



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

David F. Burke<sup>1\*</sup>, Rosemary G. Mantell<sup>2</sup>, Catherine E. Pitt<sup>2</sup> and David J. Wales<sup>2\*</sup>

<sup>1</sup> EMBL-EBI, Wellcome Genome Campus, Hinxton, United Kingdom, <sup>2</sup> Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

## 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

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

## 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; N6e 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.

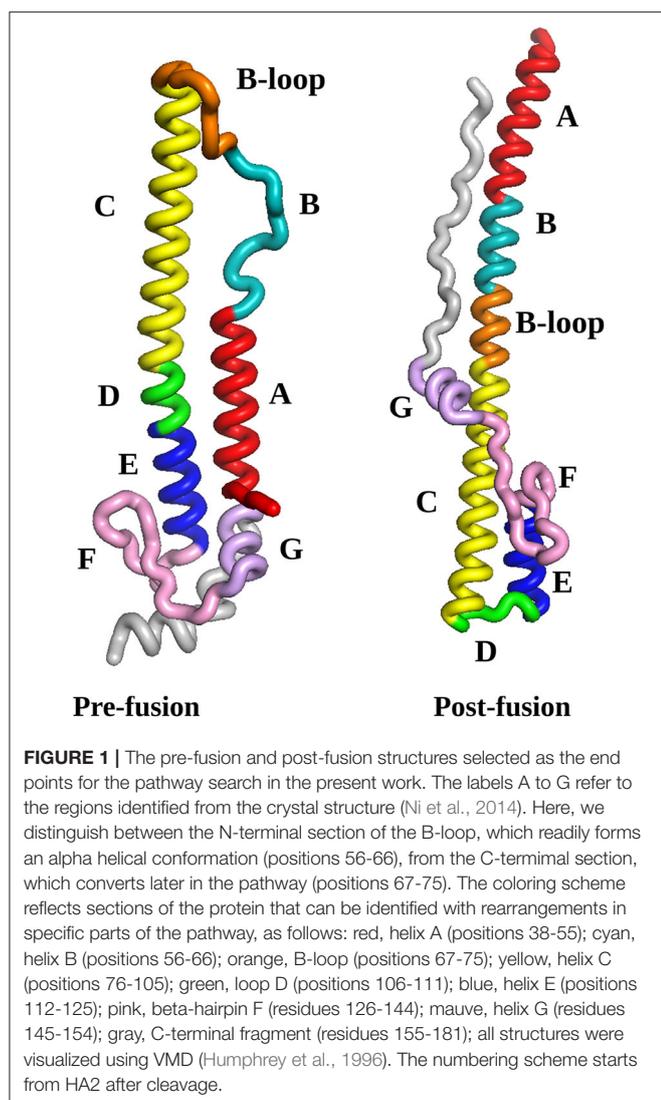
### 2.2. Force Field

We used the OPTIM<sup>1</sup> 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 (G6tz 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.

<sup>1</sup>OPTIM: A Program for Geometry Optimisation and Pathway Calculations. Available online at: <http://www.wales.ch.cam.ac.uk/software.html>.



### 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; N6e 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<sup>2</sup> (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 J6nsson, 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 J6nsson, 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<sup>3</sup> 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<sup>-6</sup> 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; R6der 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 (R6der 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 (R6der and Wales, 2018). In brief, QCI works by defining an effective

<sup>2</sup>CudaLBFGS. Available online at: <https://github.com/jwetzl/CudaLBFGS> (accessed Oct 1, 2013).

<sup>3</sup>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; Coutsiaris 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 (GoPERMDIST), 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 5 Å from the center of coordinates of the permuting set, and an alignment tolerance of 0.5 Å 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<sup>2</sup>. 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<sup>2</sup>, respectively. These values were found to perform well in systematic benchmarks for smaller systems (Röder and Wales, 2020).

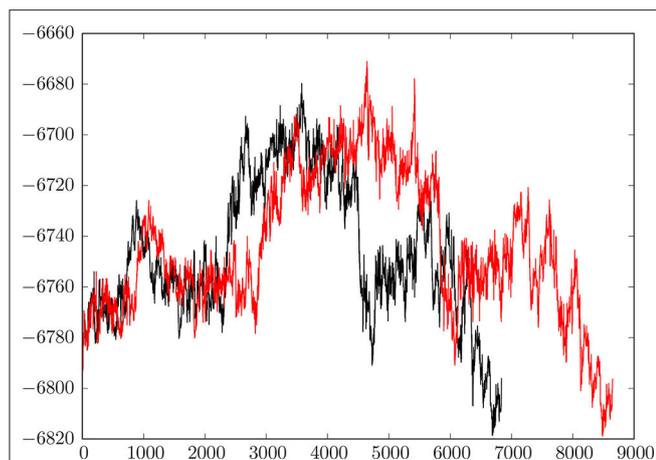
### 3.2. Refining the Kinetic Transition Network

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; N6e 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.<sup>4</sup> 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  $\beta$  is to adjacent minimum  $\alpha$ .  $P_{\alpha\beta}$  is calculated as  $k_{\alpha\beta} / \sum_{\gamma} k_{\gamma\beta}$ , where the sum is over all the pathways out of  $\beta$  and  $k_{\alpha\beta}$  is the rate constant for transitions from  $\beta$  to  $\alpha$ . 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

<sup>4</sup>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

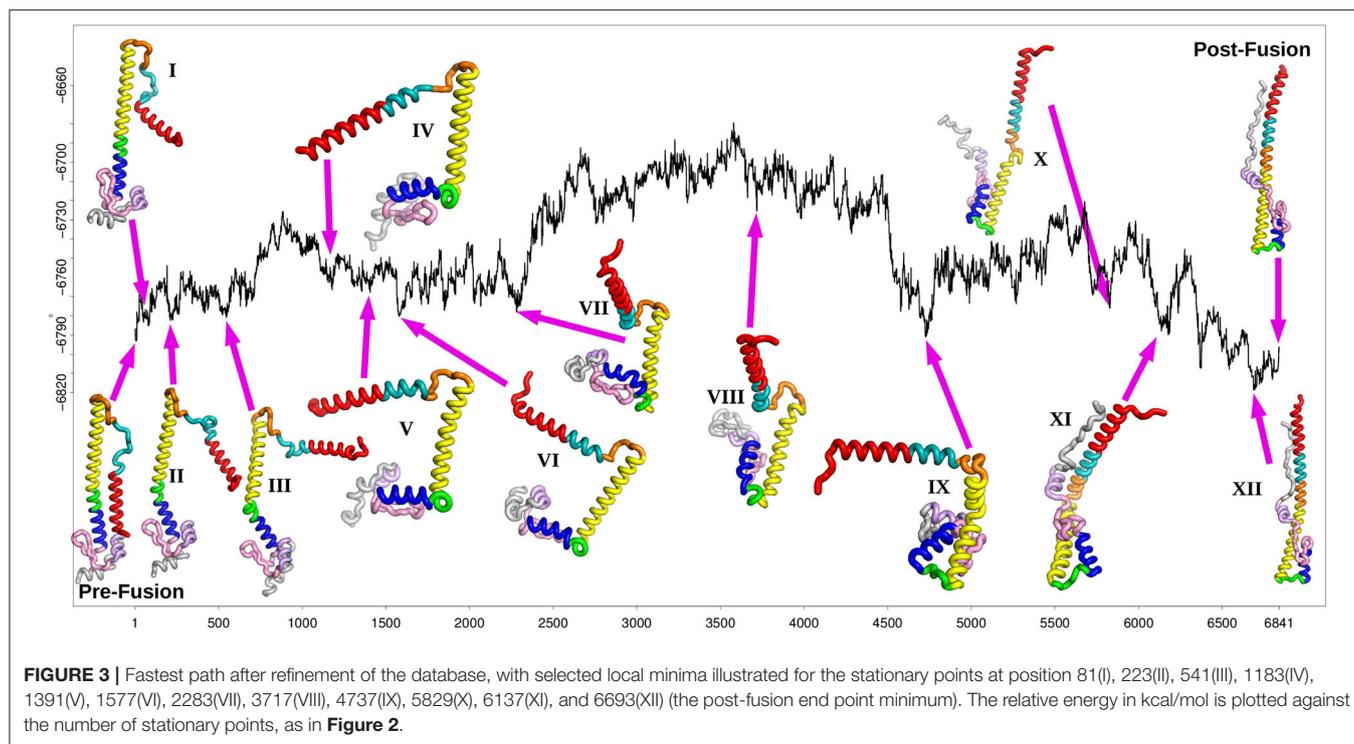
is essential. This refinement constitutes the majority of the computational effort.

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; Née 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



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).

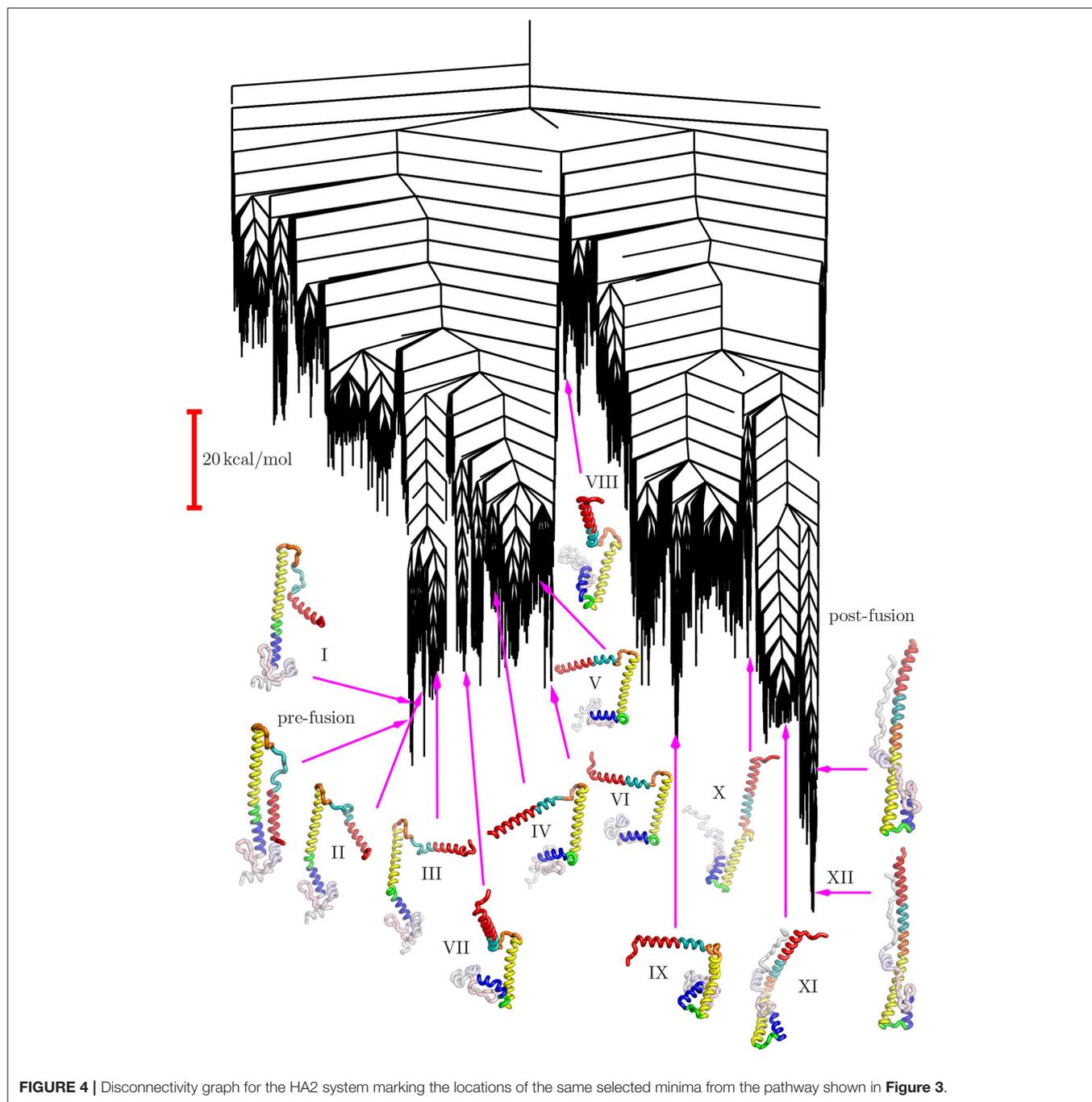
The next minimum belongs to the highest energy part of the path, which corresponds to a relatively shallow local funnel structure. The structural change involves helices E and G moving to point in the opposite direction from the prefusion state (VIII). As the C-terminal peptide extends to interact with 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 shown 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  $\beta$ -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.jpcl.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<sub>2</sub> 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., Stockbine, 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, D. J. (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., Epanand, R. F., Leikina, E., and Epanand, R. M. (2011). The final conformation of the complete ectodomain of the HA<sub>2</sub> 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<sub>10</sub>H<sub>10</sub> 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.
- MacPhar, 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
- Malolepsza, 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
- Malolepsza, 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 Deuffhard, 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 mechanically 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 committer 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.

*Copyright © 2020 Burke, Mantell, Pitt and Wales. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*