 mol%), and P(*o*-tolyl)$_3$ (1.75 mg, 10 mol%) were dissolved in toluene (4 mL). The degassed binary mixture was stirred at 180°C for 2 h and subjected to cross-coupling Stille reaction under optimized microwave irradiation. The synthesized crude polymers were purified by precipitation into methanol, followed by Soxhlet extraction to remove unreacted monomers, undesired byproducts, and low-molecular-weight materials, using acetone, hexane, and dichloromethane in succession. The polymer solution was precipitated from methanol, and the solid was filtered to obtain polymer **(D-σ-A)$_n$** as a black solid [M_n = 24.04 kDa, polydispersity index (PDI) = 2.4]. Elemental Anal. Calcd. for $(C_{324}H_{352}N_6O_2S_{22})_n$: C: 76.82; H: 6.96; N: 1.66; S: 13.92. Found: C: 76.18; H: 6.83; N: 1.71; S:13.62.

Compound (D-π-A)$_n$

In a Schlenk tube, **M1** (130 mg, 44.1 mmol), **M3** (100 mg, 44.1 mmol), and $Pd(PPh_3)_4$ (5.0 mg, 10 mol%) were dissolved in toluene (4 mL). The synthesis method and reaction conditions were the same as described for the microwave synthesis of **(D-σ-A)$_n$**. **(D-π-A)$_n$** is a black solid (M_n = 10.92 kDa, PDI = 2.3). Elemental Anal. Calcd. for $(C_{304}H_{324}N_6O_2S_{20})_n$: C: 76.14; H: 6.86; N: 1.78; S, 9.55. Found: C:75.87; H: 6.76; N: 1.83; S: 9.32.

The chemical structures of **(D-σ-A)$_n$** and **(D-π-A)$_n$** were verified by gel permeation chromatography, ^{1}H nuclear magnetic resonance (NMR; **Supplementary Figures 2, 3**), and elemental analyses (EA). The M_n values of the intermediate **(D-σ-A)$_n$** block and the final **(D-π-A)$_n$** were 20.0 and 10.9 kg mol^{-1}, respectively, and the corresponding PDI values were 2.4 and 2.3, respectively. Because the molecular weight of the repeating units consisting of the large monomers is over 5,000 g/mol, the molar ratio of the **(D-σ-A)$_n$** and **(D-π-A)$_n$** repeating units was determined to be 1:1 from NMR spectra and theoretical EA calculations.

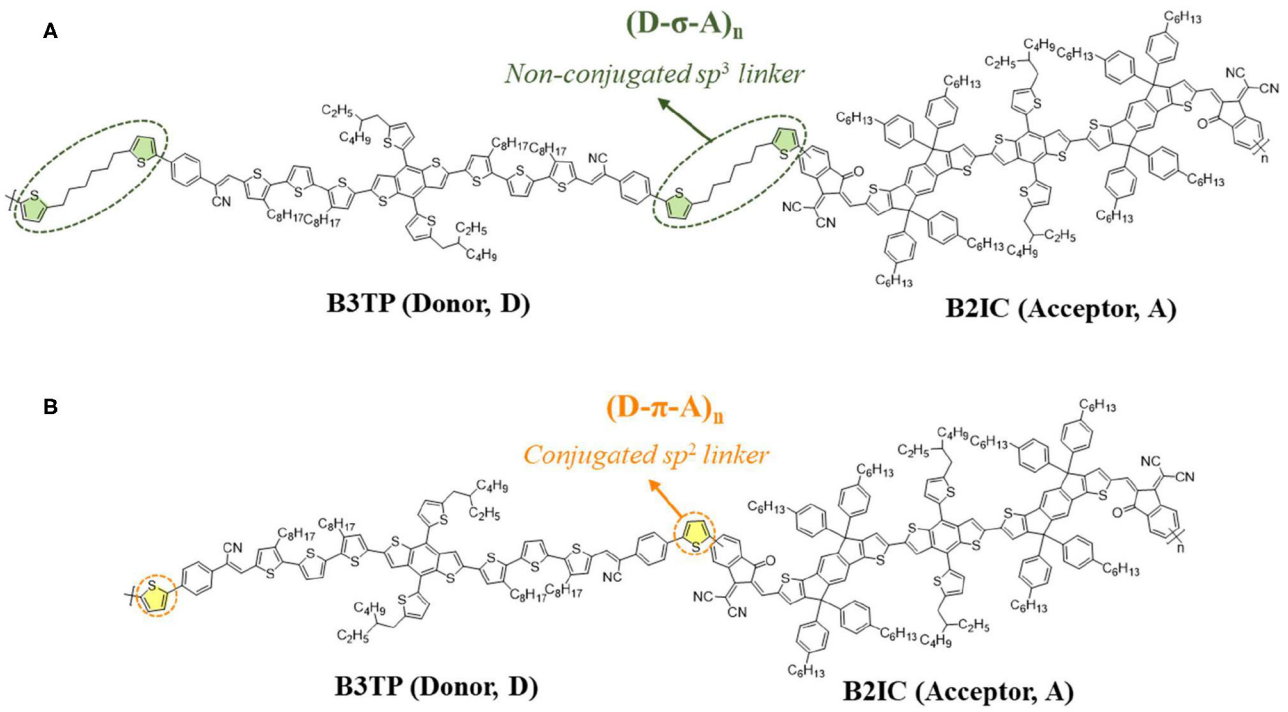

FIGURE 1 | Molecular structures of **(A)** $(\mathbf{D\text{-}\sigma\text{-}A})_n$ and **(B)** $(\mathbf{D\text{-}\pi\text{-}A})_n$ polymers.

RESULTS AND DISCUSSION

Design, Synthesis, and Characterization

The acceptor-donor-acceptor (A-D-A) backbone architecture of donor and acceptor monomers has drawn particular interest because of the easily tuned energy levels of these species. The benzo[1,2-*b*:4,5-*b*]dithiophene (BDT) unit used as the main core of π-extended donor and acceptor monomers in this study has become one of the most widely used polymer backbones because of its rigid planar conjugated structure, which can enhance electron delocalization and promote cofacial π-π stacking interactions in the solid state (Wang et al., 2018).

A design strategy for preparing effective A-D-A type donor monomers with narrow bandgaps involves increasing the electron donating ability of the donor monomers by extending the effective π conjugation length. The extended π-conjugated structure with three thiophenes increases the stiffness of the donor backbone. **B3TP**, as an A-D-A type conjugated donor monomer, was designed as shown in **Figure 1**. In **B3TP**, alkyl-substituted terthiophene moieties are added at both ends of the BDT core (Yang et al., 2018) and electron-withdrawing phenylacetonitrile groups are attached as the end-capped group, which guarantees a symmetric and planar structure. As an effective π-extended acceptor monomer, **B2IC** was designed in an acceptor-donor1-donor2-donor1-acceptor (A-D1-D2-D1-A) structure, as shown in **Figure 1**. Such A-D1-D2-D1-A type acceptors have rarely been reported. **B2IC** consists of the BDT core (D2) with indaceno[2,1-*b*:6,5-*b*]dithiophene moieties as an additional donor unit (D1) and the end-capping groups of 1,1-dicyanomethylene-3-indanone (IC) as an accepting unit (A).

In dyad systems, photoinduced energy transfer or photoinduced electron transfer is interesting photophysical phenomena that significantly depend on the linker type and length. To study such phenomena, we synthesized $(\mathbf{D\text{-}\sigma\text{-}A})_n$ and $(\mathbf{D\text{-}\pi\text{-}A})_n$ polymers by connecting the donor (**B3TP**) and acceptor (**B2IC**) units using an sp^2 type conjugated linker (thiophene) and sp^3 type non-conjugated linker. A detailed investigation of the photophysical properties is presented in the following sections.

Theoretical Studies of π-Extended Monomers, (D-σ-A) and (D-π-A)

To study the electronic and optical properties, density functional theory (DFT) calculations [B3LYP/6-31g(d) and B3LYP-d3/6-31g(d)] were carried out. First, the optimized structures of **B3TP**, **B2IC**, **D-σ-A**, and **D-π-A** were obtained. The frontier orbitals (HOMO and LUMO) and natural transition orbitals (NTOs) were calculated. **Figures 2A,B** show the optimized structures and the HOMO and LUMO of **B3TP** and **B2IC**, respectively. The donor monomer, **B3TP** (A-D-A type), and the acceptor monomer, **B2IC** (A-D1-D2-D1-A type), are both planar and fully conjugated. Both **B3TP** and **B2IC** exhibit intramolecular charge transfer upon electronic transition (**Supplementary Figure 4**).

The optimized structures and frontier orbitals of **D-σ-A** and **D-π-A** are shown in **Figure 2**. Note that the folded

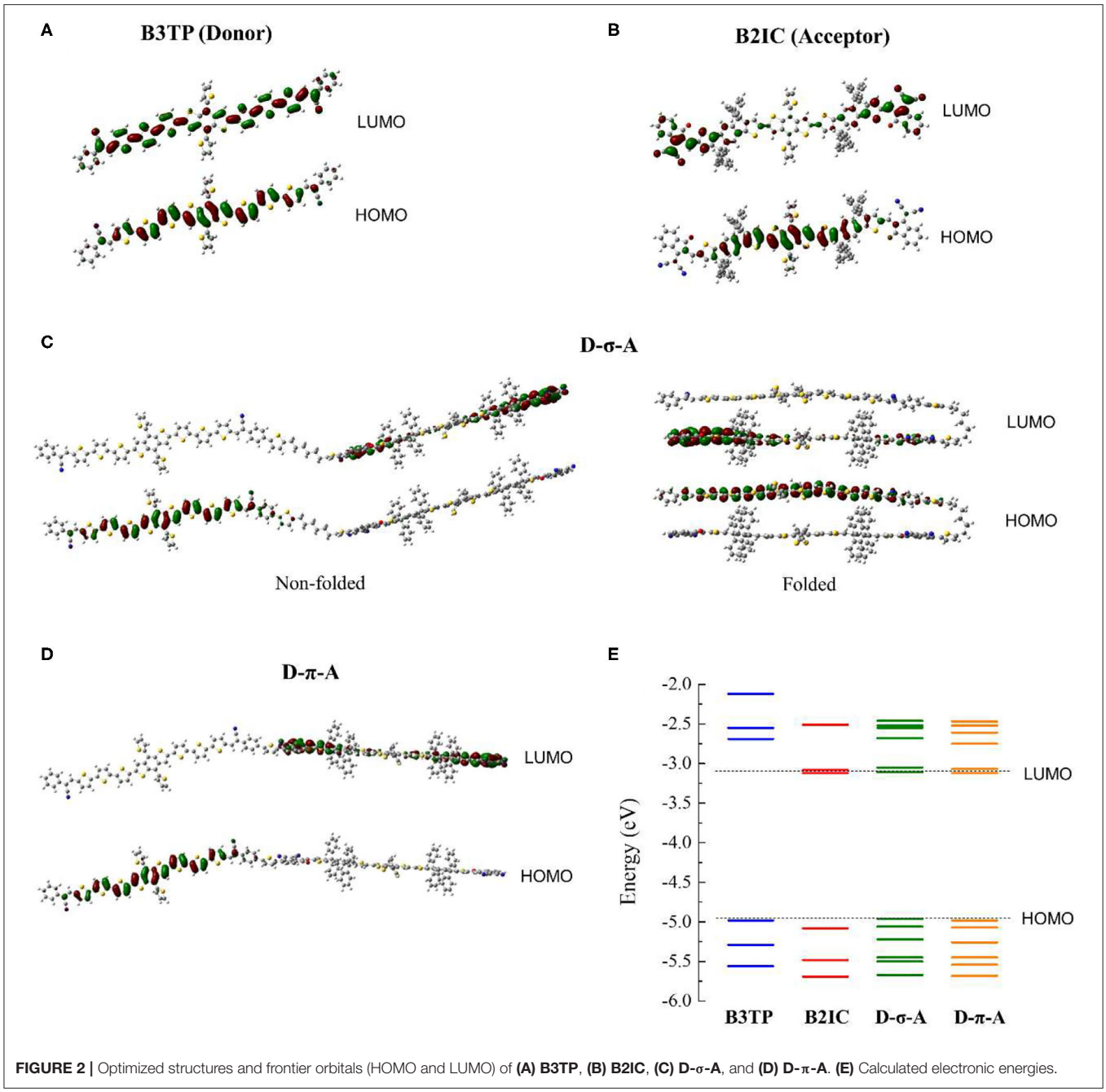

FIGURE 2 | Optimized structures and frontier orbitals (HOMO and LUMO) of **(A) B3TP**, **(B) B2IC**, **(C) D-σ-A**, and **(D) D-π-A**. **(E)** Calculated electronic energies.

structure of **D-σ-A** in **Figure 2** is also energetically stable and can possibly be populated. Because **D-σ-A** and **D-π-A** have the same donor and acceptor units, but different linkers (sp^3 vs. sp^2) between the donor and acceptor units, the calculated electronic energies of the HOMO and LUMO levels are almost the same (**Figure 2E**). For both **D-σ-A** and **D-π-A**, the HOMO and LUMO are spatially localized at the π-extended donor and acceptor units, respectively, and exhibit strong charge transfer characteristics, which is directly associated with the photophysical properties of $\mathbf{(D\text{-}\sigma\text{-}A)_n}$ and $\mathbf{(D\text{-}\pi\text{-}A)_n}$, as discussed hereinafter.

Spectroscopic Studies of π-Extended Monomers, $\mathbf{(D\text{-}\sigma\text{-}A)_n}$, and $\mathbf{(D\text{-}\pi\text{-}A)_n}$

The UV-visible absorption and steady-state emission spectra of **B3TP**, **B2IC**, $\mathbf{(D\text{-}\sigma\text{-}A)_n}$, and $\mathbf{(D\text{-}\pi\text{-}A)_n}$ in toluene are shown in **Figure 3**. The UV-visible absorption spectra of $\mathbf{(D\text{-}\sigma\text{-}A)_n}$ and $\mathbf{(D\text{-}\pi\text{-}A)_n}$ are similar, with two absorption peaks near 480 and 660 nm, representing the donor and acceptor units (**B3TP** and **B2IC**), respectively. The emission spectrum of **B3TP** (**Figure 3B**) exhibits an intense peak at 590 nm (λ_{ex} = 520 nm), whereas **B2IC** shows very weak emission at 710 nm. **B3TP** is nearly planar and structurally rigid, which seems to give a relatively

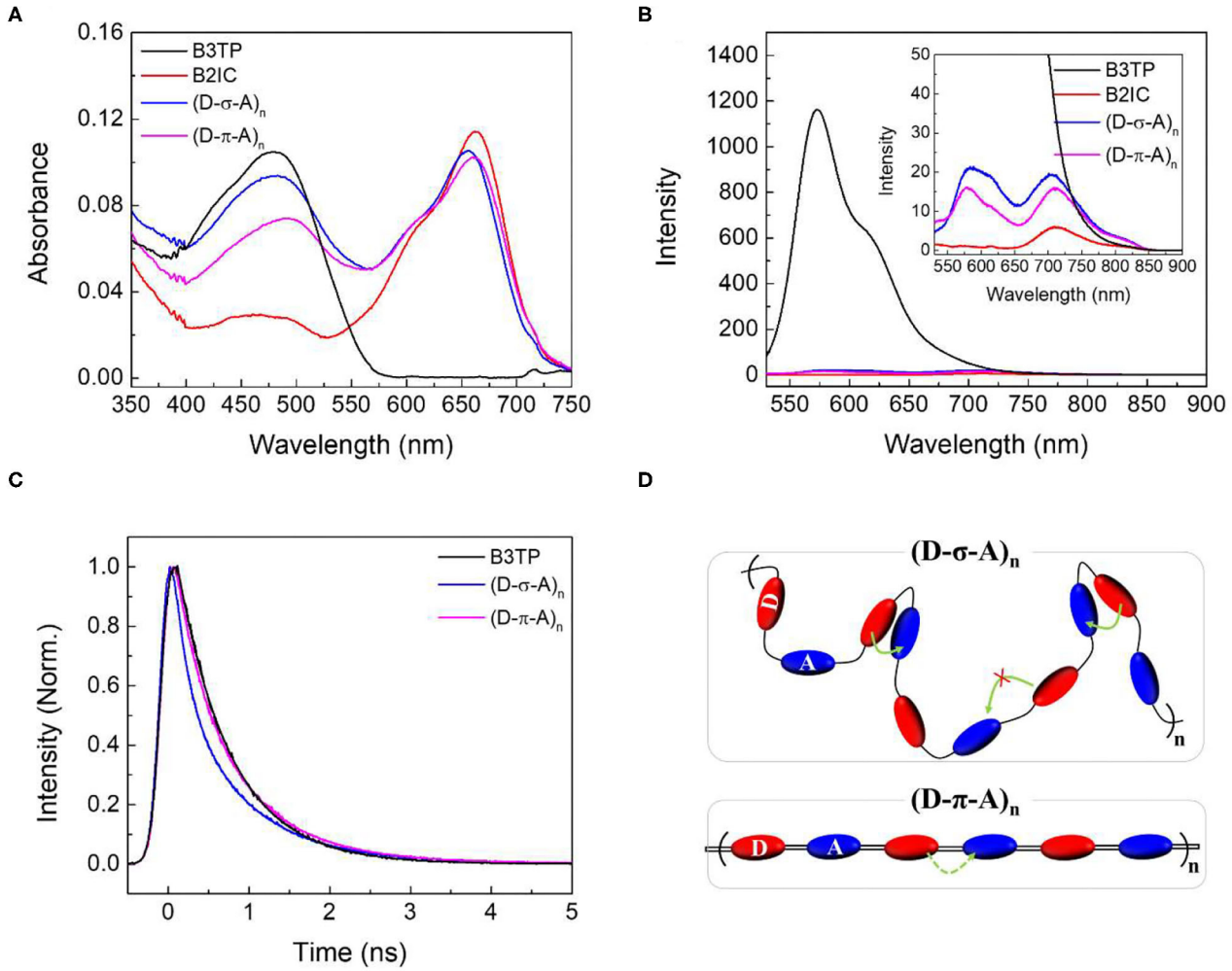

FIGURE 3 | (A) UV-visible absorption spectra and **(B)** steady-state emission spectra of **B3TP**, **B2IC**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in toluene. Steady-state emission spectra were measured at the excitation of $\lambda_{ex} = 520$ nm. **(C)** Time-resolved fluorescence signals were obtained at 590 nm at the excitation of $\lambda_{ex} = 520$ nm. The concentration of **B3TP**, **B2IC**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in toluene is 0.2 μM. **(D)** Schematic illustration of **(D-σ-A)$_n$** and **(D-π-A)$_n$** in toluene and the energy transfer between π-extended donor and acceptor units. Intramolecular Förster resonance energy transfer in toluene is indicated by curved arrows.

large fluorescence quantum yield. In contrast, **B2IC** contains 8 phenyl groups in the donor (D1) units, which are flexible and freely rotatable in toluene and might be responsible for significant fluorescence quenching.

As shown in the inset of **Figure 3B**, **(D-σ-A)$_n$** and **(D-π-A)$_n$** exhibit two emission peaks at 590 and 710 nm ($\lambda_{ex} = 520$ nm), resulting from **B3TP** and **B2IC**, respectively. However, the overall fluorescence of **(D-σ-A)$_n$** and **(D-π-A)$_n$** is significantly quenched when compared with that of **B3TP**. This observation can be readily explained based on the fluorescence properties of **B3TP** and **B2IC** as follows. The donor units (**B3TP**) in **(D-σ-A)$_n$** and **(D-π-A)$_n$** are directly excited by 520 nm photons. The excited donor units in **(D-σ-A)$_n$** and **(D-π-A)$_n$** undergo Förster resonance energy transfer (FRET) to the acceptor units (**B2IC**) in **(D-σ-A)$_n$** and **(D-π-A)$_n$**, owing to which, the fluorescence is quenched. FRET is readily possible in **(D-σ-A)$_n$** and **(D-π-A)$_n$** because the emission spectrum of the donor (**B3TP**) and UV-visible absorption spectrum of the acceptor (**B2IC**) overlap (**Supplementary Figure 5**), and the distance between the donor and acceptor units in the polymer backbones is sufficiently small.

To further understand the fluorescence quenching, the TRF signals of **B3TP**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in toluene were measured using the time-correlated single-photon counting method, as shown in **Figure 3C**. **B3TP**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in toluene were excited by 520 nm laser pulses and their TRF signals were measured at 590 nm to monitor the emission of the donor monomers (**B3TP**). The TRF signal of **(D-σ-A)$_n$** (**Figure 3C**) decayed faster than that of **B3TP** and **(D-π-A)$_n$** in toluene. The TRF signals were fitted to a multi-exponential function, and the average fluorescence lifetimes are summarized in **Table 1**. The average fluorescence lifetimes ($\tau_{avg} = 0.67$ ns) of **B3TP** and **(D-π-A)$_n$** in toluene are the same within experimental error, even though the initial TRF decay of **(D-π-A)$_n$** is slightly

TABLE 1 | Exponential fit of time-resolved fluorescence signals, $S(t) = \sum_i A_i \exp(-t/\tau_i)$, of **B3TP**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in toluene and films.

	B3TP		**(D-σ-A)$_n$**		**(D-π-A)$_n$**	
	Toluene	**Film**	**Toluene**	**Film**	**Toluene**	**Film**
A_1	0.94	0.77	0.46	0.98	0.30	0.97
τ_1	0.64 ns	0.16 ns	0.17 ns	0.10 ns	0.32 ns	0.11 ns
A_2	0.06	0.23	0.52	0.02	0.69	0.03
τ_2	1.1 ns	0.54 ns	0.79 ns	0.64 ns	0.79 ns	0.70 ns
A_3	–	–	0.02	–	0.01	–
τ_3	–	–	2.4 ns	–	2.8 ns	–
τ_{avg}	0.67 ns	0.25 ns	0.54 ns	0.11 ns	0.67 ns	0.13 ns

The average lifetime was calculated as $\tau_{avg} = \sum_i A_i\tau_i / \sum_i A_i$.

faster than that of **B3TP**. The average fluorescence lifetime (τ_{avg} = 0.54 ns) of **(D-σ-A)$_n$** in toluene is slightly shorter than that of **B3TP** and **(D-π-A)$_n$**. The difference in the TRF signals of **(D-σ-A)$_n$** and **(D-π-A)$_n$** in toluene results from their structural difference, caused by the linkers between the donor and acceptor units, as illustrated in **Figure 3D**. In **(D-π-A)$_n$**, the π-conjugated linker (sp^2) between the donor and acceptor units is rigid, which fixes the donor and acceptor far from each other, as shown in **Figure 3D**. Therefore, FRET between the donor and acceptor units in **(D-π-A)$_n$** is less likely. In contrast, in **(D-σ-A)$_n$**, the sp^3 alkyl chain between the donor and acceptor units is flexible; thus, a folded structure can feasibly be formed as shown in **Figure 2C**. Thus, intramolecular FRET between the donor and acceptor units in the folded sections of **(D-σ-A)$_n$** is likely, as schematically illustrated in **Figure 3D**. The optical properties of **B3TP**, **B2IC**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in the films were further studied, as shown in **Figure 4**. The UV-visible absorption spectra of **B3TP**, **B2IC**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in the films were slightly red-shifted relative to those in toluene (**Supplementary Figure 6**). The spectral features indicate that the donor and acceptor units are closely packed and intermingled in the films, as illustrated in **Figure 5A**. This structural arrangement in the films leads to a significant change in the fluorescence properties of **(D-σ-A)$_n$** and **(D-π-A)$_n$**. The steady-state emission spectra of **(D-σ-A)$_n$** and **(D-π-A)$_n$** in **Figure 4B** show the emission peak of the acceptor monomers (**B2IC**). This indicates completely efficient FRET from the large donor to acceptor units in **(D-σ-A)$_n$** and **(D-π-A)$_n$**. In the films, the donor and acceptor units are closely located so that intermolecular FRET from the donor and acceptor units in **(D-σ-A)$_n$** and **(D-π-A)$_n$** becomes very efficient. **Figure 4C** shows the TRF signals of **B3TP**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in the films. The TRF signal of **B3TP** in the films is found to decay faster than that of **B3TP** in toluene because of the "aggregation causing quenching" effect. The TRF signal of **(D-π-A)$_n$** in the films decays much faster than that of **(D-π-A)$_n$** in toluene, resulting from additional intermolecular (or interchain) FRET between the π-extended donor and acceptor units. The TRF signals of the films were fitted by a multi-exponential function, and the average fluorescence lifetimes are summarized in **Table 1**.

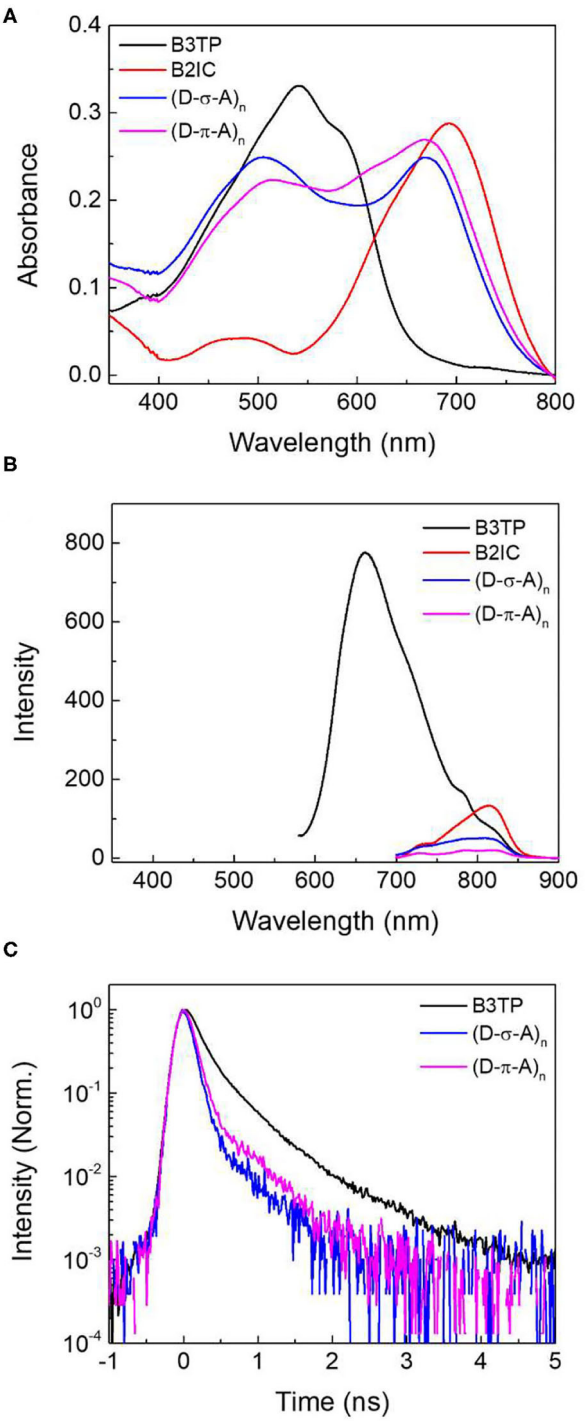

FIGURE 4 | (A) UV-visible absorption spectra and **(B)** steady-state emission spectra of **B3TP**, **B2IC**, **(D-σ-A)$_n$**, and **(D-π-A)$_n$** in films. The steady-state emission spectra were measured at the excitation of λ_{ex} = 520 nm. **(C)** Time-resolved fluorescence signals were obtained at 670 nm (λ_{ex} = 520 nm).

By using a simple kinetics model (**Figure 5B**), the intramolecular and intermolecular (or interchain) FRET rate constants can be estimated (**Supplementary Table 1**). As

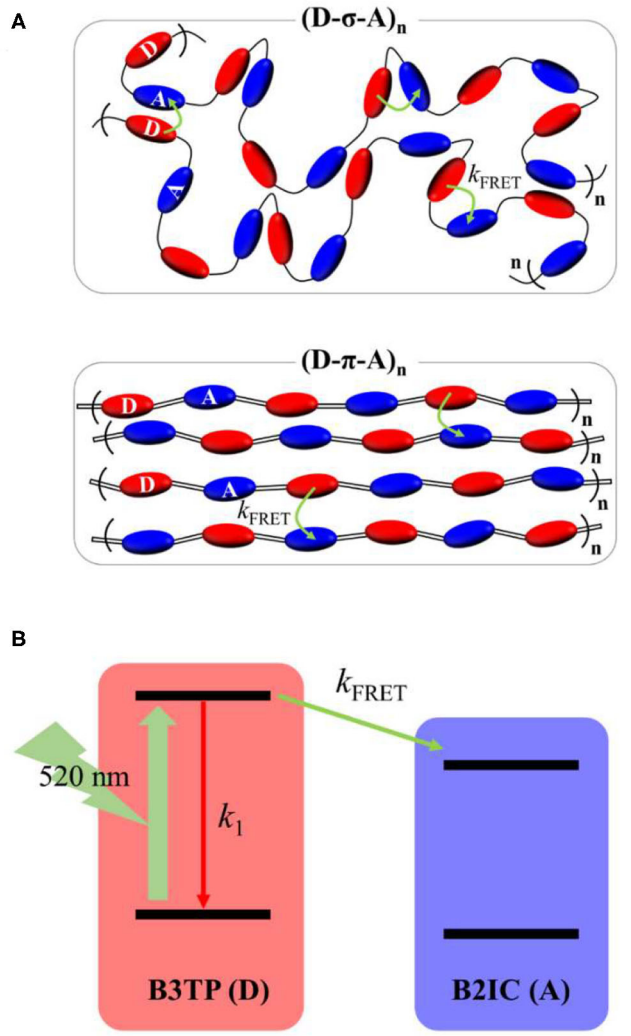

FIGURE 5 | (A) Schematic illustration of **(D-σ-A)$_n$** and **(D-π-A)$_n$** in films and Förster resonance energy transfer (FRET) from the donor (**B3TP**, **D**) to acceptor (**B2IC**, **A**) units. Intramolecular and intermolecular (or interchain) FRET are indicated by curved arrows. **(B)** Kinetic model illustrating FRET in **(D-σ-A)$_n$** and **(D-π-A)$_n$**.

shown in **Table 1**, the average fluorescence lifetime of **(D-σ-A)$_n$** (τ_{avg} = 0.54 ns) in toluene is shorter than that of **B3TP** (τ_{avg} = 0.67 ns) in toluene because of the intramolecular FRET from the donor units (**B3TP**, **D**) to the acceptor units (**B2IC**, **A**). The intramolecular FRET rate constant of **(D-σ-A)$_n$** in toluene was estimated to be $k_{FRET} = 0.36$ ns^{-1}. In the same way, the FRET rate constants of **(D-σ-A)$_n$** and **(D-π-A)$_n$** in the films were determined to be $k_{FRET} = 5.09$ and 3.69 ns^{-1}, respectively. The FRET rate constant was found to be much greater in the films than in toluene. This is because the donor and acceptor units in **(D-σ-A)$_n$** and **(D-π-A)$_n$** are much more closely located in the films, and intramolecular and intermolecular FRET are both very efficient in the films, as illustrated in **Figure 5A**.

CONCLUSIONS

To study the role of linkers in energy transfer, we designed and synthesized two dyad polymer systems. First, the π-extended donor monomers (**B3TP**, **D**) with an A-D-A structure and the π-extended acceptor monomers (**B2IC**, **A**) with an A-D1-D2-D1-A structure were synthesized. These donor and acceptor units were linked together by two different linkers (non-conjugated alkyl and π-conjugated linker) to synthesize two types of D-A repeating polymers, **(D-σ-A)$_n$** and **(D-π-A)$_n$**.

The photophysical properties of **(D-σ-A)$_n$** and **(D-π-A)$_n$** were comprehensively investigated by UV-visible absorption, steady-state emission, and TRF spectroscopy, in conjunction with DFT calculations. DFT calculations show that the π-extended donor and acceptor monomers (**B3TP** and **B2IC**) are planar; **D-σ-A** and **D-π-A** have similar energy profiles, originating from the donor and acceptor units, and exhibit intramolecular charge transfer characteristics upon electronic excitation. A folded structure is possible for **(D-σ-A)$_n$** because of the flexible non-conjugated alkyl linkers. **(D-σ-A)$_n$** and **(D-π-A)$_n$** exhibit similar UV-visible absorption properties. Their fluorescence was found to be quenched due to FRET from the large donor to acceptor units in **(D-σ-A)$_n$** and **(D-π-A)$_n$**. In toluene, FRET is more efficient for **(D-σ-A)$_n$** than for **(D-π-A)$_n$**. It is likely that the donor and acceptor units are more closely located in the folded structure of **(D-σ-A)$_n$**, which increases the speed of FRET. In the films, the fluorescence of **(D-σ-A)$_n$** is more significantly quenched than that of **(D-π-A)$_n$**, and intramolecular and intermolecular (or interchain) FRET occur more efficiently. The FRET rate constants were determined using a simple kinetics model.

For the two conjugated polymers bearing π-extended donor and acceptor monomers, the linkers between the donor and acceptor units are shown to play an important role in determining the overall structure of the polymers, thereby influencing the optical properties. The flexible alkyl linkers in the repeating group of the polymers lead to more structural freedom in solution, while the rigid π-conjugated linkers provide less flexibility to the fully conjugated polymers. This structural flexibility can significantly impact the optical properties, and can be tailored depending on the application of the donor- and acceptor-based conjugated polymers for a wide range of research fields.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

DC and SP conceived the research. Chemical synthesis was carried out by NK, SHP, and MC. Spectroscopic experiments and DFT calculations were done by YK. YK and SP analyzed

the spectroscopic data. NK, YK, DC, and SP wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Research Foundation of Korea (NRF-2020R1I1A1A01066897, 2019R1A2C2002647, and 2019R1A6A1A11044070) and Korea University-Future Research Grant (KU-FRG).

ACKNOWLEDGMENTS

MALDI-TOF mass spectra were obtained at the Korea Basic Science Institute (Seoul Center).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2020.605403/full#supplementary-material

REFERENCES

Badgurjar, D., Sudhakar, K., Jain, K., Kalantri, V., Venkatesh, Y., Duvva, N., et al. (2016). Ultrafast intramolecular photoinduced energy transfer events in benzothiazole–borondipyrromethene donor–acceptor dyads. *J. Phys. Chem. C* 120, 16305–16321. doi: 10.1021/acs.jpcc.6b03668

Caprasecca, S., and Mennucci, B. (2014). Excitation energy transfer in donor-bridge-acceptor systems: a combined quantum-mechanical/classical analysis of the role of the bridge and the solvent. *J. Phys. Chem. A* 118, 6484–6491. doi: 10.1021/jp502815r

Duvanel, G., Grilj, J., and Vauthey, E. (2013). Ultrafast long-distance excitation energy transport in donor-bridge-acceptor systems. *J. Phys. Chem. A* 117, 918–928. doi: 10.1021/jp311540x

Feng, G., Li, J., He, Y., Zheng, W., Wang, J., Li, C., et al. (2019). Thermal-driven phase separation of double-cable polymers enables efficient single-component organic solar cells. *Joule* 3, 1765–1781. doi: 10.1016/j.joule.2019.05.008

Jones, G. A., and Bradshaw, D. S. (2019). Resonance energy transfer: from fundamental theory to recent applications. *Front. Phys.* 7:100. doi: 10.3389/fphy.2019.00100

Jose, J., Ueno, Y., Castro, J. C., Li, L., and Burgess, K. (2009). Energy transfer dyads based on Nile Red. *Tetrahed. Lett.* 50, 6442–6445. doi: 10.1016/j.tetlet.2009.08.130

Kuss-Petermann, M., Wolf, H., Stalke, D., and Wenger, O. S. (2012). Influence of donor-acceptor distance variation on photoinduced electron and proton transfer in rhenium(I)-phenol dyads. *J. Am. Chem. Soc.* 134, 12844–12854. doi: 10.1021/ja3053046

Lee, D. H., Lee, J. H., Kim, H. J., Choi, S., Park, G. E., Cho, M. J., et al. (2017). (D)n–σ-(A)m type partially conjugated block copolymer and its performance in single-component polymer solar cells. *J. Mater. Chem. A* 5, 9745–9751. doi: 10.1039/C7TA01819C

Li, C., Wu, X., Sui, X., Wu, H., Wang, C., Feng, G., et al. (2019). Crystalline cooperativity of donor and acceptor segments in double-cable conjugated polymers toward efficient single-component organic solar cells. *Angew. Chem. Int. Ed. Engl.* 58, 15532–15540. doi: 10.1002/anie.201910489

Murphy, C. B., Zhang, Y., Troxler, T., Ferry, V., Martin, J. J., and Jones, W. E. (2004). Probing Förster and dexter energy-transfer mechanisms in fluorescent conjugated polymer chemosensors. *J. Phys. Chem. B* 108, 1537–1543. doi: 10.1021/jp0301406

Scholes, G. D. (2003). Long-range resonance energy transfer in molecular systems. *Annu. Rev. Phys. Chem.* 54, 57–87. doi: 10.1146/annurev.physchem.54.011002.103746

Schwartz, B. J. (2003). Conjugated polymers as molecular materials: how chain conformation and film morphology influence energy transfer and interchain interactions. *Annu. Rev. Phys. Chem.* 54, 141–172. doi: 10.1146/annurev.physchem.54.011002.103811

Villamaina, D., Bhosale, S. V., Langford, S. J., and Vauthey, E. (2013). Excited-state dynamics of porphyrin-naphthalenediimide-porphyrin triads. *Phys. Chem. Chem. Phys.* 15, 1177–1187. doi: 10.1039/C2CP43595K

Wang, L., Pan, C., Chen, Z., Zhou, X., Gao, C., and Wang, L. (2018). A study of the thermoelectric properties of benzo[1,2-b:4,5-b′]dithiophene–based donor–acceptor conjugated polymers. *Polym. Chem.* 9, 4440–4447. doi: 10.1039/C8PY00812D

Wang, S., Bohnsack, M., Megow, S., Renth, F., and Temps, F. (2019). Ultrafast excitation energy transfer in a benzimidazole-naphthopyran donor-acceptor dyad. *Phys. Chem. Chem. Phys.* 21, 2080–2092. doi: 10.1039/C8CP05054F

Wiebeler, C., Plasser, F., Hedley, G. J., Ruseckas, A., Samuel, I. D., and Schumacher, S. (2017). Ultrafast electronic energy transfer in an orthogonal molecular dyad. *J. Phys. Chem. Lett.* 8, 1086–1092. doi: 10.1021/acs.jpclett.7b00089

Wu, P., and Brand, L. (1994). Resonance energy transfer: methods and applications. *Anal. Biochem.* 218, 1–13. doi: 10.1006/abio.1994.1134

Yang, Y., Wang, K., Li, G., Ran, X., Song, X., Gasparini, N., et al. (2018). Fluorination triggered new small molecule donor materials for efficient as-cast organic solar cells. *Small* 14:e1801542. doi: 10.1002/smll.201801542

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 07 December 2020
doi: 10.3389/fchem.2020.587824

Slow Transition Path Times Reveal a Complex Folding Barrier in a Designed Protein

*Alexander Mehlich[1], Jie Fang[2], Benjamin Pelz[1], Hongbin Li[2] and Johannes Stigler[3]**

[1] Physics Department E22, Technische Universität München, Garching, Germany, [2] Department of Chemistry, University of British Columbia, Vancouver, BC, Canada, [3] Gene Center Munich, Ludwig-Maximilians-Universität München, Munich, Germany

De-novo designed proteins have received wide interest as potential platforms for nano-engineering and biomedicine. While much work is being done in the design of thermodynamically stable proteins, the folding process of artificially designed proteins is not well-studied. Here we used single-molecule force spectroscopy by optical tweezers to study the folding of ROSS, a *de-novo* designed 2x2 Rossmann fold. We measured a barrier crossing time in the millisecond range, much slower than what has been reported for other systems. While long transition times can be explained by barrier roughness or slow diffusion, we show that isotropic roughness cannot explain the measured transition path time distribution. Instead, this study shows that the slow barrier crossing of ROSS is caused by the population of three short-lived high-energy intermediates. In addition, we identify incomplete and off-pathway folding events with different barrier crossing dynamics. Our results hint at the presence of a complex transition barrier that may be a common feature of many artificially designed proteins.

Keywords: protein folding, transition path analysis, artificial protein, transition state barrier, roughness, Kramers rate theory

OPEN ACCESS

Edited by:
Pilar Cossio, University of Antioquia, Colombia

Reviewed by:
Rohit Satija, University of California, Berkeley, United States
Jae-Yeol Kim, National Institutes of Health (NIH), United States

****Correspondence:***
Johannes Stigler
stigler@genzentrum.lmu.de

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Chemistry

Received: *27 July 2020*
Accepted: *26 October 2020*
Published: *07 December 2020*

Citation:
Mehlich A, Fang J, Pelz B, Li H and Stigler J (2020) Slow Transition Path Times Reveal a Complex Folding Barrier in a Designed Protein. Front. Chem. 8:587824. doi: 10.3389/fchem.2020.587824

INTRODUCTION

The protein-folding problem has fascinated scientists for more than half a century (Anfinsen et al., 1961). To learn and understand more about the reaction that takes place when an unfolded polypeptide chain tries to "find" its correctly folded protein structure, the concept of energy landscapes is a powerful theoretical framework (Onuchic et al., 1997). Within this framework, three key parameters govern the typical timescales of an observed reaction: the height of a free energy barrier, which needs to be overcome, the curvature or stiffness at the top of that barrier, and the diffusion coefficient (Hänggi et al., 1990).

Experimentally, various techniques have been used to verify the applicability of energy landscape theory to protein folding. Besides single-molecule Förster resonance energy transfer (smFRET) (Borgia et al., 2012; Chung et al., 2012; Soranno et al., 2012; Chung and Eaton, 2013), another well-established tool to measure the folding of individual proteins is single-molecule force spectroscopy (SMFS), where force can act as both a denaturant and a readout of protein conformational changes (Rief et al., 1997; Junker et al., 2009; Gebhardt et al., 2010; Yu et al., 2012). In the case of optical tweezers, specific attachment strategies and stable setups have provided access to a large range of timescales and enabled measurements of single molecules over tens of minutes (Stigler et al., 2011; Rognoni et al., 2014) at a temporal resolution to resolve timescales

of only a few tens of microseconds (Žoldák et al., 2013; Neupane et al., 2016). Recent studies have shown that with this technique it is possible to extract and evaluate transition paths [i.e., the rare events of actual barrier crossing, which allow the direct estimation of diffusion coefficients or the determination of transition path velocities (Neupane et al., 2017, 2018)]. In addition, the theoretical foundation used to analyze and correctly interpret experimentally measured transition paths is constantly improving (Hummer, 2004; Cossio et al., 2015, 2018; Covino et al., 2019).

A complementary approach to understanding protein folding tackles the problem from the opposite end (i.e., by *de-novo* designing protein sequences that fold into the desired shape). Because of their far-reaching potential in pharmaceutical applications, these artificial proteins have received wide interest from science and industry (Kuhlman and Bradley, 2019). In general, the *de-novo* design of proteins is an optimization problem where candidate sequences of amino acids are scored according to an energy model and a target fold. The optimization often comes at a high computational cost, as different side-chain packing can yield the same general fold, but widely different energy scores. While designers have been very successful in creating thermodynamically stable proteins as building blocks, there is little known about how these proteins compare to natural counterparts in terms of folding or conformational dynamics.

In this study, we characterize the timescales of folding and barrier crossing of one of the first fully designed artificial proteins with a topology that is abundantly found in nature, the 2x2 Rossmann fold, which can serve as a scaffold for designed enzymes (Koga et al., 2012).

MATERIALS AND METHODS

Protein Expression and Sample Preparation

The sequence of the 2x2 Rossmann fold protein (ROSS) was based on the Di-II-10 sequence of Koga et al. (2012) where the part of the designed protein comprises 100 amino acids (bold in sequences below). We modified this sequence with additional n- and c-terminal cysteines and a c-terminal His_6-tag. The variant S49Cc lacked the n-terminal cysteine and harbored two cysteines at position 49 and the c-terminus. Protein expression and purification were performed as described previously (Fang et al., 2013).

ROSS: MACK **MLLYVLIISN DKKLIEEARK MAEKANLELR TVKTEDELKK YLEEFRKESQ NIKVLILVSN DEELDKAKEL AQKMEIDVRT RKVTSPDEAK RWIKEFSEEG** GSKCLE HHHHHH.

S49Cc: MASKGS **MLLYVLIISN DKKLIEEARK MAEKANLELR TVKTEDELKK YLEEFRKE*C*Q NIKVLILVSN DEELDKAKEL AQKMEIDVRT RKVTSPDEAK RWIKEFSEEG** GSSGKCLE HHHHHH.

Bead-DNA-protein-bead-DNA dumbbells were generated as described previously (Mehlich et al., 2015). In short, maleimide-modified oligonucleotides were attached to the cysteines of the purified protein via a disulfide bond to form a protein-oligonucleotide construct. Next, biotin or digoxigenin modified DNA handles equipped with a single-stranded overhang complementary to the maleimide-modified oligonucleotides on their 3′ end were hybridized to this protein-oligonucleotide. The resulting DNA-protein-DNA hybrid was then bound to streptavidin or anti-digoxigenin functionalized 1 μm diameter silica beads for optical tweezers measurements.

Optical Trap Setup and Measurement Modes

A custom-built dual optical trap setup with back-focal plane detection was used for all force spectroscopic measurements (von Hansen et al., 2012). The two measurement modes comprise either stretch-relax cycles in constant-velocity mode or measurements at constant trap distances in passive-mode.

Before recording, signals were filtered with an eighth order Butterworth filter with a 3 dB-frequency set at 100 kHz. Data were acquired at 200 kHz. The stiffness of each trap was typically set to about 0.2 pN/nm. Constant-velocity measurements were performed at 500 nm/s.

Polymer Models and Length Coordinates

To model the force compliance of the polymer linkers that connect the protein to the beads we used previously published polymer models (Bustamante et al., 1994; Wang et al., 1997). Upon stretching and while the protein remains folded, the mechanical response of a trapping construct is dominated by stretching of the DNA handles which were modeled using an extensible Worm-Like Chain (eWLC) model (Wang et al., 1997)

$$F(\xi_D) = \frac{k_B T}{p_D}\left(\frac{1}{4}\left(1-\frac{\xi_D}{L_D}+\frac{F}{K}\right)^{-2}-\frac{1}{4}+\frac{\xi_D}{L_D}-\frac{F}{K}\right) \quad (1)$$

with DNA extension ξ_D, force F, thermal energy $k_B T$, DNA persistence length p_D, DNA contour length L_D, and stretch modulus K.

The additional force-extension of unfolded protein was modeled using a Worm-Like Chain (WLC) model in series with the above response of the DNA handles;

$$F(\xi_p) = \frac{k_B T}{p_p}\left(\frac{1}{4}\left(1-\frac{\xi_p}{L_p}\right)^{-2}-\frac{1}{4}+\frac{\xi_p}{L_p}\right) \quad (2)$$

where ξ_p is the additional extension of the unfolded amino acid chain of the protein, p_p is the protein persistence length, and L_p is the protein contour length.

To fit our experimental data, p_p was fixed at 0.7 nm. Typical values for the DNA fits were $p_D \approx 31$ nm, $L_D \approx 364$ nm, and $K \approx 200$ pN. Experiments were conducted at a temperature of $T \approx 298$ K.

Instead of the molecular extension ξ_p, which depends on the applied force, we used the force-independent contour length of the unfolded polypeptide L_p as a reaction coordinate. At a trap distance d, omitting the comparatively small extension of the folded protein, the extension of the unfolded polypeptide is given

by $\xi_p = d - \frac{F}{k_c} - \xi_D(F)$, where $k_c = \left(k_1^{-1} + k_2^{-1}\right)^{-1}$ is the combined spring constant of the two beads and $x = \frac{F}{k_c}$ is the combined bead deflection (see **Supplementary Figure 1**). The contour length L_p was then obtained from ξ_p using Equation (2). This conversion allowed us to compare events at different forces in a more straightforward manner (Puchner et al., 2008).

Integrals over the polymer models were used to convert energy profiles at different force biases as described (Ramm et al., 2014).

Experimental Rate Constants and Lifetime Distributions

Rate constants for unfolding and refolding from constant-velocity experiments were determined using cycles with 500 nm/s pulling speed and applying the method by Oberbarnscheidt et al. (2009).

Folding/unfolding rate constants from passive-mode measurements were obtained from hidden Markov models (HMMs) as described previously (Geier et al., 2007; Stigler et al., 2011; Stigler and Rief, 2012). In brief, HMMs were used to determine the likeliest sequence of hidden states that describes the observed trajectory of bead deflections. The lifetimes for each detected state were well-described by a single exponential distribution. Transition rate constants between states i and j were then determined using $k_{ij} = N_{ij}/(\tau_i \cdot \sum_k N_{ik})$, where τ_i is the average lifetime of state i and N_{ij} is the number of detected transitions between states i and j.

Force-dependent folding/unfolding rate constants were fitted to a model (Equation (S6), see **Supplementary Information**) that incorporates the energetic contributions of the non-linear polymer linkers to yield transition state positions and extrapolate force-free transition rate constants (Schlierf et al., 2007; Stigler et al., 2011).

The Relation Between Rates and Shape Parameters of an Energy Landscape

A relation between the transition rate k_{ij} from state i to j and the shape of the barrier with a height $\Delta G^{i\mathrm{TS}}$ separating these two states is given by Kramers' rate theory in the Smoluchowski limit [i.e., the overdamped case where $(D \cdot \beta)^{-1} \gg \omega_{\mathrm{TS}}$ (Kramers, 1940; Hänggi et al., 1990)]:

$$k_{ij} = \frac{D \cdot \beta \cdot \omega_i \cdot \omega_{\mathrm{TS}}}{2\pi} \cdot \exp\left(-\beta \cdot \Delta G^{i\mathrm{TS}}\right) \quad (3)$$

Here, D is the diffusion coefficient, β is the inverse thermal energy ($\beta^{-1} = k_{\mathrm{B}}T$), and ω_i and ω_{TS} represent the oscillation frequencies around the minimum of state i and around the barrier top of the transition state which needs to be passed to reach state j. We used this relation for barrier height reconstruction or expected rate calculation under the assumption of symmetric and harmonic barrier shapes. This assumption implies that $\omega_i = \omega_{\mathrm{TS}}$ and $\omega_{\mathrm{TS}}^2 = \left|\partial_x^2 G\right| = 4 \cdot \frac{\Delta G^{i\mathrm{TS}}}{\left(\Delta x^{i\mathrm{TS}}\right)^2}$, where $x^{i\mathrm{TS}}$ is the distance between a state i and its adjoining transition state TS (Mehlich, 2018).

Deconvolution of Equilibrium Distributions

For deconvolution (i.e., the removal of thermal noise-broadening) of measured bead deflection distributions $P_{\mathrm{msmt}}(x)$, we used the same algorithm as described in Ramm et al. (2014). In brief, we numerically minimized the function

$$\mathcal{O} = \langle \left| \ln\left(\hat{P}_{\mathrm{prot}}(x) \otimes \Psi_x(x)\right) - \ln\left(P_{\mathrm{msmt}}(x)\right) \right| \rangle_x + \lambda \cdot \sum_i \left[\beta \cdot \partial_x^2 \hat{G}(x)|_{\hat{x}_i}\right]^2 \quad (4)$$

where $\otimes$ represents the convolution operator and $\Psi_x(x)$ is the force-dependent point-spread function (PSF) to estimate the true probability distribution $\hat{P}_{\mathrm{prot}}(x) = \exp\left(-\beta \cdot \hat{G}(x)\right)$. The only unknown quantity was the deconvolved energy landscape $\hat{G}(x)$ of the protein. The smoothing penalty parameter was set to $\lambda = 10^{-2}\mathrm{nm}^4$.

The shape of the PSF was determined from equilibrium thermodynamics calculations (see Ramm et al., 2014 for a detailed description). In brief, at a given trap distance d, the mechanical free energy of stretching the polymer linkers and displacing the beads from their traps is

$$\mathcal{H}(x) = \int_0^{d-x} F_{D+p}(\xi)\, d\xi + \tfrac{1}{2} k_c x^2, \quad (5)$$

where $F_{D+p}(\xi)$ is the inverse of $\xi_D(F) + \xi_p(F)$ (Equations (1), (2)). The theoretical PSF of deflection values for a given trap distance and a given unfolded contour length was then determined by

$$\Psi_x(x) = \frac{e^{-\beta\mathcal{H}(x)}}{\int e^{-\beta\mathcal{H}(x)} dx} \quad (6)$$

Because of the nonlinearity of the eWLC and WLC linkers, the theoretical PSF is asymmetric and well-approximated by a skewed Gaussian distribution (See Azzalini, 1985 for a description of the skewed Gaussian distribution). **Supplementary Figure 2** shows the theoretical PSF in contour-length space for the states N and U, at typical forces in passive-mode experiments and in stretch-relax experiments, which we parameterized by a scale parameter σ and a skewness parameter γ. Because the force is not constant in passive-mode experiments, $\sigma = \sigma(x)$ and $\gamma = \gamma(x)$ are position-dependent (Gebhardt et al., 2010). In our analysis, $\sigma(x)$ and $\gamma(x)$ were determined by linear interpolation between values obtained from fits to the distribution of the folded state and the unfolded state.

Determination of Transition Path Time

Hidden Markov models (HMM) were used to identify state positions, state occupancies, and dwell times for transitions between folded (N) and unfolded (U) states of ROSS as described elsewhere (Stigler et al., 2011). To assess the actual duration of these identified transitions, we applied a 3-state model where the two outer states were fixed at the identified N and U positions and a third obligatory intermediate state T was introduces and fixed right in the middle between N and U, similar to an approach

already used elsewhere (Chung et al., 2012; Sturzenegger et al., 2018). The dwell times measured for T were interpreted as the required transition path time to cross the barrier between the folded and unfolded state.

For validation, the HMM method was also applied with, instead of one central intermediate T, two intermediates at the positions of I_1 and I_2. Here, transition path times were derived as the sum of the dwell times spent within I_1 and I_2 along each transition between N and U. The two methods resulted in practically identical estimates of the transition path time (TPT) from passive-mode measurements ($\langle\tau^{I_1+I_2}\rangle = 1.6 \pm 0.1$ ms and $\langle\tau^{T}\rangle = 1.5 \pm 0.1$ ms). **Supplementary Figures 3A,B** shows a comparison between results obtained from both methods. We note that the HMM method does not strictly prohibit boundary re-crossing. When imposing absorbing boundaries, we found that the detected transition paths often excluded stretches that lie visibly on a path (**Supplementary Figure 3C**) and the corresponding TPTs were consequently underestimated. This underestimation was due to random bead fluctuation noise that did not reflect changes in the molecular length and was especially prominent at high bandwidth. Filtering to lower bandwidth suppressed this noise and resulted in TPTs that agreed with both the HMM method (see above) and the method of transition averaging (see below).

Alternatively, transition path times were determined by selecting individual transitions and fitting the high-bandwidth bead relaxation data to a sigmoidal equation $f(t) = A/\left(1+\exp\left(\frac{t-t_0}{\tau_s}\right)\right)$, to determine the time t_0 at which a transition occurred. The transitions were then aligned according to their t_0 values and averaged. The characteristic timescale τ_s of the transition was then obtained from a sigmoidal fit to the aligned and averaged data (**Supplementary Figures 4B,C**). To be able to compare the characteristic timescales of the obligatory intermediate model and the sigmoidal model, we performed simulations of transitions between two states with an obligatory intermediate in the center (**Supplementary Figure 4D**, inset). The duration of the intermediate dwell was chosen from an exponential distribution with a time constant τ. Mimicking experiments, the transitions were then aligned, averaged, and fitted to a sigmoidal equation to determine the characteristic timescale of the sigmoidal model τ_s. A conversion factor between the two timescales τ and τ_s was then obtained by repeating the simulation for different intermediate durations (**Supplementary Figure 4E**).

Transition Path Time Analysis

An approximation for the average transition path time $\langle\tau_{TP}\rangle$ which is required to cross a harmonic barrier of height $\Delta G^{\ddagger} > 2\ k_B T$ by one-dimensional diffusion is given by Chung et al. (2009):

$$\langle\tau_{TP}\rangle \approx \frac{\ln\left(2e^{\gamma_E}\cdot\beta\cdot\Delta G^{\ddagger}\right)}{\beta\cdot D\cdot\omega_{TS}^2} \quad (7)$$

where $\gamma_E \approx 0.577$ is Euler's constant.

The following model approximates the cumulative distribution function of transition path times over a harmonic barrier (Zhang et al., 2007; Chaudhury and Makarov, 2010):

$$P(\tau_{TP}) \approx \int_0^{\tau_{TP}} \frac{\omega_K\cdot\sqrt{\beta\cdot\Delta G^{\ddagger}}}{1-\mathrm{erf}\left(\sqrt{\beta\cdot\Delta G^{\ddagger}}\right)} \cdot \frac{\exp\left[-\beta\cdot\Delta G^{\ddagger}\cdot\coth\left(\omega_K\cdot\frac{\tau}{2}\right)\right]}{\sinh\left(\omega_K\cdot\frac{\tau}{2}\right)\cdot\sqrt{2\pi\cdot\sinh(\omega_K\cdot\tau)}} d\tau \quad (8)$$

where $\omega_K = \beta\cdot D\cdot\omega_{TS}^2$ and erf() represents the error function.

Off-rates from intermediates were determined by fitting dwell time distributions to a model that describes the escape rate with two determinant rates (Rief et al., 2000):

$$p(\tau_{1+2}) = \frac{k_1\cdot k_2}{k_1-k_2}\cdot\left[\exp\left(-k_2\cdot\tau_{1+2}\right) - \exp\left(-k_1\cdot\tau_{1+2}\right)\right] \quad (9)$$

where, in our use case k_1 and k_2 represent off-rates out of two adjoining and predominant high-energy intermediate states located along a transition path and τ_{1+2} reflects the combined dwell time of these two intermediates measured as an effective overall transition path time.

All diffusion coefficients reported are represented with contour length as a reaction coordinate.

Simulation of Transition Path Time Distributions

Simulations of transition path trajectories were performed by integrating the discrete-time Langevin equation in a 1D potential as described previously (Ramm et al., 2014). To gather statistics, transition path times were collected using a flux sampling scheme (Zhang et al., 2007; Laleman et al., 2017).

RESULTS

Low-Bandwidth Stretch-Relax Experiments of an Artificial Protein Suggest Two-State Behavior

To study the folding and unfolding transition path times of an artificially designed globular protein, we introduced terminal cysteines into the sequence of ROSS, a 100 amino acid (aa), fully artificially designed protein that adopts a 2x2 Rossmann fold (Koga et al., 2012) (**Figure 1A**, inset). We then fused the protein termini to DNA handles and assembled bead-DNA-protein-DNA-bead dumbbells for force spectroscopic measurements in a custom-built optical tweezer instrument.

To characterize the mechanical behavior of ROSS, we subjected the protein-DNA dumbbell to stretch-relax cycles at 500 nm/s (**Figure 1A**). In the majority of cycles, ROSS displays apparent two-state unfolding behavior with unfolding forces at ≈17 pN and a contour length increase of 34.7 ± 0.9 nm, in excellent agreement with an expected length of 34.5 nm for the unfolding of 100 aa. Upon relaxation, ROSS readily refolds at a force of ≈ 6 pN back to the native contour length of 0 nm. In addition to the described behavior, we also observed, in 23% of cycles, a population of unfolding events at reduced forces of ≈ 10 pN, with a similar contour length increase of 33.8 ± 0.9 nm. An unfolding force histogram (**Figure 1B**) shows the

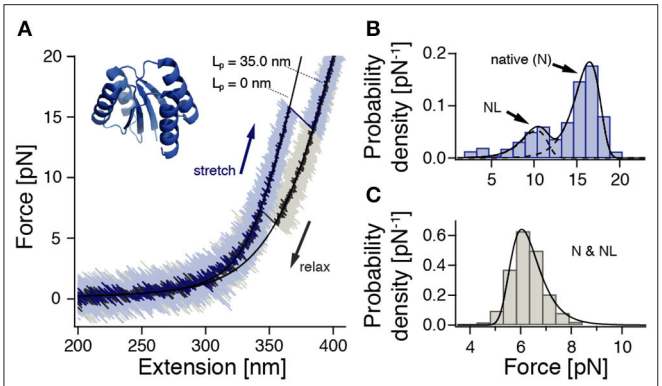

FIGURE 1 | ROSS shows apparent two-state folding with minor populations of a native-like (NL) configuration. Cf. **Supplementary Figure 10** for a zoom into the unfolding transition. **(A)** Stretch-relax cycle of ROSS recorded in a 500 nm/s constant-velocity experiment. Black lines are polymer model fits to the stretches of the folded and unfolded conformations. **(B)** Unfolding force distribution (500 nm/s, $n = 691$) with fits [Equation (S1)] to two distinct populations. **(C)** Refolding force distribution (500 nm/s, $n = 651$). Fit according to Equation (S2).

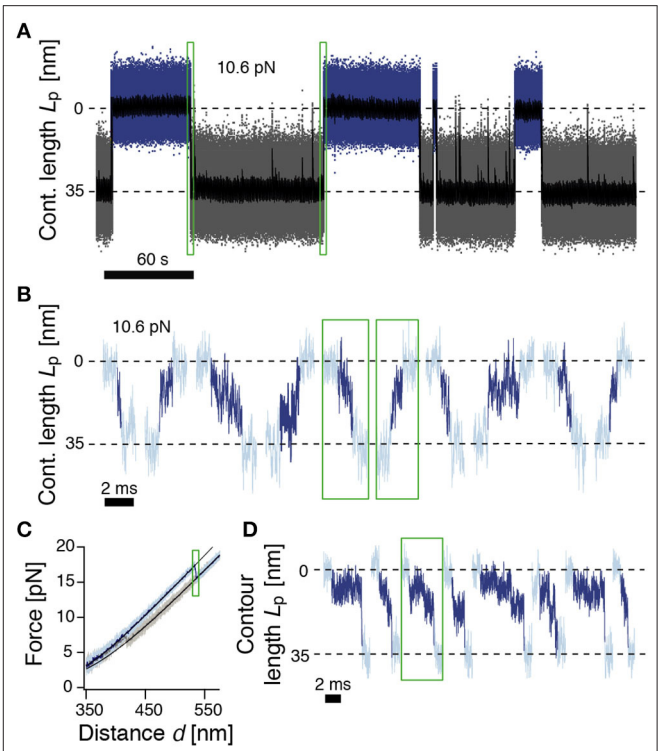

FIGURE 2 | ROSS shows slow transitions between the native and unfolded states. **(A)** Passive-mode trajectory of equilibrium fluctuations between the native (blue, 0 nm) and unfolded states (gray, 35 nm of unfolded contour length). Black: Filtered trace to guide the eye. The indicated force is acting on the native state. **(B)** Extracted transitions between the native and unfolded states from passive-mode experiments. Dark colors show the identified trajectories during barrier crossing. Green boxes highlight exemplary positions in trace (A). **(C)** Force-distance curve at a pulling speed of 500 nm/s. Black solid lines are polymer model fits to the folded and unfolded conformations. **(D)** Unfolding transitions from force-distance curves. Dark colors indicate trajectories during barrier crossing. The trajectory in the green box corresponds to the trajectory shown in (C).

occurrence of this second mechanically weaker state, which we call the native-like (NL) conformation. Refolding forces followed the expected behavior of a single pathway (**Figure 1C**). To obtain the equilibrium free energy difference between the native and unfolded states, we employed the Crooks fluctuation theorem (Crooks, 1999; Collin et al., 2005) (**Supplementary Figure 5**). We found a stability of the native state of $-27 \pm 2\ k_BT$, in good agreement with a reported value of $-25.2\ k_BT$ in chemical denaturation (Koga et al., 2012).

Transition Path Times Suggest a Rough Transition Barrier

We sought to investigate the transition path times (TPTs) for the unfolding and folding in greater detail. To this end, we held the tethered protein under tension in passive mode and followed the folding/unfolding transitions at high bandwidth over several minutes (**Figure 2A**). ROSS displayed long dwells in the native (N, 0 nm contour length) and unfolded (U, 34.7 nm contour length) states of up to a minute in duration. In addition, we also observed transient short-lived excursions from U toward the native structure, which we attribute to unsuccessful folding events. Using a model that incorporates the energetic contributions of linker stretching (see **Methods**) we determined a free energy difference between the N and U states of $-28 \pm 3\ k_BT$, confirming our earlier result from stretch-relax cycles ($-27 \pm 2\ k_BT$) and reported values from chemical denaturation ($-25.2\ k_BT$) (Koga et al., 2012).

We next took a closer look into the transitions between the N and U states (**Figure 2B**). To identify parts of the trajectory on a barrier-crossing transition path, we used hidden Markov modeling with a transient state between N and U and considered parts of the trajectory that were classified into this transient state as barrier crossing events. We observed a diverse ensemble of trajectories, ranging from 140 μs to 5.5 ms in duration. The average folding duration (1.4 ± 0.2 ms) and unfolding duration (1.5 ± 0.2 ms) were identical, as expected (Neupane et al., 2016).

Interestingly, the transitions were generally not unidirectional but showed frequent changes in direction and velocity as well as discernible pauses. Notably, the measured transition path times for ROSS were much slower than values obtained for a DNA hairpin (≈ 27 μs) (Neupane et al., 2016), the artificial protein α_3D (≈ 15 μs) (Chung and Eaton, 2013), or dimeric prion protein PrP (≈ 500 μs) (Neupane et al., 2016). The transition path duration of ROSS was also much slower than the predicted speed limit for the folding of a protein of 100 aa (≈ 1 μs) (Kubelka et al., 2004). To validate the resulting exceptionally slow barrier crossing time, we averaged aligned folding or unfolding transitions and fitted the resulting average relaxation response to a sigmoidal function (**Supplementary Figure 4**). The equivalent durations from this method were (0.7 ± 0.1 ms) for folding and (1.1 ± 0.1 ms) for unfolding, close to our previous results. Interestingly, transitions into non-native conformations were much faster and averaged ≈ 0.2 ms (**Supplementary Figure 4C**). Taken together, our observation of an exceptionally slow folding/unfolding time hint at the presence of a rugged barrier.

We suspected that the observed barrier crossing trajectories may be convoluted by thermal fluctuations of the beads that do not directly reflect changes in the protein end-to-end distance. We reasoned that this effect may be minimized at higher forces when the tension in the tether is higher and force fluctuations are thus minimized. To this end, we deliberately drove the system out of equilibrium with constant-velocity stretch-relax cycles and recorded the relaxation of the beads back to their equilibrium position at high bandwidth (**Figure 2C**). We note that during folding/unfolding, the trap position movement was ≈2 nm, much smaller than the observed bead relaxation amplitudes (≈15 nm). Hence, we do not anticipate significant bias from the active pulling protocol. **Figure 2D** shows representative relaxation trajectories of unfolding transitions that occurred at forces higher than 15 pN, thus fully excluding transitions that originate from the NL state. Contrary to expectations from a two-state model, and confirming our results from passive mode, the beads did not rapidly settle back into their relaxed position but displayed complicated trajectories for several milliseconds until relaxation, again hinting at a rugged barrier separating states N and U.

Transition Path Times of ROSS Depend on the Applied Force

The transition path times obtained from stretch-relax cycles probe the barrier at a different force bias and are thus not necessarily the same as those obtained from passive mode (Gladrow et al., 2019). Whereas, the N and U basins in our passive-mode experiments are approximately at the same energetic level, the U basin in typical stretch unfoldings lies ≈32 $k_B T$ lower than the N basin. We, therefore, investigated whether the transition path times at higher forces, where the equilibrium is shifted toward the unfolded state, differs from the situation at low forces.

In the case of a symmetric harmonic barrier of significant height, the average transition path time is $\langle \tau_{TP} \rangle \approx \frac{\ln(2e^{\gamma_E} \cdot \beta \cdot \Delta G^{\ddagger})}{\beta \cdot D \cdot \omega_{TS}^2}$ (Chung et al., 2009; Chaudhury and Makarov, 2010). Even though under increased force bias the approximation of a symmetric harmonic barrier no longer holds, the model predicts that $\langle \tau_{TP} \rangle$ is largely unaffected by a shift to higher forces due to its logarithmic dependency on the barrier height, a result that was also confirmed experimentally for a DNA hairpin (Neupane et al., 2016).

We collected a dataset of the transition path times of ROSS folding/unfolding from passive-mode experiments and stretch-relax cycles where transitions occurred in a range between 5 and 20 pN (**Figure 3A**, also see **Supplementary Figure 6** for an annotated and expanded version that also contains NL transitions). Surprisingly, we observed a clear correlation of increasing τ_{TP} with higher force ($r = 0.53 \pm 0.06$, 99 % *CI* from bootstrapping), in stark disagreement with predictions from a harmonic barrier model. N→U transitions that occurred at ≈17 pN were much slower (3.9 ± 0.2 ms) than passive-mode N→U transitions at ≈10 pN (1.5 ± 0.2 ms, see above) or the corresponding reverse reaction U→N at ≈6 pN (0.9 ± 0.1 ms) (**Supplementary Figure 6**). To understand the significance of this observation, we performed diffusion simulations in a 1D potential for the case of a smooth harmonic barrier and the case of a barrier harboring a weakly stable intermediate (**Supplementary Figure 7**). Corroborating our suspicion, $\langle \tau_{TP} \rangle$ was largely independent of force bias in the case of a harmonic barrier (**Supplementary Figure 7B**), but was force-dependent in the case of a barrier containing a weak intermediate (**Supplementary Figure 7D**). Together, these results confirmed our previous observation that ROSS folds over a rough barrier.

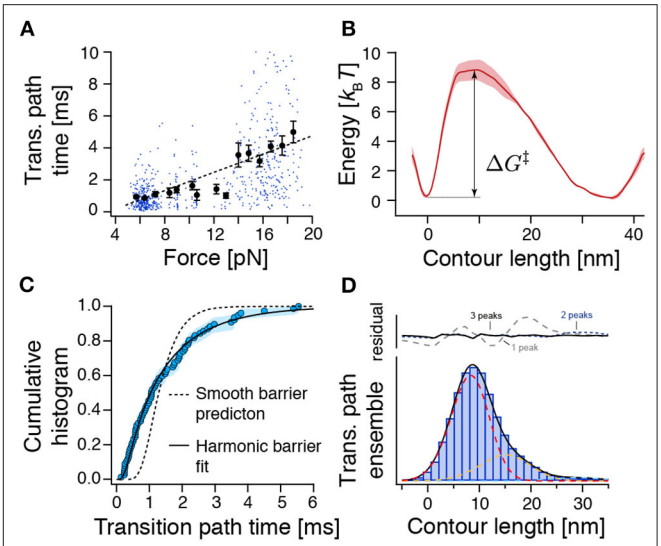

FIGURE 3 | The force-dependence and distribution of transition path times as well as the transition path ensemble show unexpected results for crossing a single and smooth barrier. **(A)** Transition path time (TPT) versus force plot for measured TPTs between N and U from both constant-velocity and passive-mode experiments ($n = 654$). While dots represent individual data points, points with error bars represent average transition path times within their respective force range. The dashed line is a linear fit to guide the eye. The related **Supplementary Figure 6** shows all detected transitions color coded according to their experimental origin. **(B)** Reconstructed free energy landscape from deconvolution of bead fluctuation histograms obtained from passive-mode experiments. **(C)** The distribution of experimental transition path times from passive-mode experiments (circles, $n = 77$) can be fitted to an approximation model for a central harmonic barrier [Equation (8), continuous line], albeit with unrealistic fit parameters (see text). Flux sampling simulations of one-dimensional diffusion over the deconvolved barrier shown in (B) predict a much narrower distribution (dashed line) that is incompatible with the experimental data ($p = 5 \times 10^{-6}$, KS test). **(D)** Position histogram of transition paths extracted from force-ramp experiments (bars, $n = 85$ transitions). Dashed lines represent fitted point-spread functions of three intermediates (see **Methods**). Black line: the sum of fitted point-spread functions. Upper panel: Fitting residuals scenarios of one, two, or three intermediates. $\Delta G^{\ddagger}$ represents the height of the energy barrier as in Equation (7).

Transition Path Times Are Incompatible With a Smooth Barrier

To get a more detailed picture of the folding barrier, we reconstructed the free energy landscape by deconvolving the bead fluctuation histograms obtained from passive-mode (constant-trap-separation) trajectories (Woodside et al., 2006; Gebhardt et al., 2010; Hinczewski et al., 2013; Ramm et al., 2014) (**Figure 3B**). However, because the time spent transitioning is an order of magnitude smaller than the time spent in the folded and unfolded states (≈0.01 %), the reconstructed energy landscape

emphasized the N and U basins but failed to pick up any deviations from a smooth barrier profile. We found a single asymmetric barrier with $\Delta G^{\ddagger} = 8.8 \pm 0.8\ k_B T$. Our measured barrier crossing time allowed us to estimate the diffusion constant of the reaction using (Hummer, 2004)

$$D = \frac{1}{\langle \tau_{TP} \rangle} \int_{x_A}^{x_B} e^{-\beta G(x)} \Phi_A(x)\, \Phi_B(x)\, dx \int_{x_A}^{x_B} e^{\beta G(x)} dx, \quad (10)$$

where Φ_A and Φ_B are committor functions (Hummer, 2004). We found $D = 10^{4.6\pm0.1}\mathrm{nm}^2/\mathrm{s}$. A comparable value ($D = 10^{4.3\pm0.1}\mathrm{nm}^2/\mathrm{s}$) was obtained when we used the approximation of a central harmonic barrier $D \approx \frac{\ln(2e^{\gamma_E}\cdot\beta\cdot\Delta G^{\ddagger})}{\beta\cdot\langle\tau_{TP}\rangle\cdot\omega^2}$ (Chung et al., 2009; Chaudhury and Makarov, 2010). An estimation of the diffusion constant based on the barrier height and the measured equilibrium transition rate constants between N and U using Kramers' theory (Kramers, 1940) ($D \approx 2\pi k \frac{e^{\Delta G^{\ddagger}}}{\beta\omega^2}$) yielded a similar value of $D = 10^{4.6\pm0.4}\ \mathrm{nm}^2/\mathrm{s}$. Notably, this experimental diffusion constant is several orders of magnitude lower than the value expected for a protein that folds at its speed limit ($\tau_{TP} \approx 1\ \mu s$, $D \approx 10^{7.7}\ \mathrm{nm}^2/\mathrm{s}$) (Kubelka et al., 2004). Slowed diffusion can be interpreted in terms of a rough energy landscape (Zwanzig, 1988) $D = D_0 e^{-\left(\frac{\varepsilon_{RMS}}{k_B T}\right)^2}$, where ε_{RMS} is a measure of the roughness of the energy profile. Based on our measurements, we estimate $\varepsilon_{RMS} \approx 2.7\ k_B T$, which amounts to a sizeable ≈10% of the folding free energy.

The previous analysis only considered the average transition path time $\langle\tau_{TP}\rangle$ and predicted a slow diffusion coefficient. We next also tested if the distribution of experimental transition path times is compatible with a one-dimensional diffusion model for the deconvolved barrier profile (**Figure 3C**). However, we found that the model (flux sampling calculation with $D = 10^{4.6}\mathrm{nm}^2/\mathrm{s}$, dashed line) predicts a much narrower distribution than the experimental data. An approximation model for a central harmonic barrier [Equation (8), continuous line] could be fitted to the data. Nevertheless, the obtained parameters of $D \approx 10^{3.8}\ \mathrm{nm}^2/\mathrm{s}$ and $\Delta G^{\ddagger} \approx 0.2\ k_B T$ were incompatible with the smooth-barrier estimations obtained earlier ($D = 10^{4.6\pm0.1}\mathrm{nm}^2/\mathrm{s}$, $\Delta G^{\ddagger} = 8.8 \pm 0.8\ k_B T$). Notably, a similar discrepancy has been observed also for PrP (Neupane et al., 2016). We conclude that the smooth barrier obtained from deconvolution (**Figure 3B**) is incompatible with the wide distribution of observed transition path times.

Weak Transient Intermediates Cause Slow Transition Paths

To understand the extraordinarily slow TPTs of ROSS, we took a closer look at the ensemble of trajectories during barrier crossing $p(x|\mathrm{TP})$. **Figure 3D** shows a position histogram of transition paths obtained from force-ramp experiments as shown in **Figure 2D**. Interestingly, the measured histogram was peaked at ≈9 nm, which is unexpected for a symmetric barrier (Hummer, 2004) but is in good agreement with the location of the identified barrier position from deconvolution (**Figure 3B**). The same indication of an asymmetric barrier was found in passive-mode experiments (**Supplementary Figure 8**).

We suspected that the protein populates weakly metastable on-pathway intermediates that slow down the overall transition and thus give rise to the notion of a rough energy landscape. We reasoned that the population of intermediates may produce elongated pauses at certain positions along the reaction coordinate of unfolded contour length that should be visible in transition path ensembles. Indeed, when comparing $p(x|\mathrm{TP})$ with the point-spread function (PSF) for intermediates, we noticed that at least three intermediates are necessary to fit the experimental $p(x|\mathrm{TP})$ (**Figure 3D**). While it is technically not possible to assign thermodynamically correct population densities to the intermediates, because of the non-equilibrium way the transition paths to produce $p(x|\mathrm{TP})$ are selected, the described procedure nevertheless clearly hints at the population of intermediates.

While intermediates were hidden in deconvolution of the equilibrium distribution (**Figure 3B**), the intermediates can be readily seen when the deconvolution procedure was applied to the transition path ensemble $p(x|\mathrm{TP})$ (**Supplementary Figure 9**), where the selection emphasizes parts of the trajectory that cross the barrier. However, because of this selection, the obtained probability distribution cannot be Boltzmann inverted to obtain a reconstruction of the barrier.

Reconstruction of the Energy Barrier

We, therefore, reverted to approximating the energy barrier based on force-dependent transition rate constants between N and U. WLC-model fits to high-force unfolding trajectories revealed the positions of three high-energy intermediates I_1-I_3 (**Figures 4A,B**, **Supplementary Figure 10**). We obtained $L_p(I_1) = 8.3 \pm 1.1$ nm, $L_p(I_2) = 17.6 \pm 1.9$ nm, $L_p(I_3) = 26.5 \pm 2.1$ nm, where $L_p(N) = 0$ nm.

To identify the position of the dominant transition barrier we determined the force-dependent rate constants for folding and unfolding (**Figure 4C**), where we used both data from passive-mode and stretch-relax experiments, and fitted a model that incorporates the energetic contributions of all compliances in the system (Schlierf et al., 2007). Notably, this model assumes that transition state positions are independent of the applied force. Therefore, kinks in the force-dependent rate plots or corresponding deviations from the model indicate a switch of the dominant transition barrier (Schlierf et al., 2010; Rico et al., 2019). Based on unfolding and refolding events observed during constant-velocity measurements, we identified three force ranges where extracted rates could be fitted locally (**Figure 4C**, continuous lines). In addition, rate constants extracted from passive-mode experiments suggest that there must be at least one more force range with a different predominant transition state since these rates are substantially lower than any of the three fits would suggest. Taken together, this yields four required transition states to interpret our data, in agreement with our finding of three intermediate states I_1, I_2, and I_3 between N and U. The force at which the TS_3:TS_4 transition state switch occurs, lies at around 6.3 pN (force of U), which is directly given by the intersection of the respective rate fits. The transition state switches TS_2:TS_3 and TS_1:TS_2 occur around 7.5 pN and 13 pN, respectively. The slopes from the fitted rate constants indicate transition state

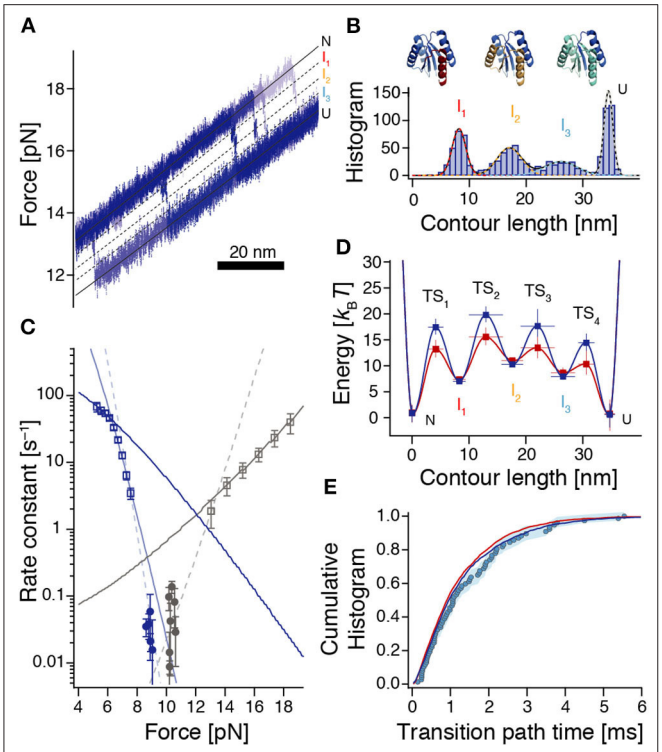

FIGURE 4 | Reconstruction of the rough energy landscape of ROSS. **(A)** Example of six high-resolution unfolding trajectories from constant-velocity experiments used to reveal positions of high-energy intermediates. Black lines indicate the corresponding positions. **(B)** Contour length gain histogram based on WLC-fits to unfolding events from force-ramp experiments ($n = 1010$ from 304 unfolding events). Colored dashed lines are fits to the individual peaks resulting from a Gaussian multi-peak fit to the entire histogram. Top: Assignment of unfolded structural elements of ROSS to the intermediates. **(C)** Force-dependent rates for folding (dark blue) and unfolding (gray) of ROSS extracted from constant-velocity (squares) and passive-mode experiments (circles). Continuous lines are local fits [Equation (S6)] to the data. Dashed lines interpolate between the accessible force ranges of constant-velocity and passive-mode experiments and illustrate the notion of switches of the dominant transition barrier in different force regimes. **(D)** Reconstructed energy landscape for ROSS. For the dark blue landscape, we assumed that folding transitions are fundamentally limited by speed limit protein folding ($D \approx 10^{7.7}$ nm^2/s); for the dark red landscape, we assumed that the speed of our observation is fundamentally limited by the bead diffusion in our optical trap ($D \approx 10^6$ nm^2/s). See **Supplementary Tables 1, 2** for parameters of the energy landscapes. **(E)** The experimentally derived transition path time distribution (circles, same data as in **Figure 3C**) is well-reproduced by the distributions predicted from the energy landscapes shown in (D) [dark blue ($p = 0.22$, KS test) and red ($p = 0.12$, KS test) lines, respectively].

positions $TS_1 = 3.1 \pm 0.3$ nm between N and I_1, $TS_3 = 19.4 \pm 1.1$ nm between I_2 and I_3, as well as $TS_4 = 30.5 \pm 1.3$ nm between I_3 and U. Due to the very slow folding kinetics and the narrow accessible force range for measurements in passive-mode, no position could directly be derived from our data for TS_2. However, interpolations between rate constants obtained from stretch-relax cycles and passive-mode experiments suggest that the TS_2 position is situated at $\approx 10 - 12$ nm which is right between I_1 and I_2 (**Figure 4C**, dashed fits).

To obtain values for the energies of the intermediates and transition states, we employed a piecewise reconstruction method where we determined individual effective barrier heights for each dominant barrier position [i.e., at different force biases, based on the measured dwell time distribution and Kramers' rate equation (Hänggi et al., 1990, **Supporting Methods**)]. Reasoning that the experimentally determined transition path times are likely dominated by metastable dwells in high-energy intermediates, we fitted the transition path time distribution to a model where the barrier crossing time is determined by the sequential population of the three intermediates I_1, I_2, and I_3. The corresponding rate constants for the escape rates from I_1, I_2, and I_3 were then translated into barrier heights using Kramers' rate equation, where we assigned the longest dwell to I_1 and the shortest dwell to I_3, in line with our observations (**Figures 2B,D, 3D**). Finally, we used the experimentally determined rate constants between N and U and the switch of the dominant transition barrier at different force bias to determine the missing barrier heights (**Supplementary Figure 11** and **Supporting Methods**).

All reconstruction steps depend on an assumption of the diffusion constant D. Since bumps in the energy profile slow down the observed transitions, our experiments imply that a realistic intrinsic D must be higher than the smooth barrier estimation of $D \approx 10^{4.6}$ nm^2/s. Here we did the reconstruction for two cases: The assumption that folding is fundamentally limited by the folding speed limit ($D \approx 10^{7.7}$ nm^2/s) and the assumption that the speed of our observations is fundamentally limited by the diffusion of the beads in the optical trap ($D \approx 10^6$ nm^2/s). The resulting barrier reconstructions mostly only differ in the barrier heights and are shown in **Figure 4D**.

A Barrier Harboring High-Energy Intermediates Agrees With Pulling Variants and Reproduces Experimental Transition Path Time Distributions

To verify our barrier reconstruction, we again used flux sampling simulations to determine the distribution of transition path times, based on the reconstructions shown in **Figure 4D**. In contrast to the smooth barrier case, this barrier profile yielded a much wider transition path time distribution that matched the experimental data very well (**Figure 4E**). In addition, we also performed Langevin dynamics simulations within the energy landscapes of **Figure 4D** to show that the ensembles of transition paths predicted from the reconstructions agree with experimental distributions (**Supplementary Figure 12**).

Corroborating evidence for the validity of our energy landscape reconstruction comes from a directional pulling mutant of ROSS, S49Cc, where force was only applied between residue 49 and the c-terminus. This variant also harbored an intermediate, which matched the I_1 intermediate of ROSS both in length (8.0 ± 0.9 nm) and in energy (-15 ± 2 k_BT) (ROSS: 8.3 ± 1.1 nm and -15 ± 3 k_BT, respectively; see **Supplementary Figures 13, 14**). In addition, the obligatory intermediate of S49Cc attains its maximal population at ≈13 pN, with an average dwell time of ≈2 ms. This closely resembles the properties of the corresponding I_1 intermediate of ROSS,

which also reaches its maximum population at ≈13 pN, with a comparable lifetime (**Supplementary Figure 15**). Taken together, these results suggest that typical ROSS unfolding occurs via the unfolding of the c-terminal half of the protein, where the length of the first intermediate is compatible with the unraveling of the c-terminal helix.

As shown earlier, metastable intermediates in the barrier may cause a shift in the average transition path time with varying force (cf. **Supplementary Figure 7**). We determined the predicted variation of $\langle \tau_{TP} \rangle$, based on the energy landscapes of **Figure 4D** analytically and with simulations. While both energy profiles correctly predicted an increase of $\langle \tau_{TP} \rangle$ with increasing force, both also predict a subsequent decrease above about 13 pN (**Supplementary Figure 15**). While this behavior was directly observed in the variant S49Cc (**Supplementary Figure 14C**), the same trend reversal was not apparent in experimental data of ROSS (**Figure 3A**), suggesting that perhaps, a one-dimensional description of the energy barrier may not be adequate for ROSS, or the non-equilibrium protocol used for obtaining transition path times at high forces may probe different transition paths (Gladrow et al., 2019).

DISCUSSION

In summary, ROSS exhibits remarkably slow transitions over the barrier between the folded and unfolded states, much slower than the transitions of other systems observed in optical tweezers (Neupane et al., 2016; Hoffer et al., 2019). FRET experiments generally have yielded even faster transition path times (Chung et al., 2012; Chung and Eaton, 2013). However, timescales are not directly comparable between different experimental techniques, because of the time-limiting effect of the measurement apparatus (Cossio et al., 2015, 2018; De Sancho et al., 2018). The best-characterized system in optical tweezers, DNA hairpins, generally displays much faster transitions than proteins (Neupane et al., 2016), likely because their transitions, owing to the experimental unzipping geometry, are well-described by a one-dimensional diffusion model (Neupane et al., 2012).

Slow transition path times have been discussed in the context of internal friction, frustration, or energy profile roughness (Liu et al., 2009; Wensley et al., 2010; Borgia et al., 2012; Chung et al., 2015), but our analysis revealed that none of these can explain the wide transition path time distribution of ROSS. Instead, we found evidence of a series of on-pathway intermediates that are briefly populated during barrier crossing. A similar mechanism has been observed for the coupled folding and binding of IDPs (Sturzenegger et al., 2018).

The slow transition path times of ROSS raise the interesting possibility that its impeded folding may be a consequence of its origin in artificial design. Indeed, most design algorithms, such as Rosetta optimize an equilibrium low energy state of the native target structure, compared to non-native interactions (Koga et al., 2012), but not the pathway of folding. The high-energy intermediates observed in our measurements may be caused by improper packing of individual alpha or beta secondary structures. The population of these non-native contacts, especially under denaturing conditions, may effectively hinder efficient folding. On the other hand, high-energy intermediates may also act as "checkpoints" in the folding landscape that can aid folding along a specific pathway, at the expense of overall cooperativity.

Notably, Rosetta structure prediction simulations indicate that ROSS may indeed adopt a non-native fold that is energetically very close to the native design structure (Koga et al., 2012). It is unclear if we observe this conformation in our experiments. However, possible candidates are the NL states and fast probing events observed in both variants. Interestingly, the fraction of NL states in the pulling variant S49Cc, where force is only applied across the c-terminal half, is significantly reduced from ≈23 to ≈7% ($p < 0.01$, binomial test), suggesting that NL states may be caused by non-native interactions between the n-terminal and the c-terminal halves.

Interestingly, the extremely slow transition path time of ROSS was only apparent in transitions between the native and unfolded conformations. Other transitions, such as incomplete folding events (**Supplementary Figure 4C**), were faster by almost an order of magnitude. In the context of a one-dimensional energy landscape, these probing events may correspond to partial barrier crossings from U to I_1. Indeed, our model predicts that such a partial crossing of the barrier lasts ≈150 μs, close to experimental values. However, this scenario fails to explain why under passive-mode conditions, there appear to be more probing events than successful folding transitions, requiring that the barrier TS_1 must be higher than TS_2. Therefore, an alternative explanation may be that the energy landscape of ROSS is multi-dimensional and our reconstruction only captures a projection of the true energy landscape.

A scenario of multidimensionality may result in position-dependent diffusion along the projected experimentally accessible energy landscape and, under certain circumstances, also result in a distribution of TPTs that is compatible with experiments. **Supplementary Figure 16** illustrates this at the example of the smooth barrier of **Figure 3B**, where the diffusion coefficient was set to $D = 10^{7.7}$ nm^2/s everywhere except in a close neighborhood around I_1 where we set $D = 10^{3.4}$ nm^2/s. However, while this model can also reproduce the experimental TPT distribution (**Supplementary Figure 16B**), it fails to explain the experimentally observed force dependence of $\langle \tau_{TP} \rangle$ (**Supplementary Figure 16C**). While it may be possible to build a position-dependent diffusivity model that also reproduces a force-dependent $\langle \tau_{TP} \rangle$ (e.g., by requiring a force-dependent change of D), the basis for this is rather speculative. On the contrary, the reconstructions of **Figure 4D** reproduce both the distribution and force-dependence of TPTs, without the need for additional assumptions.

Finally, our results allow us to infer a structural interpretation of the rearrangements during barrier crossing (**Figure 4B**, top). Unfolding of ROSS starts via I_1 which can be attributed to the unfolding of the c-terminal α-helix., while the proximate β-sheet remains bound to the hydrophobic core. After I_1, unfolding continues via I_2 which can be attributed to the unfolding of

the first three structural elements αβα starting from the c-terminus. The last and very weak intermediate I_3 most likely represents a conformation where only the n-terminal β-sheet and the successive α-helix remain folded, a possible folding seed for ROSS.

In summary, here we have presented a single-molecule folding study of the artificially designed protein ROSS. The protein showed exceptionally slow transition path times that are incompatible with a smooth transition barrier but could be explained by the presence of high-energy intermediates. Our results illustrate that high-energy folding intermediates slow the barrier crossing of all reaction systems and raise the question of whether uncharacteristically slow transition path times may be a trait of many protein structures that have been optimized for folding stability instead of folding speed (Basak et al., 2019).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AM conducted experiments. JS and AM analyzed data. BP, JF, and HL contributed reagents and methods. JS conceived of the study and wrote the paper with input from all authors. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The gene for 2x2 ROSS was a kind gift by David Baker. JS acknowledges the support by the LMU Center for Nanoscience CeNS and a DFG Emmy Noether grant STI673/2-1. HL acknowledges the support of an Alexander von Humboldt fellowship.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2020.587824/full#supplementary-material

REFERENCES

Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *PNAS* 47, 1309–1314. doi: 10.1073/pnas.47.9.1309

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* 12, 171–178.

Basak, S., Nobrega, R. P., Tavella, D., Deveau, L. M., Koga, N., Tatsumi-Koga, R., et al. (2019). Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed βα protein. *PNAS* 116, 6806–6811. doi: 10.1073/pnas.1818744116

Borgia, A., Wensley, B. G., Soranno, A., Nettels, D., Borgia, M. B., Hoffmann, A., et al. (2012). Localizing internal friction along the reaction coordinate of protein folding by combining ensemble and single-molecule fluorescence spectroscopy. *Nat. Commun.* 3, 1195–1199. doi: 10.1038/ncomms2204

Bustamante, C., Marko, J. F., Siggia, E. D., and Smith, S. (1994). Entropic elasticity of lambda-phage DNA. *Science* 265, 1599–1600. doi: 10.1126/science.8079175

Chaudhury, S., and Makarov, D. E. (2010). A harmonic transition state approximation for the duration of reactive events in complex molecular rearrangements. *J. Chem. Phys.* 133, 034118–034112. doi: 10.1063/1.3459058

Chung, H. S., and Eaton, W. A. (2013). Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* 502, 685–688. doi: 10.1038/nature12649

Chung, H. S., Louis, J. M., and Eaton, W. A. (2009). Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *PNAS* 106, 11837–11844. doi: 10.1073/pnas.0901178106

Chung, H. S., McHale, K., Louis, J. M., and Eaton, W. A. (2012). Single-molecule fluorescence experiments determine protein folding transition path times. *Science* 335, 981–984. doi: 10.1126/science.1215768

Chung, H. S., Piana-Agostinetti, S., Shaw, D. E., and Eaton, W. A. (2015). Structural origin of slow diffusion in protein folding. *Science* 349, 1504–1510. doi: 10.1126/science.aab1369

Collin, D., Ritort, F., Jarzynski, C., Smith, S. B., Tinoco, I., and Bustamante, C. (2005). Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* 437, 231–234. doi: 10.1038/nature04061

Cossio, P., Hummer, G., and Szabo, A. (2015). On artifacts in single-molecule force spectroscopy. *PNAS* 112, 14248–14253. doi: 10.1073/pnas.1519633112

Cossio, P., Hummer, G., and Szabo, A. (2018). Transition paths in single-molecule force spectroscopy. *J. Chem. Phys.* 148, 123309–123311. doi: 10.1063/1.5004767

Covino, R., Woodside, M. T., Hummer, G., Szabo, A., and Cossio, P. (2019). Molecular free energy profiles from force spectroscopy experiments by inversion of observed committors. *J. Chem. Phys.* 151:154115. doi: 10.1063/1.5118362

Crooks, G. E. (1999). Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E.* 60, 2721–2726. doi: 10.1103/PhysRevE.60.2721

De Sancho, D., Schönfelder, J., Best, R. B., Perez-Jimenez, R., and Muñoz, V. (2018). Instrumental effects in the dynamics of an ultrafast folding protein under mechanical force. *J. Phys. Chem. B* 122, 11147–11154. doi: 10.1021/acs.jpcb.8b05975

Fang, J., Mehlich, A., Koga, N., Huang, J., Koga, R., Gao, X., et al. (2013). Forced protein unfolding leads to highly elastic and tough protein hydrogels. *Nat. Commun.* 4:2974. doi: 10.1038/ncomms3974

Gebhardt, J. C., Bornschlögl, T., and Rief, M. (2010). Full distance-resolved folding energy landscape of one single protein molecule. *PNAS* 107, 2013–2018. doi: 10.1073/pnas.0909854107

Geier, F., Fengos, G., Felizzi, F., and Iber, D. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. 3rd ed. Cambridge: Cambridge University Press.

Gladrow, J., Ribezzi-Crivellari, M., Ritort, F., and Keyser, U. F. (2019). Experimental evidence of symmetry breaking of transition-path times. *Nat. Commun.* 10:55. doi: 10.1038/s41467-018-07873-9

Hänggi, P., Talkner, P., and Borkovec, M. (1990). Reaction-rate theory: fifty years after Kramers. *Rev. Modern Phys.* 62:251. doi: 10.1103/RevModPhys.62.251

Hinczewski, M., Gebhardt, J. C., Rief, M., and Thirumalai, D. (2013). From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *PNAS* 110, 4500–4505. doi: 10.1073/pnas.1214051110

Hoffer, N. Q., Neupane, K., Pyo, A. G. T., and Woodside, M. T. (2019). Measuring the average shape of transition paths during the folding of a single biological molecule. *PNAS* 116, 8125–8130. doi: 10.1073/pnas.1816602116

Hummer, G. (2004). From transition paths to transition states and rate coefficients. *J. Chem. Phys.* 120, 516–523. doi: 10.1063/1.1630572

Junker, J. P., Ziegler, F., and Rief, M. (2009). Ligand-dependent equilibrium fluctuations of single calmodulin molecules. *Science* 323, 633–637. doi: 10.1126/science.1166191

Koga, N., Tatsumi-Koga, R., Liu, G., Xiao, R., Acton, T. B., Montelione, G. T., et al. (2012). Principles for designing ideal protein structures. *Nature* 491, 222–227. doi: 10.1038/nature11600

Kramers, H. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* 7, 284–304. doi: 10.1016/S0031-8914(40)90098-2

Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004). The protein folding "speed limit." *Curr. Opin. Struct. Biol.* 14, 76–88. doi: 10.1016/j.sbi.2004.01.013

Kuhlman, B., and Bradley, P. (2019). Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* 20, 681–697. doi: 10.1038/s41580-019-0163-x

Laleman, M., Carlon, E., and Orland, H. (2017). Transition path time distributions. *J. Chem. Phys.* 147, 214103–214110. doi: 10.1063/1.5000423

Liu, F., Nakaema, M., and Gruebele, M. (2009). The transition state transit time of, W. W. domain folding is controlled by energy landscape roughness. *J. Chem. Phys.* 131, 195101–195110. doi: 10.1063/1.3262489

Mehlich, A. (2018). *Transition Paths of Protein-Folding Probed With Optical Tweezers* (Aachen: Shaker Verlag).

Mehlich, A., Austen, K., Ringer, P., Grashoff, C., and Rief, M. (2015). Evaluation of molecular tension sensors using single-molecule force spectroscopy and live cell FRET imaging. *Protocol Exchange*. doi: 10.1038/protex.2015.095. [Epub ahead of print].

Neupane, K., Foster, D. A., Dee, D. R., Yu, H., Wang, F., and Woodside, M. T. (2016). Direct observation of transition paths during the folding of proteins and nucleic acids. *Science* 352, 239–242. doi: 10.1126/science.aad0637

Neupane, K., Hoffer, N. Q., and Woodside, M. T. (2018). Measuring the local velocity along transition paths during the folding of single biological molecules. *Phys. Rev. Lett.* 121:018102. doi: 10.1103/PhysRevLett.121.018102

Neupane, K., Ritchie, D. B., Yu, H., Foster, D. A., Wang, F., and Woodside, M. T. (2012). Transition path times for nucleic Acid folding determined from energy-landscape analysis of single-molecule trajectories. *Phys. Rev. Lett.* 109:068102. doi: 10.1103/PhysRevLett.109.068102

Neupane, K., Wang, F., and Woodside, M. T. (2017). Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation. *PNAS* 114, 1329–1334. doi: 10.1073/pnas.1611602114

Oberbarnscheidt, L., Janissen, R., and Oesterhelt, F. (2009). Direct and model free calculation of force-dependent dissociation rates from force spectroscopic data. *Biophys. J.* 97, L19–L21. doi: 10.1016/j.bpj.2009.08.015

Onuchic, J. N., Luthey-Schulten, Z., and Wolynes, P. G. (1997). Theory of protein folding: the energy landscape perspective. *Ann. Rev. Phys. Chem.* 48, 545–600. doi: 10.1146/annurev.physchem.48.1.545

Puchner, E. M., Franzen, G., Gautel, M., and Gaub, H. E. (2008). Comparing proteins by their unfolding pattern. *Biophys. J.* 95, 426–434. doi: 10.1529/biophysj.108.129999

Ramm, B., Stigler, J., Hinczewski, M., Thirumalai, D., Herrmann, H., Woehlke, G., et al. (2014). Sequence-resolved free energy profiles of stress-bearing vimentin intermediate filaments. *PNAS* 111, 11359–11364. doi: 10.1073/pnas.1403122111

Rico, F., Russek, A., González, L., Grubmüller, H., and Scheuring, S. (2019). Heterogeneous and rate-dependent streptavidin-biotin unbinding revealed by high-speed force spectroscopy and atomistic simulations. *PNAS* 116, 6594–6601. doi: 10.1073/pnas.1816909116

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M., and Gaub, H. E. (1997). Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276, 1109–1112. doi: 10.1126/science.276.5315.1109

Rief, M., Rock, R. S., Mehta, A. D., Mooseker, M. S., Cheney, R. E., and Spudich, J. A. (2000). Myosin-V stepping kinetics: a molecular model for processivity. *PNAS* 97, 9482–9486. doi: 10.1073/pnas.97.17.9482

Rognoni, L., Möst, T., Zoldák, G., and Rief, M. (2014). Force-dependent isomerization kinetics of a highly conserved proline switch modulates the mechanosensing region of filamin. *PNAS* 111, 5568–5573. doi: 10.1073/pnas.1319448111

Schlierf, M., Berkemeier, F., and Rief, M. (2007). Direct observation of active protein folding using lock-in force spectroscopy. *Biophys. J.* 93, 3989–3998. doi: 10.1529/biophysj.107.114397

Schlierf, M., Yew, Z. T., Rief, M., and Paci, E. (2010). Complex unfolding kinetics of single-domain proteins in the presence of force. *Biophys. J.* 99, 1620–1627. doi: 10.1016/j.bpj.2010.06.039

Soranno, A., Buchli, B., Nettels, D., Cheng, R. R., Müller-Späth, S., Pfeil, S. H., et al. (2012). Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *PNAS* 109, 17800–17806. doi: 10.1073/pnas.1117368109

Stigler, J., and Rief, M. (2012). Hidden markov analysis of trajectories in single-molecule experiments and the effects of missed events. *Chemphyschem* 13, 1079–1086. doi: 10.1002/cphc.201100814

Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C., and Rief, M. (2011). The complex folding network of single calmodulin molecules. *Science* 334, 512–516. doi: 10.1126/science.1207598

Sturzenegger, F., Zosel, F., Holmstrom, E. D., Buholzer, K. J., Makarov, D. E., Nettels, D., et al. (2018). Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat. Commun.* 9:4708. doi: 10.1038/s41467-018-07043-x

von Hansen, Y., Mehlich, A., Pelz, B., Rief, M., and Netz, R. R. (2012). Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using Bayesian inference. *Rev. Sci. Instr.* 83:095116. doi: 10.1063/1.4753917

Wang, M. D., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1997). Stretching DNA with optical tweezers. *Biophys. J.* 72, 1335–1346. doi: 10.1016/S0006-3495(97)78780-0

Wensley, B. G., Batey, S., Bone, F. A., Chan, Z. M., Tumelty, N. R., Steward, A., et al. (2010). Experimental evidence for a frustrated energy landscape in a three-helix-bundle protein family. *Nature* 463, 685–688. doi: 10.1038/nature08743

Woodside, M. T., Anthony, P. C., Behnke-Parks, W. M., Larizadeh, K., Herschlag, D., and Block, S. M. (2006). Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* 314:1001. doi: 10.1126/science.1133601

Yu, H., Liu, X., Neupane, K., Gupta, A. N., Brigley, A. M., Solanki, A., et al. (2012). Direct observation of multiple misfolding pathways in a single prion protein molecule. *PNAS* 109, 5283–5288. doi: 10.1073/pnas.1107736109

Zhang, B. W., Jasnow, D., and Zuckerman, D. M. (2007). Transition-event durations in one-dimensional activated processes. *J. Chem. Phys.* 126, 074504–074519. doi: 10.1063/1.2434966

Žoldák, G., Stigler, J., Pelz, B., Li, H., and Rief, M. (2013). Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy. *PNAS* 110, 18156–18161. doi: 10.1073/pnas.1311495110

Zwanzig, R. (1988). Diffusion in a rough potential. *PNAS* 85, 2029–2030. doi: 10.1073/pnas.85.7.2029

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Molecular Biosciences

MINI REVIEW
published: 13 January 2021
doi: 10.3389/fmolb.2020.607158

Transcriptional Riboswitches Integrate Timescales for Bacterial Gene Expression Control

*Catherine E. Scull, Shiba S. Dandpat, Rosa A. Romero and Nils G. Walter**

Department of Chemistry, Single Molecule Analysis Group and Center for RNA Biomedicine, University of Michigan, Ann Arbor, MI, United States

Transcriptional riboswitches involve RNA aptamers that are typically found in the 5′ untranslated regions (UTRs) of bacterial mRNAs and form alternative secondary structures upon binding to cognate ligands. Alteration of the riboswitch's secondary structure results in perturbations of an adjacent expression platform that controls transcription elongation and termination, thus turning downstream gene expression "on" or "off." Riboswitch ligands are typically small metabolites, divalent cations, anions, signaling molecules, or other RNAs, and can be part of larger signaling cascades. The interconnectedness of ligand binding, RNA folding, RNA transcription, and gene expression empowers riboswitches to integrate cellular processes and environmental conditions across multiple timescales. For a successful response to an environmental cue that may determine a bacterium's chance of survival, a coordinated coupling of timescales from microseconds to minutes must be achieved. This review focuses on recent advances in our understanding of how riboswitches affect such critical gene expression control across time.

Keywords: RNA polymerase, RNA folding, riboswitch control of gene expression, transcription, structural dynamics

OPEN ACCESS

Edited by:
Rene A. Nome,
State University of Campinas, Brazil

Reviewed by:
Bibekanand Mallick,
National Institute of Technology Rourkela, India
Kristine Bourke Arnvig,
University College London, United Kingdom

****Correspondence:***
Nils G. Walter
nwalter@umich.edu

Specialty section:
This article was submitted to Protein and RNA Networks, a section of the journal Frontiers in Molecular Biosciences

Received: *22 September 2020*
Accepted: *11 December 2020*
Published: *13 January 2021*

Citation:
Scull CE, Dandpat SS, Romero RA and Walter NG (2021) Transcriptional Riboswitches Integrate Timescales for Bacterial Gene Expression Control. Front. Mol. Biosci. 7:607158. doi: 10.3389/fmolb.2020.607158

INTRODUCTION

Over the past several decades, RNA has emerged as a key player beyond a "message" between DNA and protein. Non-coding RNAs (ncRNAs) are vital for countless cellular mechanisms, such as ribozyme mediated catalysis of RNA processing reactions, RNA mediated gene silencing, and stabilization of phase separated particles, to name a few (Eddy, 2001; Walter and Engelke, 2002; Eulalio et al., 2007; Serganov and Patel, 2007; Carthew and Sontheimer, 2009; Palazzo and Lee, 2015; Ravikumar et al., 2019; Herviou et al., 2020; Tollerson and Ibba, 2020). The functionality of RNA can be attributed to its propensity to fold into a variety of different structures on a rugged free-energy landscape (Chen and Dill, 2000; Thirumalai et al., 2001; Mustoe et al., 2014). To form functional, dynamic structures, RNA must overcome internal electrostatic repulsion of its phosphate backbone to form stable hydrogen-bonding between heteroatoms in the form of both Watson-Crick and non-Watson-Crick base pairs and sugar-base interactions (Rich, 2009). In the cell, these structural and conformational transitions occur in the presence of counterions, metabolites, small molecules, and proteins that form a plethora of interactions with functional RNAs to achieve critical cellular outcomes (Winkler et al., 2003; Mandal and Breaker, 2004; Breaker, 2012; Frieda and Block, 2012; Suddala et al., 2019; Chauvier et al., 2020; Zhang, 2020). A complete study of the structural dynamics of RNA is crucial for understanding its role beyond its canonical

function as a coding messenger RNA (mRNA). Emerging biophysical techniques such as single-molecule microscopy, multi-dimensional NMR, and most recently, near-atomic resolution cryo-electron microscopy (cryo-EM) are yielding new insights into the time-dependent evolution of RNA structures (Tinoco et al., 2010; Ray et al., 2019; Chang et al., 2020).

Riboswitches are a group of dynamic ncRNA motifs that exist almost exclusively in prokaryotes (Blount and Breaker, 2006), although there have been a few riboswitches described in eukaryotes that appear to modulate splicing (McCown et al., 2017), and more recently some riboswitch-like elements were discovered in viral genomes (Chahal et al., 2019). Riboswitches are usually found upstream, in the 5′ untranslated region of mRNAs, where they regulate transcription and translation through binding of their cognate ligand to their aptamer domain (**Figure 1A**). Aptamers have evolved to bind diverse ligands, from small molecules (often metabolites like s-adenosyl-methionine and $preQ_1$), to cations (such as Mn^{2+}), anions (such as F^-), and even other RNAs (such as tRNAs) (Poiata et al., 2009; Suddala et al., 2015; Widom et al., 2018; Chauvier et al., 2019; Zhang, 2020). Upon binding to their cognate ligands, riboswitches alter the secondary structure of a downstream domain, termed the expression platform, which turns "on" or "off" either transcription termination or translation initiation (**Figure 1A**, Widom et al., 2018). This review primarily focuses on the function of riboswitches involved in transcription regulation, encompassing events on timescales ranging from ligand binding to RNA folding, RNA transcription, and far-reaching cellular gene expression control.

Transcriptional riboswitch activity can essentially be broken down into 4 steps that occur on distinct timescales: (1) ligand influx, which can be as fast as the rate of diffusion, (2) ligand binding to the RNA, establishing specific hydrogen bonding, stacking and ionic interactions (nanoseconds to single seconds), (3) alterations in RNA secondary structure (milliseconds to multiple seconds), and (4) regulation of transcription elongation/termination and their downstream biological consequences (seconds to minutes) (**Figure 1B**) (Al-Hashimi and Walter, 2008). Until recently, studies of riboswitch folding were often executed in the absence of the transcriptional machinery (Duesterberg et al., 2015), despite the reality that riboswitch folding *in vivo* occurs co-transcriptionally (Frieda and Block, 2012). This leaves a gap between the field's understanding of riboswitch activity from molecular to cellular levels. RNA structures have recently been revealed to impact active site conformations and transcription activity of bacterial RNA polymerase (RNAP) in both *cis* and *trans* (Sedlyarova et al., 2017; Kang et al., 2018). A study by the Walter lab unveiled that the nascent $preQ_1$ riboswitch's secondary structure directly influences pausing behavior of the transcribing RNAP (Widom et al., 2018). This study illustrates that both the template DNA and RNAP have a significant impact on riboswitch folding, and vice versa. Thus, to achieve proper control of transcript synthesis and ultimately protein expression, the four steps of transcriptional riboswitch activity must be kinetically coupled (Ray et al., 2019).

COORDINATION OF RIBOSWITCH ACTIVITY FOR ADVANTAGEOUS BIOLOGICAL OUTCOMES

The timescales associated with gene regulation vary widely in both bacteria and eukaryotes, and are dependent on fine-tuned cellular sensitivity to external environmental signals (Hargrove et al., 1991; Shamir et al., 2016). Due to the complexity and compartmentalization of eukaryotic cells, gene regulation is relatively isolated, both spatially and temporally (Mandal and Breaker, 2004; Ralston, 2008). By contrast, in simpler organisms such as bacteria, with generally little membrane-enclosed sub-cellular compartmentalization, the colocalization of transcription and translation of their genes (Ralston, 2008) engenders regulation through direct coupling of processes. This makes bacteria a simple yet elegant model to study gene regulation (Proshkin et al., 2010; Kohler et al., 2017). Leveraging this coupled system, bacteria have evolved a variety of motifs within the nascent mRNA, including riboswitches and specific sequence elements, that induce transcriptional pausing and backtracking (Zhang et al., 2010; Perdrizet et al., 2012; Steinert et al., 2017). Furthermore, *Escherichia coli* (*E. coli*) bacteria have even been shown to directly couple transcription and translation, with ribosomes binding to mRNA during active transcription elongation by RNAP (Kohler et al., 2017; O'Reilly et al., 2020). This feature allows for the precise orchestration of gene regulation through the formation of the tightly coupled and highly efficient machinery termed the "expressome" (Proshkin et al., 2010; Kohler et al., 2017; O'Reilly et al., 2020; Washburn et al., 2020). The signal of a small ligand affecting the local structure of a riboswitch can then be transduced into a profound change in expressome function through a wave of kinetic selection, creating a system analogous to the struggle of "David vs. Goliath," where a tiny metabolic ligand has the ability to control the activity of the giant expressome (Ray et al., 2019).

The connection of fast intermolecular reactions (ligand influx and binding) to relatively slow global gene regulation is a critical modulator for sustaining the life of prokaryotes. For example, the Mn^{2+} sensing riboswitch found in *Salmonella* modulates the uptake of the transition metal ion Mn^{2+}, which is required for the virulence of this pathogenic bacterium (Shi et al., 2014). *Salmonella* is an intracellular pathogen that is phagocytized by host immune cells and resides in specialized cellular compartments known as *Salmonella*-containing vacuoles (SCVs) (Zaharik et al., 2004). SCVs contain host transmembrane transporters (such as Nramp1) that remove divalent cations from the vacuole to starve the pathogen's supply of essential cofactors (Forbes and Gros, 2001; Shi et al., 2014). For this reason, once internalized by SCVs, *Salmonella* must carefully balance intracellular concentrations of divalent cations to maintain sufficient but permissive concentration of Mn^{2+} ions. (Forbes and Gros, 2001; Shi et al., 2014). This balance is mediated in part by a Mn^{2+} sensitive riboswitch found upstream of a gene coding for mntH, a Mn^{2+} specific transporter (**Figure 2**, blue Mn^{2+} transport protein). When *Salmonella* is deficient in Mn^{2+}, transcription elongation of the mntH gene becomes permissive,

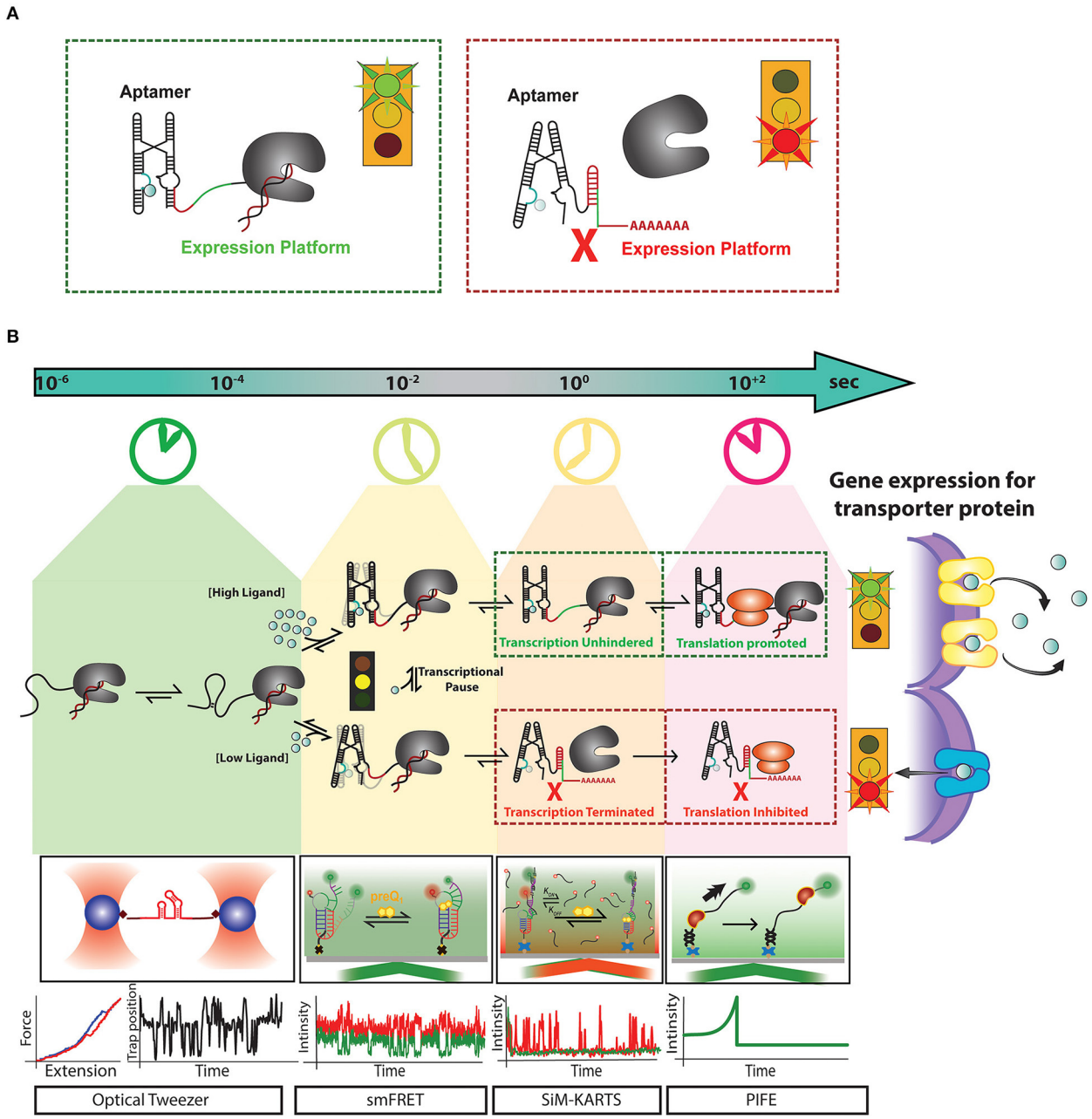

FIGURE 1 | Recent advances in single-molecule techniques have allowed for the study of transcriptional riboswitches on a variety of biologically relevant timescales: **(A)** Transcriptional riboswitches, such as the Mn^{2+} riboswitch, consist of a ligand binding, or "aptamer" region (cyan) that controls the shape of the "expression platform" region (red) to ultimately control transcription termination by the transcribing RNAP (gray). **(B)** Riboswitch mediated upregulation of a gene happens co-transcriptionally where events like RNA folding leading to aptamer formation occur at a timescale of micro-second to milliseconds. After partial aptamer formation, ligand binding events compete with tertiary folding events which happen in the timescale of milliseconds to seconds. Events such as transcriptional pausing play an essential role in regulating the ligand binding and RNA folding events at the co-transcriptional level. For a riboswitch upregulating gene expression for metal ion transporter proteins, at high concentration metal ion, the riboswitch binds them as ligand and from anti-terminator promoting transcription and releases ribosome binding site available for initiating translation. The formation of transporter proteins due to this signaling releases excess metal ions out of cytoplasm to avoid toxicity. In contrast, when there is a low concentration of metal ions required for cell function, riboswitch for terminator hairpin to prevent transcription and sequesters ribosome binding site to block translation. Single-molecule techniques like optical tweezer have been very effective in measuring events at a faster timescale to monitor stepwise RNA folding and unfolding. smFRET has been adapted as a useful tool to monitor the interaction between RNA folding and ligand binding. Single-molecule methods like Single-Molecule Kinetic Analysis of RNA Transient Structures (SiM-KARTS) been used as an essential tool to probe changes in RNA structure as an alternative to SMFRET that requires site-specific labeling and can be extended to study binding events happening during co-transcriptional events. Protein induced fluorescence enhancement (PIFE) has emerged as an important tool to monitor the speed of transcription and activity of RNA polymerase happening in the timescale of seconds. These unique techniques have been essential for the study of riboswitch regulated events at different timescales. However, a combination of one or more of these techniques will be a powerful tool to decipher the real-time mechanism of the expressome at each step of gene regulation.

subsequently allowing for expression of the transporter and resulting in an increase in Mn^{2+} uptake (Shi et al., 2014). Once sufficient Mn^{2+} has been imported, binding of excess Mn^{2+} to the Mn^{2+} riboswitch turns "off" expression of mntH so that concentration levels of the divalent cation do not result in cytotoxicity (Forbes and Gros, 2001; Shi et al., 2014).

In addition to the mntH gene, *E. coli* contains an efflux transporter protein, called mntP (**Figure 2**, yellow Mn^{2+} transport protein) (Waters et al., 2011). mntP's expression can be controlled at both the transcriptional and translational levels by the *yybP–ykoY* Mn^{2+} riboswitch (Suddala et al., 2019), which induces both transcription elongation and translation of *mntP* upon binding Mn^{2+} ions at a sub-cytotoxic concentration. After expression, mntP exports excess divalent cation when its concentration exceeds permissive levels (Dambach et al., 2015). Clearly, mediation of genetic control by riboswitches is both highly dynamic and critical for maintaining a proper balance of Mn^{2+} homeostasis in bacteria.

TOWARD A HOLISTIC UNDERSTANDING OF RIBOSWITCH FUNCTION

While biochemical and genetic studies have been essential for identification of functional riboswitches, critical advances in studies *in vivo* as well as *in vitro* have allowed us to understand how small molecular interactions in riboswitches ultimately propagate into global changes in gene expression. A recently developed technique called *Re*porter *Co*upled *I*n *C*ell *S*elective 2′-*H*ydroxyl *A*cylation analyzed by *P*rimer *E*xtension (ReCo-icSHAPE) is one approach by which the coupling of ligand influx to ligand-induced folding of a translational riboswitch and the subsequent impact on the expression of a reporter gene can be monitored directly (Dutta et al., 2018). Using this strategy, a $preQ_1$ type II translational riboswitch from *Lactobacillus rhamnosus* (Lrh) was probed in *E. coli*, confirming and complementing expression studies of a GFP-coupled reporter. Through intracellular icSHAPE, this study revealed stronger $preQ_1$ mediated occlusion of the ribosome binding site than was observed in the previously described structure, which was solved in purified form via X-ray deffraction (Dutta et al., 2018). Although this approach was used here to probe a translational riboswitch, in the future, this tool may be utilized for directly monitoring the coordination of poorly understood transcriptional riboswitches. This would allow for holistic studies of all stages of riboswitch mediated gene control from the initiation of transcription to ultimate protein expression (**Figure 1B**).

Riboswitches in Prokaryotes Allow for Tight Coupling of Ligand Influx and Gene Expression

In the precisely organized gene regulatory systems of bacteria, the efficiency of riboswitch regulation is dependent on the speed of transcription, long before the fate of translation is decided (Wickiser et al., 2005; Garst and Batey, 2009). That is, folding of the nascent RNA transcript in the wake of the elongating RNAP, and in response to ligand binding, will determine the outcome of gene expression (**Figure 2**). Due to such direct coupling of RNA folding with transcription, riboswitches are considered to kinetically control the regulation of the downstream genes (Wickiser et al., 2005). In the sequence of events, the binding rate of the ligand, and the folding rate of the RNA may be faster (nanosecond to millisecond timescales) than the rate of transcription (milliseconds to seconds) (Wickiser et al., 2005; Gilbert et al., 2006; Roberts et al., 2008; Breaker, 2012; Watters et al., 2016). Hence the cellular concentration of ligand must be higher than its dissociation constant (K_D) to allow for the ligand binding kinetics to outpace and thus drive the kinetics of RNA folding. A commonly observed mechanism called the induced fit (IF) mechanism describes riboswitch ligand binding events where ligand binding occurs faster than the conformational changes of the riboswitch, whereas ligand dissociation from the unfolded RNA is slower (Gilbert et al., 2006). Alternatively, the conformation selection (CS) model describes riboswitches where the RNA conformational change occurs faster than ligand binding, and ligand dissociates too rapidly from the unfolded RNA to achieve the IF mechanism (Suddala et al., 2015). The transition between the IF and CS models is governed, on one hand, by the ligand concentration and, on the other hand, by temperature and cofactors affecting RNA folding such as the cationic micro-environment. Undoubtedly, evolutionary pressures shape the sequence composition of the riboswitch to finetune this balance to the cell's needs (Suddala and Walter, 2014; Suddala et al., 2015; Rode et al., 2018). Ligand recognition mechanisms like the CS and IF models have provided the basis for the kinetic selection of transcriptional riboswitches (Suddala and Walter, 2014).

For some of the best described riboswitches, it is thought that the IF model is the prevailing mechanism driving riboswitch folding. However, aside from the challenges to accurately distinguish the two mechanisms, there exist examples of riboswitches where it is thought that (subtle) structural rearrangement occurs faster than ligand binding, hence gene regulation occurs via CS (Suddala et al., 2015). One example of such a finely tuned riboswitch, a fluoride sensing riboswitch, was observed to fold into identical tertiary structures, with or without its ligand F^-, yet triggered gene activation only upon recognition of F^- in a narrow concentration range (Zhao et al., 2017). NMR spectroscopy revealed that in the absence of F^-, the aptamer forms a transitory state of short lifetime (somewhat confusingly termed "excited state" when only thermally-activated sampling is required), which unlocks a linchpin-gated structure that promotes transcription termination. The presence of F^- stabilizes the gated conformation toward a functional response within a narrow range of ligand concentrations over a wide range of transcription rates (Zhao et al., 2017). More generally, this example showcases how full refolding of a riboswitch between two alternate secondary structures—which would be slow and come with a thermodynamic barrier likely higher than the energetic driving force available from the RNA binding a small ligand—may often be avoided by an intricate coupling of RNA folding with transcription elongation wherein linchpin

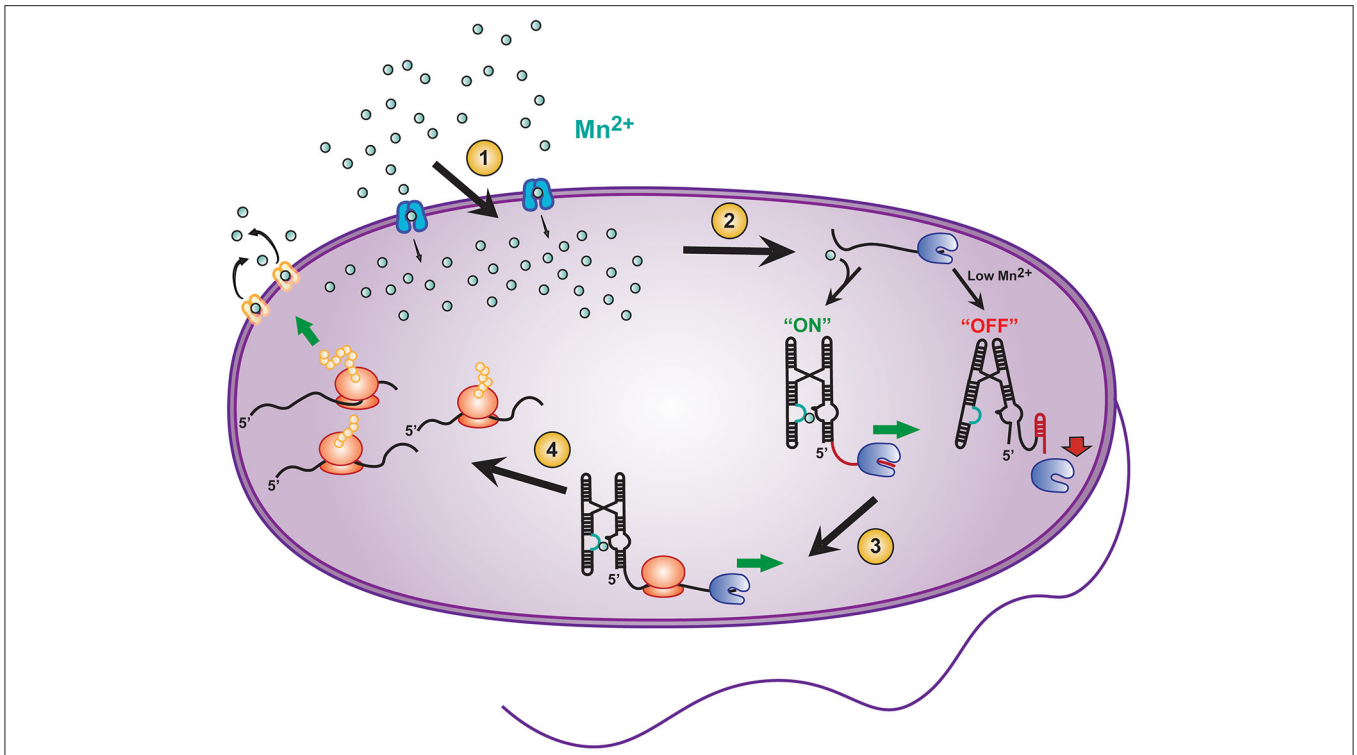

FIGURE 2 | Transcriptional riboswitches, such as the Mn^{2+} riboswitch, involve integration of various time scales to modulate gene expression and maintain cellular homeostasis in *E. coli*. A number of coordinated steps are required for riboswitch mediated maintenance of cellular Mn^{2+} homeostasis: (1) Modulation of Mn^{2+} transport into the cell by mntH (*blue transporter*), (2) binding of Mn^{2+} to the *yybP-ykoY* riboswitch on the 5′ UTR of the *mntP* gene to permit transcription elongation by RNAP (*blue packman*), (3) co-transcriptional translation of *mntP* to ultimately result in, (4) increased expression of the mntP transporter (*yellow transporter*) which exports excessive intracellular Mn^{2+}.

events gate the partitioning between the alternate structures (Zhao et al., 2017).

The *yybP–ykoY* family of Mn^{2+} ion sensing riboswitches represent another model for the CS mechanism at physiological concentrations of divalent ions (**Figures 1**, **2**) (Suddala et al., 2019; Sung and Nesbitt, 2019). They have been found to upregulate expression of Mn^{2+} homeostasis genes by binding both Mg^{2+} and Mn^{2+} ions in two adjacent metal ion binding pockets occupying a linchpin position that, once occupied, promotes transcription (Guo et al., 2018; Suddala et al., 2019). The cooperative binding of Mg^{2+} and Mn^{2+} is thought to follow a CS mechanism that stabilizes an adjoining helix P1.1, which in turn competes with a terminator stem that disrupts transcription (**Figures 1**, **2**). Structural studies demonstrate the riboswitch pre-arranging a four-way junction in the presence of millimolar concentrations of Mg^{2+} such that two of the helical arms become transiently juxtaposed, allowing for Mn^{2+} to be captured to turn the riboswitch "on" (Frieda and Block, 2012; Saba et al., 2019) (**Figures 1**, **2**). More broadly, the hierarchy of folding events can be perturbed in riboswitches by the presence of non-cognate ligands and by mutations in the ligand binding aptamer domain. This enables integration over competing metabolic signals and sequence evolution for functional adaptation. Any such perturbation must take effect on the timescale of transcription, emphasizing the kinetic role of ligand and RNA sequence specificity around the aptamer region (Price et al., 2015; Suddala et al., 2015, 2019).

Correlating the Timescales of Ligand Binding to RNA Folding

Modern biophysical techniques, such as single-molecule fluorescence resonance energy transfer (smFRET) and optical tweezers, have quickly become essential tools for monitoring ligand-dependent structural changes in the aptamer region (**Figure 1B**) (Savinov et al., 2014). smFRET in particular has been widely used to probe conformational changes in riboswitches at varying concentrations of ligand to correlate ligand binding with RNA folding (Savinov et al., 2014; Suddala and Walter, 2014; Ray et al., 2019). In other cases, such as *in vitro* evolved aptamers with more open binding pockets, the ligand itself can be labeled to monitor its binding to single RNA molecules (Elenko et al., 2009).

In most reports, however, indirect changes in RNA folding and unfolding dynamics are used to probe the mechanism of the aptamer-ligand interaction. For example, one of the smallest riboswitches, the class I $preQ_1$-sensing riboswitch follows the two ligand binding mechanisms of IF and CS dependent on the ligand and metabolite conditions as well as specific sequence adaptations found in various bacteria (Suddala et al., 2013, 2015). The transcriptional $preQ_1$ riboswitch from *Bacillus subtillis*

(*Bsu*) has been observed to favor a CS pathway where the ligand primarily binds to a pre-selected conformation of the aptamer (Suddala et al., 2015). However, the ligand recognition mechanism is fluid, as the same riboswitch can adopt instead the IF mechanism at low metabolite and Mg^{2+} concentrations (Suddala et al., 2015). This dependence of the folding pathway on the relative timescales of ligand binding and conformational dynamics of the aptamer can be identified as a kinetic coupling mechanism occurring early in the decision tree of gene regulation (**Figure 1B**). Similar kinetic control mechanisms of ligand recognition by the aptamer have been observed to be operational in multiple other riboswitches (Manz et al., 2017; Rode et al., 2018; McCluskey et al., 2019; Sung and Nesbitt, 2019).

Correlating the Timescales of Riboswitch Folding, Transcription, and Gene Expression

During bacterial transcription, both the kinetics of ligand binding and the speed of RNA transcription determine the functionality of the riboswitch beyond the ligand binding to its aptamer (Wickiser et al., 2005). The coupling observed for the rates of transcription elongation and RNA folding as it emerges from the RNAP exit tunnel in 5′-to-3′ direction highlights the importance of studying riboswitches in the context of the transcription elongation complex. In fact, it has been demonstrated in other RNA folding systems (such as bacterial ribosome biogenesis) that the co-transcriptional directionality of its folding influences an RNA's interactions with known binding partners (Duss et al., 2019; Rodgers and Woodson, 2019). Transcriptional regulatory events, such as pausing, are crucial to the balance between RNA folding and the speed of additional RNA sequence emerging in the wake of RNAP (Saba et al., 2019). Studies have shown how sequence-specific pausing allows the nascent RNA to reach an equilibrium of folded states that then can be further stabilized by RNA binding molecules (Watters et al., 2016; Widom et al., 2018; Rodgers and Woodson, 2019). Depending on cellular conditions, transcription factors such as NusA and NusG are found to stabilize and disrupt transcriptional pausing, respectively (Yakhnin et al., 2016; Guo et al., 2018; Kang et al., 2019). During transcription elongation, a cascade of faster events including RNA folding, ligand binding, and interactions of the RNA with RNAP, together with variations of transcription speed over time, integrate over the biological state of the cell to govern downstream gene regulation (**Figure 1B**). In light of the reversibility of many, and irreversibility of some, of these steps, conformational and kinetic proofreading becomes possible, adding critical layers of control over the ultimate gene expression outcome (Walter, 2019).

Several recent studies have highlighted details of the co-transcriptional nature of riboswitch folding. High-resolution optical tweezers and single-molecule force spectroscopy approaches showed that a co-transcriptionally folded adenine riboswitch undergoes transcription readthrough predominantly in the presence of adenine, while its absence leads to transcription termination (Frieda and Block, 2012). smFRET assays further demonstrated kinetic control of co-transcriptional folding of a thiamine pyrophosphate (TPP) riboswitch (Uhm et al., 2018). The isolated riboswitch aptamer was observed to fold into a translation "off" conformation independently of its TPP ligand. By contrast, transcriptional pausing allows the riboswitch to rearrange into an "on" conformation in the absence of TPP, while ligand binding steers the nascent RNA into the "off" conformation to downregulate gene expression. This work illustrated that only a brief time window between transcriptional pausing and ligand binding determines the fate of downstream gene expression (Uhm et al., 2018). In case of a F^--sensing riboswitch, co-transcriptional *S*elective 2′-*H*ydroxyl *A*cylation analyzed by *P*rimer *E*xtension (SHAPE)-seq revealed that the riboswitch is controlled by the kinetics of co-transcriptional folding, which drives the RNA into a short-lived folded state, even in the absence of F^-. Binding of ligand favors a kinetically trapped, stably folded state, which delays the nucleation of the terminator hairpin until RNAP has escaped the terminator poly(U) sequence to continue transcription (Watters et al., 2016).

To study the physical interaction between a riboswitch and RNAP, together with the role of a consensus pause sequence on co-transcriptional folding, Widom et al. (2018) performed smFRET, biochemical transcription assays, and molecular dynamics simulations on the paused elongation complex of the class III *que* pause featuring the $preQ_1$ riboswitch. This study demonstrated that, on the time scale of transcription, pausing allows the RNAP to slow down and the riboswitch aptamer to sense ligand. This ultimately stabilizes a fully folded RNA pseudoknot conformation that releases the paused RNAP. Additionally, transcription elongation rates likely play a role in riboswitch folding. Early studies in *E. coli* indicated that changes in transcription elongation rate disrupt gene expression and cell growth (Lewicki et al., 1993; Scull and Schneider, 2019). This phenomenon also exists in eukaryotes, where an alteration in the elongation rate of RNAP I disrupts ribosome biogenesis (Schneider et al., 2006, 2007), and a change in RNAP II transcription speed disrupts mRNA splicing (Brzyzek and Swiezewski, 2015), suggesting the universality of such layers of gene expression control. Changes in transcription elongation rate are modulated *in vivo* by covalent modification of the polymerase itself (Fath et al., 2001, 2004), as well as through positive and negative transcription elongation factors, such as NusA and NusG in bacteria (Herbert et al., 2010; Zhou et al., 2011). Future studies on the relationship between transcription elongation rate and riboswitch folding will likely discover additional pause-independent transcriptional regulatory mechanisms. Importantly, future studies should further probe riboswitch folding in the context of RNAP elongation rate, both in the presence and absence of the transcription termination machinery—as both RNA folding and transcription termination could potentially be influenced by RNAP elongation rate.

CONCLUSIONS AND PERSPECTIVES

Since the discovery of riboswitches in 2002, nearly 20 years ago, methods have dramatically progressed from structural to

kinetic studies. These new Technologies are paving the way for a comprehensive understanding of the underlying dynamics under a broad range of conditions and timescales. The structural organization of the aptamer domain and expression platform in the absence or presence of ligand remains critical to a foundational understanding of riboswitch function. Established methods including X-ray crystallography together with more recent advances in high-resolution cryo-EM have provided snapshots of RNA structures that have aided in identifying ligand binding sites and RNA structures (Garst and Batey, 2009; Frank, 2017; Zhang et al., 2019).

Traditional techniques like in-line probing, dimethyl sulfate (DMS) footprinting and SHAPE have allowed for monitoring of structural changes upon addition of a ligand (Soukup and Breaker, 1999; Winkler et al., 2003). Intracellular footprinting by DMS and SHAPE-seq, the latter of which was shown to be able to incorporate selection for active elongation complexes into the original SHAPE protocol (Takahashi et al., 2016; Mitchell et al., 2019), indicate global and some local conformational changes in riboswitches. However, as riboswitches fold asynchronously, smaller local changes may be missed by population averaging (Chauvier et al., 2019; Ray et al., 2019). Advances in single molecule methods such as smFRET and SiM-KARTS have further facilitated our understanding of riboswitches and their mechanisms with the ability to kinetically probe both local and global dynamics and obtain folding and unfolding rate constants (**Figure 1B**) (Chauvier et al., 2019; Ray et al., 2019). Complementary force spectroscopy experiments have enabled real-time mapping of secondary structure dynamics under perturbation, utilizing magnetic or optical tweezers (**Figure 1B**) (Frieda and Block, 2012; Tomko and Galburt, 2019). These methods have been instrumental for our understanding of riboswitch dynamics; however, recent advances in the field are increasingly shifting to allow for the study of riboswitches under more biologically relevant conditions.

The overall goal of the field of riboswitch biology remains the same: to understand the mechanisms by which riboswitches bind ligands to transduce a signal through conformational changes in the expression platform for ultimate control of gene expression. Recent studies have focused on the importance of co-transcriptional folding and how the elongation complex affects riboswitch mechanism, and vice versa (Watters et al., 2016; Ray et al., 2019; Strobel et al., 2019). Transcription rates can range from 10 to 25 nucleotides per second, and the RNA immediately starts folding directly after exiting RNAP, leaving only a short time window for riboswitches to sample alternative folding pathways in service of gene regulation (Dangkulwanich et al., 2014). Co-transcriptional studies are beginning to highlight the importance of the context of the transcription machinery and the critical role that integration of timescales plays in the mechanisms of gene regulation (**Figure 1B**) (Watters et al., 2016; Ray et al., 2019; Strobel et al., 2019). Conversely, the discovery of the functional importance of co-transcriptional riboswitch folding has driven the development of techniques such as co-transcriptional SHAPE-seq and artificial RNA elongation complex assembly (Watters et al., 2016; Strobel et al., 2019). Co-translational SHAPE-seq, in turn, has enabled high-throughput structural probing of RNAP complexes halted *in vitro* at various transcript lengths to obtain single-nucleotide resolution of the nascent RNA (Watters et al., 2016). To truly understand how ligand binding by a riboswitch couples to gene expression, future studies must increasingly monitor these events in concert and acknowledge that RNA acts as an active effector of gene regulation rather than a passive output. Integrating riboswitches into their biologically relevant contexts will require directly monitoring single molecules at a broad range of timescales and including a plethora of external cofactors that may influence folding mechanisms during and after transcription.

Technical developments on the horizon will include single molecule assays that monitor elongating RNAP complexes using total internal reflection fluorescence (TIRF) and zero-mode waveguide (ZWM) microscopy at increasingly higher throughput of individual molecules while maintaining high sensitivity (Duss et al., 2018, 2019). Protein Induced Fluorescence Enhancement (PIFE) has emerged as a way to measure transcription rates and, when coupled with FRET or fluorescent probes to monitor ligand binding and dynamics of the transcript, enables the measurement of real-time kinetics during transcription (Duss et al., 2018, 2019; Rodgers and Woodson, 2019). There is also potential for three- or four-color smFRET to monitor multiple dynamic interactions simultaneously within a single riboswitch (Lee et al., 2010). These developments will allow for kinetic measurements under varying conditions, including at RNAP pause sites and in the presence of transcription factors, divalent ions, and other intracellular factors, without the need for synchronizing individual molecules by sudden perturbation (Suddala et al., 2015; Gabizon et al., 2018; Widom et al., 2018). Extending these cutting-edge techniques to riboswitches will beget a deeper understanding of how RNA can efficiently integrate environmental cues over broad timescales to affect bacterial gene expression control and survival, with the promise of boosting our ability to suppress bacterial infections with antibiotics.

AUTHOR CONTRIBUTIONS

CS, SD, RR, and NW wrote the manuscript. CS and SD produced the figures. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by NIH/National Institute of General Medical Sciences grants: GM131922, GM062357, GM118524, and GM122803 to NW, a Michigan Life Sciences Fellowship to CS, and Rackham Merit Fellowship and Michigan Predoctoral Training in Genetics (5T32GM007544-43) to RR.

ACKNOWLEDGMENTS

We thank all the members of the Walter lab, especially Adrien Chauvier and Javier Cabello-Villegas, for their support and suggestions on this manuscript.

REFERENCES

Al-Hashimi, H. M., and Walter, N. G. (2008). RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* 18, 321–329. doi: 10.1016/j.sbi.2008.04.004

Blount, K. F., and Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* 24, 1558–1564. doi: 10.1038/nbt1268

Breaker, R. R. (2012). Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* 4:a003566. doi: 10.1101/cshperspect.a003566

Brzyzek, G., and Swiezewski, S. (2015). Mutual interdependence of splicing and transcription elongation. *Transcription* 6, 37–39. doi: 10.1080/21541264.2015.1040146

Carthew, R. W., and Sontheimer, E. J. (2009). Origins and mechanisms of miRNAs and siRNAs. *Cell* 136, 642–655. doi: 10.1016/j.cell.2009.01.035

Chahal, J., Gebert, L. F. R., Gan, H. H., Camacho, E., Gunsalus, K. C., MacRae, I. J., et al. (2019). miR-122 and ago interactions with the HCV genome alter the structure of the viral 5′ terminus. *Nucleic Acids Res.* 47, 5307–5324. doi: 10.1093/nar/gkz194

Chang, J. Y., Cui, Z., Yang, K., Huang, J., Minary, P., and Zhang, J. (2020). Hierarchical natural move monte carlo refines flexible RNA structures into cryo-EM densities. *RNA* 26, 1755–1766. doi: 10.1261/rna.071100.119

Chauvier, A., Ajmera, P., and Walter, N. G. (2020). Competition between ligand binding and transcription rate modulates riboswitch-mediated regulation of transcription. *Biophys. J.* 118:68A. doi: 10.1016/j.bpj.2019.11.545

Chauvier, A., Cabello-Villegas, J., and Walter, N. G. (2019). Probing RNA structure and interaction dynamics at the single molecule level. *Methods* 162–163, 3–11. doi: 10.1016/j.ymeth.2019.04.002

Chen, S.-J., and Dill, K. A. (2000). RNA folding energy landscapes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 646–651. doi: 10.1073/pnas.97.2.646

Dambach, M., Sandoval, M., Updegrove, T.aylor B., Anantharaman, V., Aravind, L., Waters, L.auren S., et al. (2015). The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element. *Mol. Cell* 57, 1099–1109. doi: 10.1016/j.molcel.2015.01.035

Dangkulwanich, M., Ishibashi, T., Bintu, L., and Bustamante, C. (2014). Molecular mechanisms of transcription through single-molecule experiments. *Chem. Rev.* 114, 3203–3223. doi: 10.1021/cr400730x

Duesterberg, V. K., Fischer-Hwang, I. T., Perez, C. F., Hogan, D. W., and Block, S. M. (2015). Observation of long-range tertiary interactions during ligand binding by the TPP riboswitch aptamer. *Elife* 4:e12362. doi: 10.7554/eLife.12362

Duss, O., Stepanyuk, G. A., Grot, A., O'Leary, S. E., Puglisi, J. D., and Williamson, J. R. (2018). Real-time assembly of ribonucleoprotein complexes on nascent RNA transcripts. *Nat. Commun.* 9:5087. doi: 10.1038/s41467-018-07423-3

Duss, O., Stepanyuk, G. A., Puglisi, J. D., and Williamson, J. R. (2019). Transient protein-RNA interactions guide nascent ribosomal RNA folding. *Cell 179*, 1357–1369.e16. doi: 10.1016/j.cell.2019.10.035

Dutta, D., Belashov, I. A., and Wedekind, J. E. (2018). Coupling green fluorescent protein expression with chemical modification to probe functionally relevant riboswitch conformations in live bacteria. *Biochemistry* 57, 4620–4628. doi: 10.1021/acs.biochem.8b00316

Eddy, S. R. (2001). Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2, 919–929. doi: 10.1038/35103511

Elenko, M. P., Szostak, J. W., and van Oijen, A. M. (2009). Single-molecule imaging of an *in vitro*-evolved RNA aptamer reveals homogeneous ligand binding kinetics. *J. Am. Chem. Soc.* 131, 9866–9867. doi: 10.1021/ja901880v

Eulalio, A., Behm-Ansmant, I., Schweizer, D., and Izaurralde, E. (2007). P-body formation is a consequence, not the cause, of RNA-mediated gene silencing. *Mol Cell Biol.* 27, 3970–3981. doi: 10.1128/MCB.00128-07

Fath, S., Kobor, M. S., Philippi, A., Greenblatt, J., and Tschochner, H. (2004). Dephosphorylation of RNA polymerase I by Fcp1p is required for efficient rRNA synthesis. *J. Biol. Chem.* 279, 25251–25259. doi: 10.1074/jbc.M401867200

Fath, S., Milkereit, P., Peyroche, G., Riva, M., Carles, C., and Tschochner, H. (2001). Differential roles of phosphorylation in the formation of transcriptional active RNA polymerase I. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14334–14339. doi: 10.1073/pnas.231181398

Forbes, J. R., and Gros, P. (2001). Divalent-metal transport by NRAMP proteins at the interface of host–pathogen interactions. *Trends Microbiol.* 9, 397–403. doi: 10.1016/S0966-842X(01)02098-4

Frank, J. (2017). Time-resolved cryo-electron microscopy: recent progress. *J. Struct. Biol.* 200, 303–306. doi: 10.1016/j.jsb.2017.06.005

Frieda, K. L., and Block, S. M. (2012). Direct observation of cotranscriptional folding in an adenine riboswitch. *Science* 338, 397–400. doi: 10.1126/science.1225722

Gabizon, R., Lee, A., Vahedian-Movahed, H., Ebright, R. H., and Bustamante, C. J. (2018). Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. *Nat. Commun.* 9:2930. doi: 10.1038/s41467-018-05344-9

Garst, A. D., and Batey, R. T. (2009). A switch in time: detailing the life of a riboswitch. *Biochim. Biophys. Acta* 1789, 584–591. doi: 10.1016/j.bbagrm.2009.06.004

Gilbert, S. D., Stoddard, C. D., Wise, S. J., and Batey, R. T. (2006). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J. Mol. Biol.* 359, 754–768. doi: 10.1016/j.jmb.2006.04.003

Guo, X., Myasnikov, A. G., Chen, J., Crucifix, C., Papai, G., Takacs, M., et al. (2018). Structural basis for NusA stabilized transcriptional pausing. *Mol. Cell* 69, 816–827.e4. doi: 10.1016/j.molcel.2018.02.008

Hargrove, J. L., Hulsey, M. G., and Beale, E. G. (1991). The kinetics of mammalian gene expression. *BioEssays* 13, 667–674. doi: 10.1002/bies.950131209

Herbert, K. M., Zhou, J., Mooney, R. A., Porta, A. L., Landick, R., and Block, S. M. (2010). *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *J. Mol. Biol.* 399, 17–30. doi: 10.1016/j.jmb.2010.03.051

Herviou, P., Le Bras, M., Dumas, L., Hieblot, C., Gilhodes, J., Cioci, G., et al. (2020). hnRNP H/F drive RNA G-quadruplex-mediated translation linked to genomic instability and therapy resistance in glioblastoma. *Nat. Commun.* 11:2661. doi: 10.1038/s41467-020-16168-x

Kang, J. Y., Mishanina, T. V., Landick, R., and Darst, S. A. (2019). Mechanisms of transcriptional pausing in bacteria. *J. Mol. Biol.* 431, 4007–4029. doi: 10.1016/j.jmb.2019.07.017

Kang, J. Y., Mooney, R. A., Nedialkov, Y., Saba, J., Mishanina, T. V., Artsimovitch, I., et al. (2018). Structural basis for transcript elongation control by NusG family universal regulators. *Cell 173*, 1650–1662.e14. doi: 10.1016/j.cell.2018.05.017

Kohler, R., Mooney, R. A., Mills, D. J., Landick, R., and Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science* 356, 194–197. doi: 10.1126/science.aal3059

Lee, J., Lee, S., Ragunathan, K., Joo, C., Ha, T., and Hohng, S. (2010). Single-molecule four-color FRET. *Angew. Chem. Int. Ed. Engl.* 49, 9922–9925. doi: 10.1002/anie.201005402

Lewicki, B. T., Margus, T., Remme, J., Nierhaus, K. H. (1993). Coupling of rRNA transcription and ribosomal assembly *in vivo*. Formation of active ribosomal subunits in *Escherichia coli* requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase. *J. Mol. Biol.* 231, 581 −593. doi: 10.1006/jmbi.1993.1311

Mandal, M., and Breaker, R. R. (2004). Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* 5, 451–463. doi: 10.1038/nrm1403

Manz, C., Kobitski, A. Y., Samanta, A., Keller, B. G., Jäschke, A., and Nienhaus, G. U. (2017). Single-molecule FRET reveals the energy landscape of the full-length SAM-I riboswitch. *Nat. Chem. Biol.* 13, 1172–1178. doi: 10.1038/nchembio.2476

McCluskey, K., Boudreault, J., St-Pierre, P., Perez-Gonzalez, C., Chauvier, A., Rizzi, A., et al. (2019). Unprecedented tunability of riboswitch structure and regulatory function by sub-millimolar variations in physiological Mg^{2+}. *Nucleic Acids Res.* 47, 6478–6487. doi: 10.1093/nar/gkz316

McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., and Breaker, R. R. (2017). Riboswitch diversity and distribution. *RNA* 23, 995–1011. doi: 10.1261/rna.061234.117

Mitchell, D. 3rd., Assmann, S. M., and Bevilacqua, P. C. (2019). Probing RNA structure *in vivo*. *Curr. Opin. Struct. Biol.* 59, 151–158. doi: 10.1016/j.sbi.2019.07.008

Mustoe, A. M., Brooks, C. L., and Al-Hashimi, H. M. (2014). Hierarchy of RNA functional dynamics. *Annu. Rev. Biochem.* 83, 441–466. doi: 10.1146/annurev-biochem-060713-035524

O'Reilly, F. J., Xue, L., Graziadei, A., Sinn, L., Lenz, S., Tegunov, D., et al. (2020). In-cell architecture of an actively transcribing-translating expressome. *Science* 369, 554–557. doi: 10.1126/science.abb3758

Palazzo, A. F., and Lee, E. S. (2015). Non-coding RNA: what is functional and what is junk? *Front. Genet.* 6:2. doi: 10.3389/fgene.2015.00002

Perdrizet, G. A., Artsimovitch, I., Furman, R., Sosnick, T. R., and Pan, T. (2012). Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* 109, 3323–3328. doi: 10.1073/pnas.1113086109

Poiata, E., Meyer, M. M., Ames, T. D., and Breaker, R. R. (2009). A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. *RNA* 15, 2046–2056. doi: 10.1261/rna.1824209

Price, I. R., Gaballa, A., Ding, F., Helmann, J. D., and Ke, A. (2015). Mn(2+)-sensing mechanisms of yybP-ykoY orphan riboswitches. *Mol. Cell* 57, 1110–1123. doi: 10.1016/j.molcel.2015.02.016

Proshkin, S., Rahmouni, A. R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* 328, 504–508. doi: 10.1126/science.1184939

Ralston, A. (2008). Simultaneous gene transcription and translation in bacteria. *Nat. Educ.* 1:4. Available online at: https://www.nature.com/scitable/topicpage/simultaneous-gene-transcription-and-translation-in-bacteria-1025/

Ravikumar, S., Devanapally, S., and Jose, A. M. (2019). Gene silencing by double-stranded RNA from *C. elegans* neurons reveals functional mosaicism of RNA interference. *Nucleic Acids Res.* 47, 10059–10071. doi: 10.1093/nar/gkz748

Ray, S., Chauvier, A., and Walter, N. G. (2019). Kinetics coming into focus: single-molecule microscopy of riboswitch dynamics. *RNA Biol.* 16, 1077–1085. doi: 10.1080/15476286.2018.1536594

Rich, A. (2009). The era of RNA awakening: structural biology of RNA in the early years. *Q. Rev. Biophys.* 42, 117–137. doi: 10.1017/S0033583509004776

Roberts, J. W., Shankar, S., and Filter, J. J. (2008). RNA polymerase elongation factors. *Ann. Rev. Microbiol.* 62, 211–233. doi: 10.1146/annurev.micro.61.080706.093422

Rode, A. B., Endoh, T., and Sugimoto, N. (2018). Crowding shifts the FMN recognition mechanism of riboswitch aptamer from conformational selection to induced fit. *Angew. Chem. Int. Edn.* 57, 6868–6872. doi: 10.1002/anie.201803052

Rodgers, M. L., and Woodson, S. A. (2019). Transcription increases the cooperativity of ribonucleoprotein assembly. *Cell* 179, 1370–1381.e12. doi: 10.1016/j.cell.2019.11.007

Saba, J., Chua, X. Y., Mishanina, T. V., Nayak, D., Windgassen, T. A., Mooney, R. A., et al. (2019). The elemental mechanism of transcriptional pausing. *Elife* 8:e40981. doi: 10.7554/eLife.40981

Savinov, A., Perez, C. F., and Block, S. M. (2014). Single-molecule studies of riboswitch folding. *Biochim. Biophys. Acta* 1839, 1030–1045. doi: 10.1016/j.bbagrm.2014.04.005

Schneider, D. A., French, S. L., Osheim, Y. N., Bailey, A. O., Vu, L., Dodd, J., et al. (2006). RNA polymerase II elongation factors Spt4p and Spt5p play roles in transcription elongation by RNA polymerase I and rRNA processing. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12707–12712. doi: 10.1073/pnas.0605686103

Schneider, D. A., Michel, A., Sikes, M. L., Vu, L., Dodd, J. A., Salgia, S., et al. (2007). Transcription elongation by RNA polymerase I is linked to efficient rRNA processing and ribosome assembly. *Mol. Cell* 26, 217–229. doi: 10.1016/j.molcel.2007.04.007

Scull, C. E., and Schneider, D. A. (2019). Coordinated control of rRNA processing by RNA polymerase I. *Trends Genet.* 35, 724–733. doi: 10.1016/j.tig.2019.07.002

Sedlyarova, N., Rescheneder, P., Magán, A., Popitsch, N., Rziha, N., Bilusic, I., et al. (2017). Natural RNA polymerase aptamers regulate transcription in *E. coli*. *Mol. Cell* 67, 30–43.e6. doi: 10.1016/j.molcel.2017.05.025

Serganov, A., and Patel, D. J. (2007). Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.* 8, 776–790. doi: 10.1038/nrg2172

Shamir, M., Bar-On, Y., Phillips, R., and Milo, R. (2016). SnapShot: timescales in cell biology. *Cell* 164, 1302–1302.e1. doi: 10.1016/j.cell.2016.02.058

Shi, Y., Zhao, G., and Kong, W. (2014). Genetic analysis of riboswitch-mediated transcriptional regulation responding to Mn^{2+} in Salmonella. *J. Biol. Chem.* 289, 11353–11366. doi: 10.1074/jbc.M113.517516

Soukup, G. A., and Breaker, R. R. (1999). Engineering precision RNA molecular switches. *Proc. Natl. Acad. Sci. U. S. A.* 96, 3584–3589. doi: 10.1073/pnas.96.7.3584

Steinert, H., Sochor, F., Wacker, A., Buck, J., Helmling, C., Hiller, F., et al. (2017). Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation. *eLife* 6:e21297. doi: 10.7554/eLife.21297

Strobel, E. J., Cheng, L., Berman, K. E., Carlson, P. D., and Lucks, J. B. (2019). A ligand-gated strand displacement mechanism for ZTP riboswitch transcription control. *Nat. Chem. Biol.* 15, 1067–1076. doi: 10.1038/s41589-019-0382-7

Suddala, K. C., Price, I. R., Dandpat, S. S., Janeček, M., Kührová, P., Šponer, J., et al. (2019). Local-to-global signal transduction at the core of a Mn2+ sensing riboswitch. *Nat. Commun.* 10:4304. doi: 10.1038/s41467-019-12230-5

Suddala, K. C., Rinaldi, A. J., Feng, J., Mustoe, A. M., Eichhorn, C. D., Liberman, J. A., et al. (2013). Single transcriptional and translational preQ1 riboswitches adopt similar pre-folded ensembles that follow distinct folding pathways into the same ligand-bound structure. *Nucleic Acids Res.* 41, 10462–10475. doi: 10.1093/nar/gkt798

Suddala, K. C., and Walter, N. G. (2014). Riboswitch structure and dynamics by smFRET microscopy. *Methods Enzymol.* 549, 343–373. doi: 10.1016/B978-0-12-801122-5.00015-5

Suddala, K. C., Wang, J., Hou, Q., and Walter, N. G. (2015). Mg^{2+} shifts ligand-mediated folding of a riboswitch from induced-fit to conformational selection. *J. Am. Chem. Soc.* 137, 14075–14083. doi: 10.1021/jacs.5b09740

Sung, H.-L., and Nesbitt, D. J. (2019). Single-molecule FRET kinetics of the Mn^{2+} riboswitch: evidence for allosteric Mg^{2+} control of "Induced-Fit" vs "Conformational Selection" folding pathways. *J. Phys. Chem. B* 123, 2005–2015. doi: 10.1021/acs.jpcb.8b11841

Takahashi, M. K., Watters, K. E., Gasper, P. M., Abbott, T. R., Carlson, P. D., Chen, A. A., et al. (2016). Using in-cell SHAPE-Seq and simulations to probe structure-function design principles of RNA transcriptional regulators. *RNA* 22, 920–933. doi: 10.1261/rna.054916.115

Thirumalai, D., Lee, N., Woodson, S. A., and Klimov, D. (2001). Early events in RNA folding. *Annu. Rev. Phys. Chem.* 52, 751–762. doi: 10.1146/annurev.physchem.52.1.751

Tinoco, I., Chen, G., and Qu, X. (2010). RNA reactions one molecule at a time. *Cold Spring Harb. Perspect. Biol.* 2:a003624. doi: 10.1101/cshperspect.a003624

Tollerson, R. II, and Ibba, M. (2020). Translational regulation of environmental adaptation in bacteria. *J. Biol. Chem.* 295, 10434–10445. doi: 10.1074/jbc.REV120.012742

Tomko, E. J., and Galburt, E. A. (2019). Single-molecule approach for studying RNAP II transcription initiation using magnetic tweezers. *Methods* 159–160, 35–44. doi: 10.1016/j.ymeth.2019.03.010

Uhm, H., Kang, W., Ha, K. S., Kang, C., and Hohng, S. (2018). Single-molecule FRET studies on the cotranscriptional folding of a thiamine pyrophosphate riboswitch. *Proc. Natl. Acad. Sci. U. S. A.* 115, 331–336. doi: 10.1073/pnas.1712983115

Walter, N. G. (2019). Biological pathway specificity in the cell-does molecular diversity matter? *Bioessays* 41:e1800244. doi: 10.1002/bies.201800244

Walter, N. G., and Engelke, D. R. (2002). Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist* 49, 199–203.

Washburn, R. S., Zuber, P. K., Sun, M., Hashem, Y., Shen, B., Li, W., et al. (2020). *Escherichia coli* NusG links the lead ribosome with the transcription elongation complex. *iScience* 23, 101352–101352. doi: 10.1016/j.isci.2020.101352

Waters, L. S., Sandoval, M., and Storz, G. (2011). The *Escherichia coli* MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. *J. Bacteriol.* 193, 5887–5897. doi: 10.1128/JB.05872-11

Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T., and Lucks, J. B. (2016). Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat. Struct. Mol. Biol.* 23, 1124–1131. doi: 10.1038/nsmb.3316

Wickiser, J. K., Cheah, M. T., Breaker, R. R., and Crothers, D. M. (2005). The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* 44, 13404–13414. doi: 10.1021/bi051008u

Widom, J. R., Nedialkov, Y. A., Rai, V., Hayes, R. L., Brooks, C. L. III, Artsimovitch, I., and Walter, N. G. (2018). Ligand modulates cross-coupling between riboswitch folding and transcriptional pausing. *Mol. Cell* 72, 541–552.e6. doi: 10.1016/j.molcel.2018.08.046

Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.* 10, 701–707. doi: 10.1038/nsb967

Yakhnin, A. V., Murakami, K. S., and Babitzke, P. (2016). NusG is a sequence-specific RNA polymerase pause factor that binds to the non-template DNA

within the paused transcription bubble. *J. Biol. Chem.* 291, 5299–5308. doi: 10.1074/jbc.M115.704189

Zaharik, M. L., Cullen, V. L., Fung, A. M., Libby, S. J., Kujat Choy, S. L., Coburn, B., et al. (2004). The *Salmonella enterica* serovar typhimurium divalent cation transport systems MntH and SitABCD are essential for virulence in an Nramp1G169 murine typhoid model. *Infect. Immun.* 72, 5522–5525. doi: 10.1128/IAI.72.9.5522-5525.2004

Zhang, J. (2020). Unboxing the T-box riboswitches-A glimpse into multivalent and multimodal RNA-RNA interactions. *Wiley Interdiscip. Rev. RNA* 11:e1600. doi: 10.1002/wrna.1600

Zhang, J., Lau, M. W., and Ferré-D'Amaré, A. R. (2010). Ribozymes and riboswitches: modulation of RNA function by small molecules. *Biochemistry* 49, 9123–9131. doi: 10.1021/bi1012645

Zhang, K., Li, S., Kappel, K., Pintilie, G., Su, Z., Mou, T. C., et al. (2019). Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat. Commun.* 10:5511. doi: 10.1038/s41467-019-13494-7

Zhao, B., Guffy, S. L., Williams, B., and Zhang, Q. (2017). An excited state underlies gene regulation of a transcriptional riboswitch. *Nat. Chem. Biol.* 13, 968–974. doi: 10.1038/nchembio.2427

Zhou, J., Ha, K. S., La Porta, A., Landick, R., and Block, S. M. (2011). Applied force provides insight into transcriptional pausing and its modulation by transcription factor NusA. *Mol. Cell* 44, 635–646. doi: 10.1016/j.molcel.2011.09.018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Chemistry

ORIGINAL RESEARCH
published: 05 March 2021
doi: 10.3389/fchem.2021.624937

Persistent Homology Metrics Reveal Quantum Fluctuations and Reactive Atoms in Path Integral Dynamics

Yunfeng Hu[1], *Phonemany Ounkham*[2], *Ondrej Marsalek*[3], *Thomas E. Markland*[4], *Bala Krishmoorthy*[5]* *and Aurora E. Clark*[2]*

[1]Department of Mathematics and Statistics, Washington State University, Pullman, WA, United States, [2]Department of Chemistry, Washington State University, Pullman, WA, United States, [3]Faculty of Mathematics and Physics, Charles University, Prague, Czech, [4]Department of Chemistry, Stanford University, Stanford, CA, United States, [5]Department of Mathematics and Statistics, Washington State University, Vancouver, WA, United States

OPEN ACCESS

Edited by:
Rene A. Nome,
State University of Campinas, Brazil

Reviewed by:
Nancy Makri,
University of Illinois at Urbana-Champaign, United States
Sérgio Ricardo Muniz,
University of São Paulo, Brazil

***Correspondence:**
Bala Krishnamoorthy
kbala@wsu.edu
Aurora E. Clark
auclark@wsu.edu

Specialty section:
This article was submitted to Physical Chemistry and Chemical Physics, a section of the journal Frontiers in Chemistry

Received: *01 November 2020*
Accepted: *22 January 2021*
Published: *05 March 2021*

Citation:
Hu Y, Ounkham P, Marsalek O, Markland TE, Krishmoorthy B and Clark AE (2021) Persistent Homology Metrics Reveal Quantum Fluctuations and Reactive Atoms in Path Integral Dynamics. Front. Chem. 9:624937. doi: 10.3389/fchem.2021.624937

Nuclear quantum effects (NQEs) are known to impact a number of features associated with chemical reactivity and physicochemical properties, particularly for light atoms and at low temperatures. In the imaginary time path integral formalism, each atom is mapped onto a "ring polymer" whose spread is related to the quantum mechanical uncertainty in the particle's position, i.e., its thermal wavelength. A number of metrics have previously been used to investigate and characterize this spread and explain effects arising from quantum delocalization, zero-point energy, and tunneling. Many of these shape metrics consider just the instantaneous structure of the ring polymers. However, given the significant interest in methods such as centroid molecular dynamics and ring polymer molecular dynamics that link the molecular dynamics of these ring polymers to real time properties, there exists significant opportunity to exploit metrics that also allow for the study of the fluctuations of the atom delocalization in time. Here we consider the ring polymer delocalization from the perspective of computational topology, specifically persistent homology, which describes the 3-dimensional arrangement of point cloud data, (i.e. atomic positions). We employ the Betti sequence probability distribution to define the ensemble of shapes adopted by the ring polymer. The Wasserstein distances of Betti sequences adjacent in time are used to characterize fluctuations in shape, where the Fourier transform and associated principal components provides added information differentiating atoms with different NQEs based on their dynamic properties. We demonstrate this methodology on two representative systems, a glassy system consisting of two atom types with dramatically different de Broglie thermal wavelengths, and ab initio molecular dynamics simulation of an aqueous 4 M HCl solution where the H-atoms are differentiated based on their participation in proton transfer reactions.

Keywords: path integral molecular dynamics, persistent homology, quantum delocalization, proton transfer, Wasserstein distances

1 INTRODUCTION

In recent years, path integral (PI) methods have seen significant application as a means to study nuclear quantum effects (NQEs), such as those arising from zero-point energy and tunneling, in chemical systems. In the imaginary time PI approach, each atom is described as a ring polymer composed of a set of beads where the adjacent beads interact via harmonic springs (Feynman and Hibbs, 1963; Chandler and Wolynes, 1981; Parrinello and Rahman, 1984). As the mass of the nuclei or the temperature of the system increases, the stiffness of the harmonic spring between the beads is increased, the polymer shrinks, and the ring polymer representation of the atom becomes more "localized". Conversely, for lower temperatures or for lighter particles, the weaker coupling between the beads allows the ring polymer to adopt a range of shapes reflecting the quantum mechanical delocalization in the atom's position. The quantum mechanical uncertainty in the atom's position is composed of the distribution of the centroid position and the ring polymer's spread.

NQEs have been demonstrated to affect hydrogen bond strengths, and thus the physicochemical, structural, and dynamic properties of protic solvents like water (Hardy et al., 2001; Morrone and Car, 2008; Habershon et al., 2009; Paesani and Voth, 2009; Pamuk et al., 2012; Harada et al., 2013; Ceriotti et al., 2016; Kim et al., 2017; Ruiz Pestana et al., 2018). The structure and dynamics of the species within acidic media has also received significant attention. For example, NQEs are observed to increase delocalization within protonated structures and as such enhance proton transfer within acidic systems (Ivanov et al., 2015; Marsalek and Markland, 2017; Napoli et al., 2018; Kawashima et al., 2018).

Several metrics have been proposed to characterize atomic delocalization in path integral systems. The imaginary-time mean square displacement (Berne and Thirumalai, 1986) evaluates a correlation function along the ring polymer. A set of shape metrics have also been introduced that characterize the anisotropy of the ring polymer in different chemical environments. The extension of the ring polymer is projected along a particular coordinate of interest e.g., in the case the proton transfer between two oxygen atoms projecting along the O-O coordinate (Benoit and Marx, 2005; Schran et al., 2018). By constructing idealized ellipsoid models of the bead density and their associated principal axes, an approximate shape of the distribution can be obtained (cigar-like or disk-like). Complementary, is the construction and analysis of the radius of gyration (R_g) of the ring polymer, defined as the average root mean squared distance of the replicas from the polymer center (or centroid), or related quantities such as the ratio of R_g values for different atoms, and gyration tensors (Markland et al., 2011, 2012; Dreschel-Grau and Marx, 2014; Schran et al., 2018). These shape metrics thus provide a route to analyze NQEs once a relevant atom has been identified. However, it leaves open the possibility to investigate a broader set of shape metrics to capture the changes in the *global* shape of the ring polymer and thus identify *a priori* atoms undergoing interesting changes in their "quantumness". In particular, these methods only utilize the static information obtained from a path integral molecular dynamics (PIMD) or path integral Monte Carlo (PIMC) sampling. While originally the dynamics obtained by PIMD was introduced purely as a tool to sample the quantum ensemble (Parrinello and Rahman, 1984; Tuckerman et al., 1993), methods such as centroid molecular dynamics (Cao and Voth, 1994; Jang and Voth, 1999) (CMD) and ring polymer molecular dynamics (Craig and Manolopoulos, 2004; Habershon et al., 2013) (RPMD) have demonstrated that for systems where the quantum coherence of the nuclei is rapidly damped that classical evolution under the imaginary time ring polymer Hamiltonian can be used to predict the dynamics of a quantum system. This opens the door to using the specific time series information of the global ring polymer configurations generated by these methods to identify quantum events.

Within the last decade, concepts from the mathematical field of algebraic topology have been combined with computational methods to characterize the global shape of data (Carlsson, 2009). Termed computational topology or topological data analysis (TDA), this field has seen rapid developments (Edelsbrunner and Harer, 2009). Persistent homology is a TDA method that produces compact summaries of the global shape and topology of sets of points in the form of barcodes (Ghrist, 2008). Given a collection of data sets (ring polymers representing atoms in our case), persistent homology provides an objective way to quantify and compare global shapes of the data sets by measuring distances between their barcodes. Statistical analyses on collections of such barcode distances may also be used to distinguish between different distributions. Here we apply persistent homology to study the time-dependent fluctuations of the ring polymers arising from RPMD and thermostatted RPMD (TRPMD) (Rossi et al., 2014) simulations and assess its ability to detect chemically meaningful information about NQEs.

In particular, we compare and contrast different shape and persistent homology metrics for two different chemical systems. The first system is a Kob–Andersen glass that contains two atom types of dramatically different quantum mechanical uncertainty. Not only is persistent homology able to elucidate variations in ring-polymer shape, but the Wasserstein distance between adjacent snapshots in time (which measures the change in the shape of the ring polymer), and its associated Fourier transform are found to be remarkably different for the two different atom types. In the second system, we examine the ability of the shape and persistent homology metrics to identify proton-transferring (PT) vs. non-PT H-atoms in an *ab initio* path integral simulation of an aqueous 4 M HCl solution. Again, a pronounced difference is observed in the Fourier transform of the Wasserstein distance, where PT H-atoms have significantly more fluctuation in shape than their non-PT counterparts. This observation paves the way for employing persistent homology in the study of a wide variety of chemical systems where NQEs are relevant, to not only identify atoms that have different nuclear behavior, but understand the change in quantum delocalization of an atom over time and along complex reaction coordinates.

2 COMPUTATIONAL METHODS

2.1 Static Atomic Uncertainty Metrics

In this work we consider several metrics that reflect the distribution of the distances of replicas relative to the centroid of the ring polymer as well as between the replicas themselves. Results for these quantities are included in the Supplementary Information for comparison and completeness. For a system of p replicas of N_A atoms, we denote the position of replica k of atom j as $\mathrm{r}_j^{(k)}$ and the position of the centroid of atom j as $\overline{\mathrm{r}}_j$. The gyration radius of atom j for any given configuration of the ring polymer is defined as the root mean square of the distance between the centroid and the replicas,

$$R_{g,j} = \sqrt{\frac{1}{P}\sum_{k=1}^{P}\left|\mathrm{r}_j^{(k)} - \overline{\mathrm{r}}_j\right|^2}. \tag{1}$$

This quantity is then most commonly averaged over all equivalent atoms and over the ensemble sampled by the simulation. More generally, we can examine the gyration tensor of atom j defined as

$$S_{(j)} = \frac{1}{P}\sum_{k=1}^{P}\left(\mathrm{r}_j^{(k)} - \overline{\mathrm{r}}_j\right)\left(\mathrm{r}_j^{(k)} - \overline{\mathrm{r}}_j\right). \tag{2}$$

Note that the subscript (j) of the tensor specifies index of the atom, and not its rank. The tensor can be represented by a symmetric 3×3 matrix whose off-diagonal entries give the xy, yz, and xz components of the tensor. Diagonalization of the 3×3 matrix representing the gyration tensor yields eigenvectors that describe the principal directions of the distribution of points and eigenvalues that describe the spread of the distribution in these directions. If we denote the ordered eigenvalues λ_x^2, λ_y^2, and λ_z^2, they can be used to obtain the gyration radius as $R_{g,j}^2 = \lambda_x^2 + \lambda_y^2 + \lambda_z^2$ and additional shape descriptors common in polymer and macromolecular science as follows (Mattice and Suter, 1994). The asphericity

$$b = \frac{3}{2}\lambda_z^2 - \frac{R_g^2}{2}, \tag{3}$$

describes the deviation from a fully symmetric distribution (for which $b = 0$), whereas the acylindricity

$$c = \lambda_y^2 - \lambda_x^2, \tag{4}$$

emphasizes symmetry about any two coordinate axes. Relative shape anisotropy is defined as

$$\kappa^2 = \frac{3}{2}\frac{\lambda_x^4 + \lambda_y^4 + \lambda_z^4}{\left(\lambda_x^2 + \lambda_y^2 + \lambda_z^2\right)^2} - \frac{1}{2}, \tag{5}$$

where a value of zero only occurs when all points are spherically symmetric, while a value of one is observed if all points are on a line. Metrics based on the inter-bead distances could include the distribution of all centroid to bead distances, or the distribution of pairwise distances between individual beads within the ring polymer, $\left\{\left|\mathbf{r}_j^{(k)} - \mathbf{r}_j^{(l)}\right| : 1 \le k \le l \le P\right\}$. The imaginary time mean-square displacement (iMSD) is also frequently employed to study a variety of aspects of path-integral simulations. Variations in the iMSD are characteristic of the spread of the ring polymer and also provide information about short-time dynamics (Tuckerman, 2010). It is calculated as

$$\Delta r^2(i\tau) = \left\langle \frac{1}{N_A P}\sum_{j=1}^{N}\sum_{k=1}^{P}\left|\mathrm{r}_j^{(k)} - \mathrm{r}_j^{(k+l)}\right|^2 \right\rangle, \tag{6}$$

where angle brackets denote averaging over the sampled ensemble, $\tau = l\beta\hbar/P$ and replica indices should be taken modulo p in the ring polymer.

2.2 Dynamic Shape Metrics From Persistent Homology

As an extension of the methods provided above, it is intriguing to combine the information contained within shape metrics of the polymer with its dynamic behavior. Toward this end we consider homology, the method from classical algebraic topology that captures how a space is connected. Herein, we first describe the general principles of homology and persistent homology, as well as known distance metrics to measure changes in topological features, as they are both applied to the ring polymer dynamics trajectories.

2.2.1 Persistent Homology as Metric of Shape

In the setting of homology directly amenable to computation, the space is modeled as a combinatorial object called the simplicial complex, which is a collection of vertices, edges, triangles, and higher order simplices glued together "nicely" (Munkres, 1984). For instance, a triangular mesh is a 2-dimensional simplicial complex. The ranks of the homology groups, termed *Betti numbers* and denoted by β_i in dimension i, have intuitive interpretations for small dimensions. In particular, β_0 counts the number of connected components in the object or space, β_1 counts the number of loops or holes, and β_2 counts the number of enclosed voids. Since we are interested in the global shapes of ring polymers that naturally form loops, we study the first Betti number β_1.

Persistent homology (Edelsbrunner et al., 2002) produces a more comprehensive picture (than simple homology) of the shape of space by constructing a sequence of growing simplicial complexes, rather than a single complex. Changes in β_i values are tracked across this sequence, and this information is presented in a compact form as a *barcode* (one barcode in each dimension i). Such persistent homology representations come with stability guarantees—small changes in input produce only small changes in the representation (Cohen-Steiner et al., 2007).

Given the collection of beads in a ring polymer, we consider a ball of radius r centered at each bead (**Figure 1**). We systematically grow the radius r from 0 to infinity (in this study, we measure r in Angstroms). Observe that as the radius grows, balls centered at beads that are close to each other will

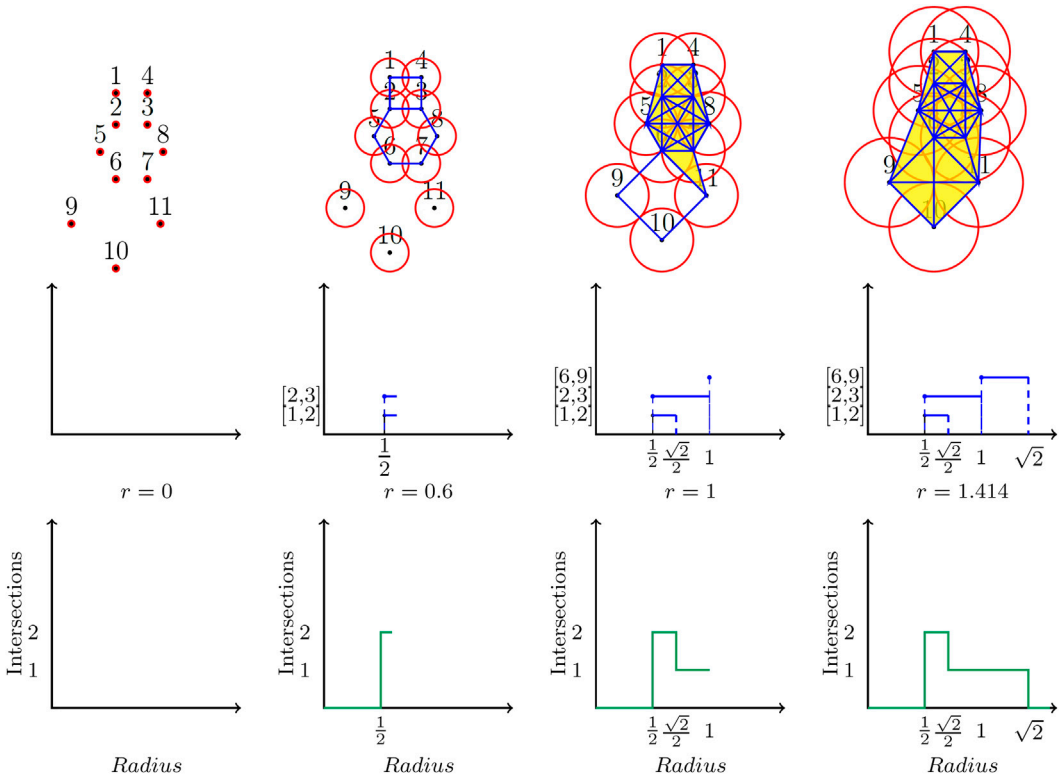

FIGURE 1 | Top Row: A ball of growing radius r (using the units of the coordinate system) is centered at each bead. Middle Row: A β_1 barcode records the birth and death of the holes. Each hole is "represented" by one of its edges, which is listed on the vertical axis ($[1,2]$, $[2,3]$, $[6,9]$). Bottom Row: A Betti sequence is constructed by sliding a vertical line at each radius and keeping track of the numbers of intersection of the line and the barcode.

intersect before those centered at beads that are farther apart. The intersections of these balls over the entire range of values of r capture all information about the global shape of the ring polymer. These intersections are used to define the Vietoris-Rips (VR) complex (Edelsbrunner and Harer, 2009) of the ring polymer. When $r = 0$, the VR complex consists of the individual points associated with the beads of the ring polymer, as the balls have no intersections. As r is increased, the intersection of a pair of balls is captured by adding the edge connecting the points to the VR complex. Triangles, tetrahedra, and higher order simplices are added to the VR complex to capture higher order intersections of balls. As the VR complex grows, small connected components merge into bigger connected components, and holes as well as voids appear and disappear.

The top row in **Figure 1** displays the construction of a VR complex for a ring polymer with 11 beads. 2D balls centered at the points (representing the beads) are shown as circles. At $r = 0$ (first figure on the left), none of the balls intersect, and hence the VR complex consists of individual points representing the beads. At $r = \frac{1}{2}$, edges added to capture pairwise intersections of the balls form two loops in the VR complex, shown as a square and a hexagon in the second figure. At $r = \frac{\sqrt{2}}{2}$, balls centered at beads one to four intersect pairwise, and hence simplices (tetrahedron 1,234, and its component triangles) are added to the VR complex to fill up the square loop, thus "killing" this feature of the topology. The barcode in the second row of **Figure 1** records the birth and death of each loop in the VR complex as the radius is increased. The hexagonal loop formed by beads 3, 2, 5, 6, 7, and 8, for instance, is born, (i.e. formed) at $r = \frac{1}{2}$, and dies, (i.e. is closed up) at $r = 1$. The complete β_1 barcode is shown in the fourth (last) figure, with all holes closed up at $r = \sqrt{2}$. We can use this barcode as the representation of the shape of the ring polymer.

2.2.2 Fluctuations in Shape

We could compare the shapes of two ring polymers by comparing their β_1 persistence barcodes. To quantify this comparison, we want to compute a distance between the barcodes. We are using the word *distance* in the mathematical sense: a distance is a function that accepts two distributions as input, and returns a nonnegative real number which measures how close the two distributions are. To this end, we want to convert each barcode to a vector with the same number of entries, and then compute the distance between the corresponding vectors. We build a *Betti sequence* by sliding a vertical line across each radius value and keeping track of the intersection of the line and the barcode (bottom row, **Figure 1**). For example, a vertical line at $r = \frac{1}{2}$ intersects the barcode twice as there are 2 bars at $r = \frac{1}{2}$. Thus the Betti sequence is constructed by recording the number of bars at each radius. Hence, denoting $h(r)$ to be the number of intersections at radius r, we define the Betti sequence as

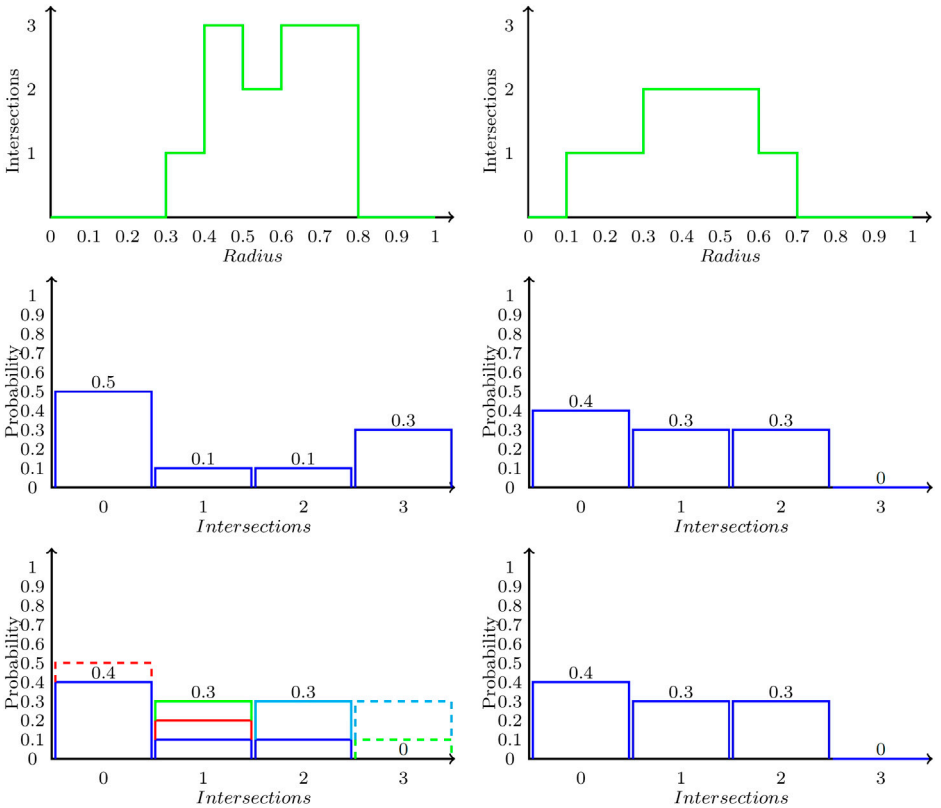

FIGURE 2 | Illustration of Wasserstein distance computation between two Betti sequence probability distributions. Top row shows two Betti sequences. The second row shows the corresponding Betti sequence probability distributions P **(left)** and Q **(right)**. The last row shows the optimal transformation of P into Q, which consists of three steps: moving the red box from bin 0 to one contributing $0.1 \times |0 - 1| = 0.1$, the green box from bin three to one contributing $0.1 \times |1 - 3| = 0.2$, and the cyan box from bin three to two contributing $0.2 \times |2 - 3| = 0.2$. Therefore, the Wasserstein distance is $0.1 + 0.2 + 0.2 = 0.5$.

$\{h(r)\}_{r \in [0,1]}$. The choice of upper bound of $r = 1$ is motivated by the observation that for all ring polymers we considered in this study, the holes were closed well before the radius value of $r = 1$ in each chemical system ($r = 1$ Bohr for the Kob–Andersen glass and $r = 1$ Å for the proton transfer example). In other words, $h(r) = 0$ for $r \geq 1$. The resolution at which we cover $[0, 1]$ is guided by the barcodes—we increment r in steps fine enough to distinguish births and deaths of each bar. We used 400 steps for the Kob–Anderson glass system and 1,000 steps for the second system studying proton transfer.

In the final step, we normalize this Betti sequence to create the *Betti sequence probability distribution* indexed by the number of intersections. We use the maximum number of intersections observed in any ring polymer as the common number of intersections used in all cases, thus standardizing the lengths of all Betti sequence probability distributions. Each such distribution adds up to a total probability of 1, by definition. Note that there are other vectorizations of persistence barcodes or diagrams known, e.g., persistence landscapes (Bubenik, 2015) and persistence images (Adams et al., 2017). These constructions are arguably more general than our Betti sequences. At the same time, we found the Betti sequences simpler to compute, and they served our purpose in this study of ring polymers efficiently.

Comparison of two different ring polymer shapes can be made by calculating the Wasserstein distance (WD) between the Betti sequence probability distributions. Stability results have recently been presented for WD of persistent barcodes (Skraba and Turner, 2020). The Wasserstein distance (Villani, 2009), also termed the *Earth Mover's Distance* (Rubner et al., 2000) is a metric that measures the distance between the two normalized distributions as the cost of transforming one into another. We present the definition and then illustrate steps in the WD computation using **Figure 2**. More generally, let $K = \{K(i)\}_{i=1}^{p}$ and $L = \{L(j)\}_{j=1}^{q}$ be two normalized probability distributions. Let d_{ij} be the distance between the bins i in K and j in L. In the formal setting of WD, this distance could be measured in units of length between actual piles of earth. Subsequently, the WD is also specified in units of length by default. But more generally, d_{ij} could be set as the difference between probability measures in the corresponding bins i and j, and hence need not be measured in units of length (*vide infra*). We consider all possible transformations of K into L. We represent such a transformation by the matrix of values $[f_{ij}]$ with f_{ij} denoting the mass, (i.e. probability) transferred from bin i in K to bin j in L. The WD between K and L is given by the optimal objective function value of the following optimization problem.

$$\min \quad \sum_{i=1}^{k}\sum_{j=1}^{l} d_{ij}f_{ij}, \tag{7}$$

$$\sum_{j=1}^{l} f_{ij} \le K(i), i = 1, \ldots, k, \tag{8}$$

$$\sum_{i=1}^{k} f_{ij} \le L(j), j = 1, \ldots, l, \tag{9}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{l} f_{ij} = 1, \tag{10}$$

$$f_{ij} \ge 0, i = 1, \ldots, k, \; j = 1, \ldots, l. \tag{11}$$

Constraints 8) and 9) specify that we cannot transport more probability out of a bin than what is available. **Equation 10** ensures we transform all of K into L. The Wasserstein distance can be computed efficiently by solving this optimization problem as a transportation problem (Ahuja et al., 1993).

We illustrate the Wasserstein distance computation on two example Betti sequence probability distributions in **Figure 2**. The two Betti sequences are presented in the top row of **Figure 2** as the number of intersections with the β_1 barcode as a function of the radius of the balls (used to construct the VR complex). The middle row of **Figure 2** presents the corresponding Betti sequence probability distributions. Recall that we normalize the Betti sequence to create the *Betti sequence probability distribution* indexed by the number of intersections. Each such distribution adds up to a total probability of 1, by definition. We then compute the Wasserstein distance to transform the probability K on the left to probability L on the right. The third row of **Figure 2** illustrates the optimal way to redistribute the area associated with the probability distributions to make the two equal. This transformation is effected by first moving the $f_{01} = 0.1$ probability from intersection 0 to intersection 1 (the red box). This move contributes $0.1 \times |0 - 1| = 0.1$ to the overall distance. In the next step, the $f_{32} = 0.2$ probability is moved from intersection 3 to intersection 2 (the cyan box). This move contributes $0.2 \times |2 - 3| = 0.2$ to the total distance. Finally, we move the $f_{31} = 0.1$ probability at intersection 3 to intersection 1 (the green box), which contributes $0.1 \times |1 - 3| = 0.2$ to the distance. Hence the Wasserstein distance between the two Betti sequences is given as $0.1 + 0.2 + 0.2 = 0.5$.

2.2.3 Application to Ring Polymers

To study the dynamic fluctuations to ring polymer shape, the β_1 barcode for each ring polymer atom representation is determined at time t and the compared to at time $t + 1$. In our application to ring polymer shape comparison, we set $k = l = B_1$. This corresponds to the largest value observed in any Betti sequence in the entire data set, i.e., the largest number of bars in the β_1 barcode of any ring polymer (at any radius value). The Wasserstein distance between time sequential Betti sequence probability distributions is then determined. For a trajectory with N snapshots, we compute the Wasserstein distance vector with $N - 1$ entries, with the tth entry specifying the Wasserstein distance between snapshots t and $t + 1$. The value d_{ij} is set to $|i - j|$.

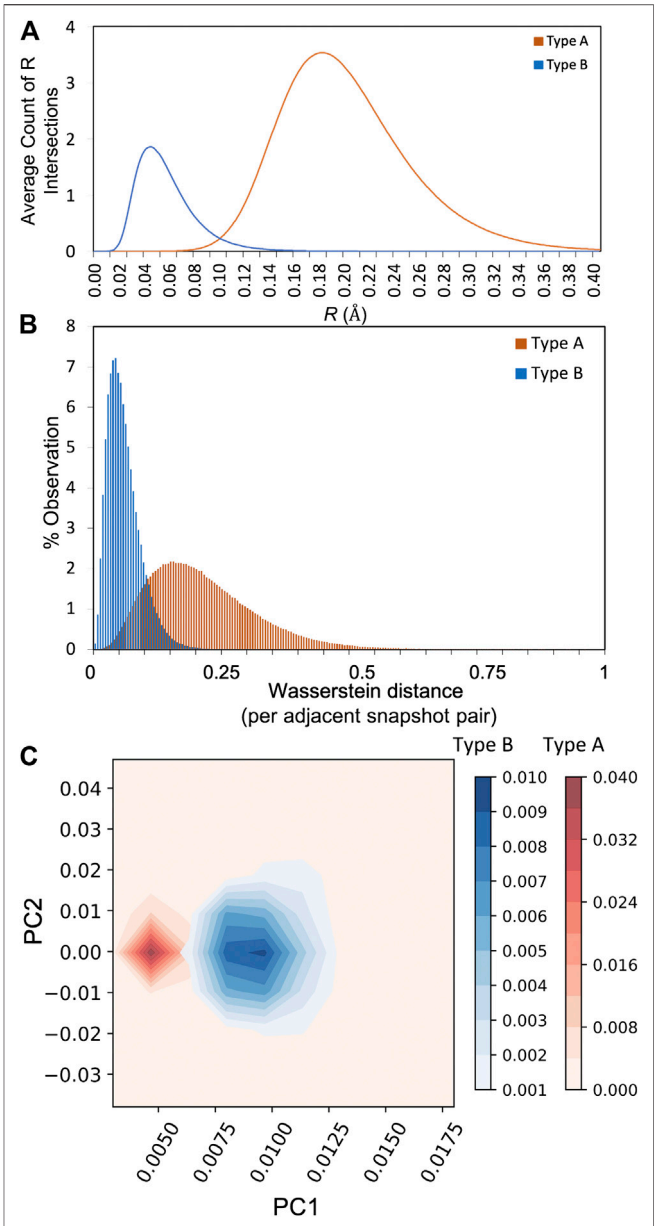

FIGURE 3 | (A) The average Betti sequence distribution of all type A and type B atoms **(B)** Distribution of Wasserstein distances between adjacent snapshots in time observed over the entire simulation trajectory **(C)** Principal components analysis capturing 90% of the total variance in the datasets using trajectory windows of ± 20 snapshots (12 ps).

$$WD = \begin{bmatrix} wd_{1,2} & wd_{2,3} & \cdots & wd_{N-1,N} \end{bmatrix}. \tag{12}$$

Finally, in recognizing that the Wasserstein distance vector WD captures the fluctuation in shape over time by measuring distances between adjacent snapshots, we performed a Fourier transform of WD, followed by principal component analysis. We then used the coefficients for the two largest frequencies for comparing the characteristic fluctuations in ring polymer shape across different chemical systems.

We have made the Python code and sample data available for the main calculations in a GitHub repository (Hu et al., 2020).

3 RESULTS AND DISCUSSION

3.1 System 1: The Kob–Andersen Glass

The RPMD simulation of the Kob–Andersen glass forming system was taken from Markland et al. (2011, 2012) and contained atoms types *A* and *B* with different degrees of atomic delocalization. This was determined by their respective *de Broglie* thermal wavelength $\Lambda = \sqrt{\frac{\beta\hbar^2}{m}}$, where the mass for particle *A* was a magnitude smaller than particle *B* and consequently particle *A* exhibited more quantum fluctuations than its counterpart. The cubic box consisted of 172 type *A* and 44 type *B* particles for a total of 216 particles. The path integrals were discretized into *p* Trotter slices (beads) with $P = 64$ for both particles, with 3,000 ps (5 x 10^6 steps with 0.6 fs time steps) of configurations saved every 1,000 steps (0.6 ps) yielding a total of 5,000 snapshots.

3.1.1 Static Atomic Uncertainty Metrics

For the extreme case of the delocalized type *A* and highly localized type *B* atoms in the Kob–Andersen glass, all metrics that evaluate shape are highly differentiated (**Supplementary Table S1**). In the case of the bead centroid metrics, the radii of gyration are 0.19 and 0.017 Å for the *A* and *B* atoms, respectively. These values belie distributions in the individual distances of the polymer beads from the centroid that are statistically very different and have nominal overlap, as illustrated in **Supplementary Figure S1A**. Analysis of the R_g tensors indicates that the *A* atoms are much more aspherical ($b = 7.67 \cdot 10^{-2}$, **Eq. 3**) than *B* atoms ($b = 5.84 \cdot 10^{-3}$), and similarly more acylindrical (**Supplementary Table S1**). Neither atom type is anisotropic, having κ^2 values of 10^{-4} – 10^{-5} (**Eq. 5**). The pair-wise bead-bead distance distributions have little overlap for *A* and *B* atoms (**Supplementary Figure S1B**), which in turn is reflected in the imaginary time mean square displacement for the *A* and *B* atoms as demonstrated in **Supplementary Figure S2**.

3.1.2 Homology and Persistent Homology Metrics

The alternative shape metric based on the Betti sequence of all *A* and *B* atoms is presented in **Figure 3**. The Betti sequence distribution, which captures the number of intersecting radii as a function of *R*, exhibits two distinctly different distributions in this case. The distribution of Wasserstein distances for the type *A* and *B* atoms are significantly different, where the *A* atoms explore a much broader shape space than *B* (meaning there is more variation in the Betti sequence distribution from one snapshot to the next for type *A*). Further, the magnitude of the Wasserstein distances are much larger for type *A* relative to *B*, meaning that from one snapshot to the next there are large changes to the shapes that the *A* ring polymers adopt. This is further demonstrated by monitoring the time evolution of Wasserstein distances, as shown in **Supplementary Figure S3**. To quantify the fluctuation in Wasserstein distances, the Fourier transform was studied for type *A* and *B* atoms, followed by principal component analysis. As illustrated in **Figure 3C**, the first principal components (PC1) for both atom types are clearly well separated, and along with PC2, are able to explain 90% of the variance. The Fourier transform was examined with different lengths of sampling duration, with no appreciable changes observed (**Supplementary Figure S4**). The Kob–Andersen glass forming system thus represents a proof of principle that a broader suite of shape metrics may be suitable for understanding shapes of ring-polymer representations of atoms, and that persistent homology metrics can reveal identifying characteristics of atoms with dramatically different quantum behavior.

3.2 System 2: Aqueous 4 M Hydrochloric Acid

Given the effectiveness of the Wasserstein distance and its Fourier transform in distinguishing atoms in the Kob–Andersen Lennard Jones system, it is thus pertinent to examine the ability of such new methods to reveal varying properties of atoms with much closer nuclear quantum behavior. To test an extreme case, where normal distance-based shape metrics do not indicate significant variations in quantum mechanical behavior, we turn to the identification of proton-transferring H-atoms in a 4 M HCl aqueous solution.

The TRPMD simulation of the 4 M HCl solution was taken from the work of Napoli et al. (2018a). The cubic box of length 14.926 Å consisted of 102 water molecules, eight excess H^+ and Cl^-. The path integrals were discretized into $P = 32$ Trotter slices for all atoms. A total of 123 ps of path integral simulations using the revPBE0-D3 hybrid functional (Adamo and Barone, 1999; Grimme et al., 2010; Goerigk and Grimme, 2011) with a 2 fs time step and sampling frequency were analyzed for a total of 61,555 snapshots. Analysis of topological properties of the ring polymers focused on the H-atoms, where they were split into two primary groups based on whether or not they underwent proton transfer during the course of the trajectory. To identify transferring H-atoms, a sequential filtration process was employed. First, the centroids of all O- and H-atoms were examined, wherein zundel cations were identified by employing an O–H distance criterion of 1.3 Å (Napoli et al., 2018; Schran et al., 2018). Within this set, time windows of 40–200 fs were then examined and any changes to the connectivity of the H-atom to different O-atoms were examined. Changes to O-atoms bonded to the H-atom was then identified, (i.e. an H-atom is connected to O_1, then forms a zundel with the H-atom shared between O_1 and O_2, and then the H-atom forms a single bond with O_2), and subsequently the connectivity of each bead of the H-atom ring polymer was analyzed to understand the timescale associated with all 32 beads changing O-atom partners. An average time of 40 fs was observed for the complete proton transfer (PT) of all 32 beads. Using this criterion (that all 32 beads must change O-atom partners) within a 40 fs time window (+ 20 fs relative to the center of the time window) a total of 2,283 proton transfer events were identified during the simulation trajectory. Non-PT H-atoms were identified as those wherein the 32 beads did not change their covalent connectivity during a 40 fs time window. In total, there are 1,267 windows of time where H-atoms do not undergo *p*T. This

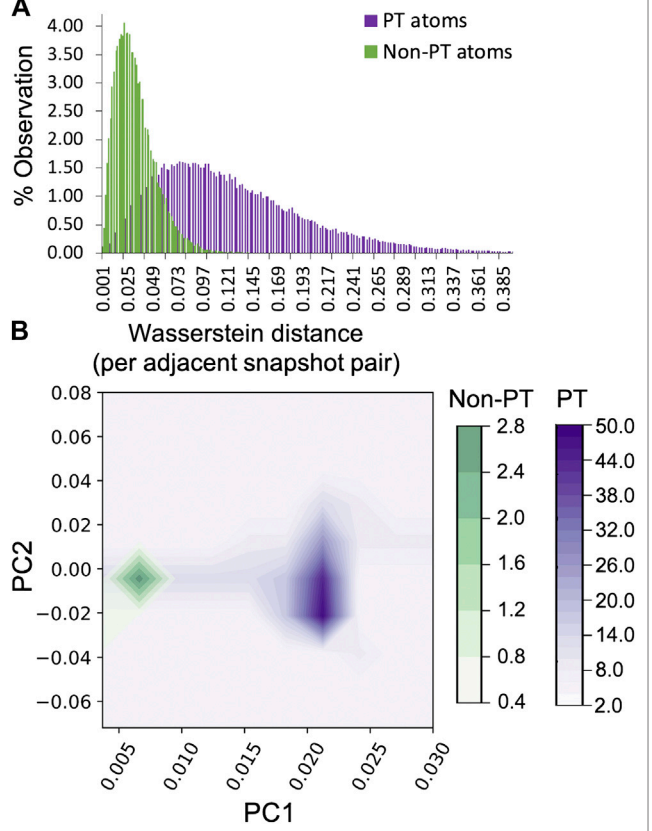

FIGURE 4 | Analysis of the Wasserstein distance characteristics of adjacent t and t + 1 snapshots for proton transferring (PT) and non-PT H-atom ring polymers **(A)** The distribution of Wasserstein distances observed over all PT and non-PT ring polymers **(B)** Principle component analysis of the PT and non-PT H-atoms capturing 90% of the total variance in the datasets.

yields a total of 98,638 snapshots and is comparable to the 2,283 × 40 = 91,320 snapshots in the proton-transferring dataset.

3.2.1 Shape Metrics

Comparison of the R_g of H-atoms undergoing PT and those chemically inert H-atoms yields nearly identical values of 0.1211 Å and 0.1206 Å, respectively (**Supplementary Table S3**). This is in good agreement with prior observations of nearly identical R_g in other room-temperature proton transferring H-atoms in formic acid (Ivanov et al., 2015). Similarly there is little discrimination in shape factors, with the shape anisotropy only being slightly larger for proton transferring atoms (κ^2 of $1.36 \cdot 10^{-2}$) relative to non-transferring counterparts (κ^2 of $8.03 \cdot 10^{-3}$; **Supplementary Table S2**). Analysis of the underlying distribution of distances between all beads and the centroid reveals significant overlap (**Supplementary Figure S5**), and indeed, applications of Student's t-test reveal the centroid distance distributions to be statistically equivalent (**Supplementary Table S3**).

Interestingly, a slightly better identification of PT and non-PT H-atoms is obtained using the bead–bead pairwise distance distribution, which passes the student t-test, however, the average values are still nearly identical, at 0.2155 Å and 0.2151 Å. A more clear delineation is further obtained in the distributions of the Betti sequences of the H-atoms undergoing proton transfer relative to the unreactive atoms (**Supplementary Figure S5**), where the average distances for reactive and unreactive atoms is 0.1286 vs. 0.1263, respectively, also being statistically significant and passing Student's t-test (**Supplementary Table S3**). This suggests that the Betti sequence distribution, which contains more information about the ring-polymer shape, is a more sensitive shape metric than the methods based on distances within the ring polymer.

3.2.2 Persistent Homology Metrics

As in the Kob–Andersen glass, the Wasserstein distances between PT and non-PT H-atoms in adjacent frames were then examined to identify the fluctuation in shape from one snapshot in the trajectory to another. Unlike the ensemble shape distributions for these atoms, the distribution of the Wasserstein distances for the two sets of H-atoms *is very well-separated* (**Figure 4A**). The PT atoms exhibit a larger range of accessible shapes, (i.e. the distribution broad), and the fluctuation in shape (which leads to a large cost for changing the Betti distribution from one snapshot to the next) is significantly larger than in the case of non-PT atoms, sampled within a similar 40 fs time window. The fluctuations themselves are illustrated in **Supplementary Figure S6**. This illustrates that the time dependent fluctuations in shape may be an alternative metric for identifying variations in nuclear quantum behavior between atom sets. Perhaps just as important is the observation that the NQEs may manifest themselves differently for the static vs. dynamic features of ring polymers. The Fourier transform of the Wasserstein distances was then performed and the two dominant principal components plotted in **Figure 4B**. In contrast to the Kob–Andersen glass system, the correlation between the first and second principal components is much higher, however they are still clearly differentiated for the PT vs. non-PT atoms. In combination, these data demonstrate that metrics based upon the fluctuation of atom delocalization in time are highly sensitive to quantum behavior, being able identify such phenomena when traditional ensemble averaged shape metrics based upon distance criterion (like the gyration radius) are inadequate.

4 CONCLUSION

As the pervasiveness of path integral methods increases within the applied computational chemistry community, new tools are needed to identify atoms where NQEs may be relevant and understand the role of NQEs in reactive processes. While a few metrics exist that identify variations in atomic position uncertainty, they are optimal for systems where the difference in uncertainty is large between different atom types. This work expands the set of available tools to study the shape of the delocalization of atomic positions, the uncertainty associated with NQEs, using persistent homology. Further, the chemical information associated with the time evolution of shape has not been investigated previously. Here, we demonstrate that compared to static distributions the time-

dependent persistent homology metrics can provide a clearer way to identify atoms where NQEs are important and to distinguish atoms of different kinds or in different chemical environments. Reactive hydrogen atoms during proton transfer exhibit much larger fluctuations in time of their ring polymer shape than non-reactive counterparts. We believe that the utility of metrics that capture the fluctuations of the atom delocalization in time is generalizable to other reactive chemical systems, and in turn that this provides a means to extract information on reactivity from the quantum behavior of the system, a topic that has not received consideration within the literature.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AC conceived the project, oversaw research and supported manuscript drafting, revisions, and final edits. BK oversaw research and supported manuscript drafting, revisions, and final edits. YH performed persistent homology analysis and supported manuscript drafting. PO performed parsing of HCl data for persistent homology analysis and supported manuscript drafting. OM and TM performed the RPMD simulations and participated in manuscript revisions and final edits.

FUNDING

This work was supported by the Department of Energy, Office of Basic Energy Sciences CTC and CPIMS programs, under Award Number DE-SC0014437. BK acknowledges funding from the National Science Foundation through grants DBI-1661348 and DMS-1819229.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2021.624937/full#supplementary-material.

REFERENCES

Adamo, C., and Barone, V. (1999). Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* 110, 6158–6170. doi:10.1063/1.478522

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., et al. (2017). Persistence images: a stable vector representation of persistent homology. *J. Machine Learn. Res.* 18, 1–35.

Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1993). *Network flows: theory, algorithms, and applications*. London, United Kingdom: Pearson.

Benoit, M., and Marx, D. (2005). The shapes of protons in hydrogen bonds depend on the bond length. *Chemphyschem* 6, 1738–4110. doi:10.1002/cphc.200400533

Berne, B. J., and Thirumalai, D. (1986). On the simulation of quantum systems: path integral methods. *Annu. Rev. Phys. Chem.* 37, 401. doi:10.1146/annurev.pc.37.100186.002153

Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Machine Learn. Res.* 16, 77–102.

Cao, J., and Voth, G. A. (1994). The formulation of quantum statistical mechanics based on the Feynman path centroid density.V.quantum instantaneous normal mode theory of liquids. *J. Chem. Phys.* 101, 6184. doi:10.1063/1.468400

Carlsson, G. (2009). Topology and data. *Bull. Amer. Math. Soc.* 46, 255–308. doi:10.1090/s0273-0979-09-01249-x

Ceriotti, M., Fang, W., Kusalik, P. G., McKenzie, R. H., Michaelides, A., Morales, M. A., et al. (2016). Nuclear quantum effects in water and aqueous systems: experiment, theory, and current challenges. *Chem. Rev.* 116, 7529–7550. doi:10.1021/acs.chemrev.5b00674

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.* 37, 103–120. doi:10.1007/s00454-006-1276-5

Craig, I. R., and Manolopoulos, D. E. (2004). Quantum statistics and classical mechanics: real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.* 121, 3368. doi:10.1063/1.1777575

Dreschel-Grau, C., and Marx, D. (2014). Quantum simulation of collective proton tunneling in hexagonal ice crystals. *Phys. Rev. Lett.* 112, 148302.

Edelsbrunner, H., and Harer, J. L. (2009). *Computational topology an introduction*. Providence, RI, United States: American Mathematical Society.

Edelsbrunner, H., Letscher, D., and Zomorodian, A. (2002). Topological persistence and simplification. *Discrete Comput. Geom.* 28, 511–533. doi:10.1007/s00454-002-2885-2

Fernández-Serra, M., and Rahman, A. (1984). Study of an F center in molten KCl. *J. Chem. Phys.* 80, 860. doi:10.1063/1.446740

Feynman, R. P., and Hibbs, A. R. (1963). *Quantum mechanics and path integrals*. 1 edn. New York, NY, United States: McGraw-Hill.

Ghrist, R. (2007). Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc.* 45, 61–76. doi:10.1090/S0273-0979-07-01191-3

Goerigk, L., and Grimme, S. (2011). A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* 13, 6670–6688. doi:10.1039/C0CP02984J

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132, 154104. doi:10.1063/1.3382344

Habershon, S., Markland, T. E., and Manolopoulos, D. E. (2009). Competing quantum effects in the dynamics of a flexible water model. *J. Chem. Phys.* 131, 024501. doi:10.1063/1.3167790

Habershon, S., Manolopoulos, D. E., Markland, T. E., and Miller, T. F. (2013). Ring-polymer molecular dynamics: quantum effects in chemical dynamics from classical trajectories in an extended phase space. *Annu. Rev. Phys. Chem.* 64, 387–413. doi:10.1146/annurev-physchem-040412-110122

Harada, Y., Tokushima, T., Horikawa, Y., Takahashi, O., Niwa, H., Kobayashi, M., et al. (2013). Selective probing of the OH or OD stretch vibration in liquid water using resonant inelastic soft-x-ray scattering. *Phys. Rev. Lett.* 111, 193001. doi:10.1103/PhysRevLett.111.193001

Ivanov, S. D., Grant, I. M., and Marx, D. (2015). Quantum free energy landscapes from ab initio path integral metadynamics: double proton transfer in the formic acid dimer is concerted but not correlated. *J. Chem. Phys.* 143, 124304. doi:10.1063/1.4931052

Jang, S., and Voth, G. A. (1999). A derivation of centroid molecular dynamics and other approximate time evolution methods for path integral centroid variables. *J. Chem. Phys.* 111, 2371. doi:10.1063/1.479515

Kawashima, Y., Sawada, K., Nakajima, T., and Tachikawa, M. (2018). A path integral molecular dynamics study on intermolecular hydrogen bond of acetic acid-arsenic acid anion and acetic acid-phosphoric acid anion clusters. *J. Comput. Chem.* 40, 172–180. doi:10.1002/jcc.25562

Kim, K. H., Pathak, H., Spah, A., Perakis, F., Mariedahl, D., Sellberg, J. A., et al. (2017). Nuclear quantum effects in water. *PRL* 119, 075502. doi:10.1103/physrevlett.119.075502

Markland, D., and Wolynes, P. G. (1981). Exploiting the isomorphism between quantum theory and classical statistical mechanics of polyatomic fluids. *J. Chem. Phys.* 74, 4078. doi:10.1063/1.441588

Markland, T. E., Morrone, J. A., Berne, B. J., Miyazaki, K., Rabani, E., and Reichman, D. R. (2011). Quantum fluctuations can promote or inhibit glass formation. *Nat. Phys.* 7, 134–137. doi:10.1038/nphys1865

Markland, T. E., Morrone, J. A., Miyazaki, K., Berne, B. J., Reichman, D., and Rabani, E. (2012). Theory and simulations of quantum glass forming liquids. *J. Chem. Phys.* 136, 074511. doi:10.1063/1.3684881

Marsalek, O., and Markland, T. E. (2017). Quantum dynamics and spectroscopy of ab initio liquid water: the interplay of nuclear and electronic quantum effects. *J. Phys. Chem. Lett.* 8, 1545–1551. doi:10.1021/acs.jpclett.7b00391

Mattice, W. L., and Suter, U. W. (1994). *Conformational theory of large molecules*. Hoboken, NJ, United States: Wiley-Interscience.

Morrone, J. A., and Car, R. (2008). Nuclear quantum effects in water. *Phys. Rev. Lett.* 101, 017801. doi:10.1103/PhysRevLett.101.017801

Munkres, J. R. (1984). *Elements of algebraic topology*. Menlo Park, CA, United States: Addison–Wesley Publishing Company.

Napoli, J. A., Marsalek, O., and Markland, T. E. (2018). Decoding the spectroscopic features and time scales of aqueous proton defects. *J. Chem. Phys.* 148, 222833. doi:10.1063/1.5023704

Paesani, F., and Voth, G. A. (2009). The properties of water: insights from quantum simulations. *J. Phys. Chem. B.* 113, 5702. doi:10.1021/jp810590c

Pamuk, B., Soler, J. M., Ramírez, R., Herrero, C. P., Stephens, P. W., Allen, P. B., et al. (2012). Anomalous nuclear quantum effects in ice. *Phys. Rev. Lett.* 108, 193003. doi:10.1103/PhysRevLett.108.193003

Rossi, M., Ceriotti, M., and Manolopoulos, D. E. (2014). How to remove the spurious resonances from ring polymer molecular dynamics. *J. Chem. Phys.* 140, 234116. doi:10.1063/1.4883861

Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The Earth Mover's Distance as a metric for image retrieval. *Int. J. Comput. Vis.* 40, 99–121. doi:10.1023/A:1026543900054

Ruiz Pestana, L., Marsalek, O., Markland, T. E., and Head-Gordon, T. (2018). The quest for accurate liquid water properties from first principles. *J. Phys. Chem. Lett.* 9, 5009–5016. doi:10.1021/acs.jpclett.8b02400

Schran, C., Brieuc, F., and Marx, D. (2018). Converged colored noise path integral molecular dynamics study of the zundel cation down to ultralow temperatures at coupled cluster accuracy. *J. Chem. Theor. Comput.* 14, 5068–5078. doi:10.1021/acs.jctc.8b00705

Sacher, Y., Krishnamoorthy, B., and Clark, A. (2020). *Persistent homology computations on atom ring polymers*. https://gitlab.com/aurora-clark-public/pershomol_ringpolymers.

Shin, E. H., Zygar, A., and Zeidler, M. D. (2001). Isotope effect on the translational and rotational motion in liquid water and ammonia. *J. Chem. Phys.* 114, 3174. doi:10.1063/1.1340584

Skraba, P., and Turner, K. (2020). *Wasserstein stability for persistence diagrams*, arXiv. 2006. 16824.

Tuckerman, M. E., Berne, B. J., Martyna, G. J., and Klein, M. L. (1993). Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *J. Chem. Phys.* 99, 2796. doi:10.1063/1.465188

Tuckerman, M. E. (2010). *Statistical mechanics: theory and molecular simulation*. Oxford, United Kingdom: Oxford Graduate Texts (Oxford University Press.

Villani, C. (2009). *Optimal transport old and new*. Berlin, Germany: Springer-Verlag Berlin Heidelberg.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership