Multifunctional energy landscape for a DNA G-quadruplex: An evolved molecular switch

Tristan Cragnolini, Debayan Chakraborty, Jiří Šponer, Philippe Derreumaux, Samuela Pasquali, and David J. Wales

Citation: The Journal of Chemical Physics **147**, 152715 (2017); doi: 10.1063/1.4997377
View online: http://dx.doi.org/10.1063/1.4997377
View Table of Contents: http://aip.scitation.org/toc/jcp/147/15
Published by the American Institute of Physics

Articles you may be interested in

Atomic clusters with addressable complexity
The Journal of Chemical Physics **146**, 054306 (2017); 10.1063/1.4974838

Defining and quantifying frustration in the energy landscape: Applications to atomic and molecular clusters, biomolecules, jammed and glassy systems
The Journal of Chemical Physics **146**, 124103 (2017); 10.1063/1.4977794

Predicting reaction coordinates in energy landscapes with diffusion anisotropy
The Journal of Chemical Physics **147**, 152701 (2017); 10.1063/1.4983727

A new class of enhanced kinetic sampling methods for building Markov state models
The Journal of Chemical Physics **147**, 152702 (2017); 10.1063/1.4984932

The threshold algorithm: Description of the methodology and new developments
The Journal of Chemical Physics **147**, 152713 (2017); 10.1063/1.4985912

A nucleotide-level coarse-grained model of RNA
The Journal of Chemical Physics **140**, 235102 (2014); 10.1063/1.4881424

Multifunctional energy landscape for a DNA G-quadruplex: An evolved molecular switch

Tristan Cragnolini,[1] Debayan Chakraborty,[1] Jiří Šponer,[2] Philippe Derreumaux,[3,4] Samuela Pasquali,[3] and David J. Wales[1,a)]

[1]*University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, United Kingdom*
[2]*Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic*
[3]*Laboratoire de Biochimie Théorique UPR 9080 CNRS, Université Paris Diderot, Sorbonne Paris Cité, IBPC 13 Rue Pierre et Marie Curie, 75005 Paris, France*
[4]*Institut Universitaire de France, Boulevard Saint-Michel, 75005 Paris, France*

(Received 26 April 2017; accepted 24 July 2017; published online 9 August 2017)

We explore the energy landscape for a four-fold telomere repeat, obtaining interconversion pathways between six experimentally characterised G-quadruplex topologies. The results reveal a *multi-funnel* system, with a variety of intermediate configurations and misfolded states. This organisation is identified with the intrinsically *multi-functional* nature of the system, suggesting a new paradigm for the classification of such biomolecules and clarifying issues regarding apparently conflicting experimental results. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4997377]

I. INTRODUCTION

Biological molecules, such as nucleic acids and proteins, are challenging systems for theory and simulations due to their complex interactions, high dimensionality, hierarchy of time scales, and interplay between different length and time scales.[1,2] This behavior requires efficient sampling methods to properly account for structural, dynamical, and thermodynamic properties.

Here we employ the potential energy landscape approach, which provides insight both at a fundamental level and as a practical route to the design of computational tools.[3] For example, visualisation of the corresponding kinetic transition networks[4–7] using disconnectivity graphs[8,9] has provided the basis for understanding self-organisation in a diverse range of different systems, including "magic number" clusters, proteins, and crystals.[10] These landscapes exhibit "funnelled" topographies, where relaxation to the lowest energy morphology involves relatively small downhill barriers for any alternative structure. Systems with a low-lying competing morphology, separated by a large barrier, often exhibit a low-temperature heat capacity signature where the global free energy minimum switches between structures, along with distinct relaxation time scales.[11,12] This effect has been observed in nucleic acid systems, showing clear partitioning of the conformational space in separate funnels.[13,14] Recent experiments further demonstrate that competition between alternative four-stranded folds dominates the folding landscape of human telomeric quadruplexes.[15–17] The landscapes of glassy systems correspond to the limit of an exponential number of low-lying minima, separated by barriers that are large compared with the thermal energy available at the glass transition temperature.[7,10,18] This sort of structural competition is often referred to as "frustration."[19,20]

The appearance of alternative favourable structures introduces the possibility of control over the molecular function with changes in the local environment, such as temperature, or binding to specific ligands. For example, intrinsically disordered proteins (IDPs) are present in many cellular protein interaction networks. In recent work, we have shown that one particular IDP, involved in apoptosis when bound to its partner, exhibits a number of alternative low energy structures in the absence of the ligand.[21] This observation suggests that IDPs may typically be associated with *multi-funnel* potential energy landscapes. Multifunnel landscapes have also been inferred for allosteric proteins, particularly in the context of symmetric multimeric systems.[22]

In the present work, we use the coarse-grained HiRE-RNA model to examine the landscape associated with six experimental motifs reported for a four-fold telomere repeat in DNA, which is known to form a G-quadruplex structure. Telomeric G-quadruplexes (GQs) have recently become a particularly active area of research since the repeat base sequences of the telomere play an important role in protecting the termini of chromosomes from damage. Quadruplex formation decreases the activity of the telomerase enzyme, which functions to maintain the length of the telomeric repeats. Malfunctioning of this protection system is associated with a large fraction of cancer conditions.[23]

DNA quadruplexes are stabilized by quartets of guanine bases, each forming two hydrogen bonds (HBs) on their Watson-Crick side with the Hoogsteen side of the next base,[24] for a total of eight HB per quartet. Coupled with the strong stacking between successive quartets, those structures are extremely stable. While the quartet structure is well known, the topology of those molecules, governed by the loop connecting the guanines, has been the subject of much debate.[25,26]

a)dw34@cam.ac.uk

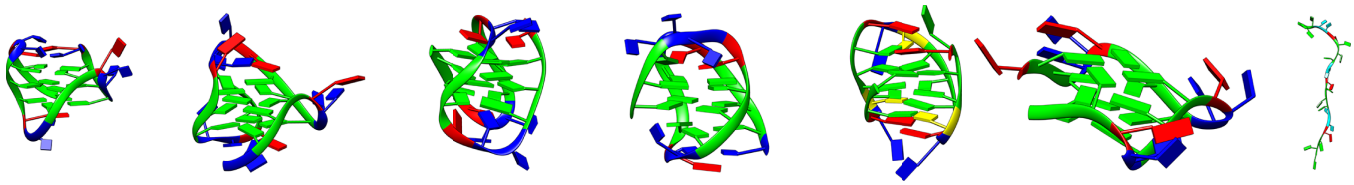

FIG. 1. Simplified representation of the starting conformations. The numbering scheme corresponds to the text, with U designating the initial unfolded structure used in the connection runs.

In the present contribution, we obtain pathways and calculate interconversion rates between all known topologies adopted by a four-fold telomere repeat of 22 bases, with sequence $A(G_3TTA)_3G_3$ (presented in Fig. 1). We find that the landscape has a multi-funnelled structure, with numerous intermediates and misfolded states. We suggest that the landscape has evolved to support the multi-functional role of the quadruplexes, which are involved in the gene expression[27] and are implicated in diseases, such as amyotrophic lateral sclerosis and fronto-temporal dementia.[28] This view suggests a new perspective for understanding the structure, dynamics, and thermodynamics of telomeres, with potential applications in future drug discovery.[29]

II. METHODS

A. The HiRE-RNA potential

We employ the coarse-grained HiRE-RNA model, developed to study the structural and dynamical properties of nucleic acids, which explicitly includes the key driving forces involved in nucleic acid folding.[30,31] The model uses six or seven particles per nucleotide, positioned on the P, O5′, C5′, C4′, and C1′ atoms, plus one or two particles at the center of mass of each aromatic ring in the nucleobases (two for purines and one for pyrimidines). The chain connectivity and local geometry are maintained with harmonic potentials for the bond lengths and angles, and dihedrals are maintained with periodic expressions in the corresponding angles. The excluded volume of the particles is represented by a fast-decaying negative exponential function. The electrostatic interaction between the phosphates is modeled using a Debye-Hückel potential, with implicit solvent and ions. The electrostatic potential contains a parameter, the Debye length, which is related to the ionic concentration in the solution. We fixed its value at 5 Å, which corresponds to an ionic strength of around 370 mM, a buffering of the electrostatics similar to ionic concentrations used in G-quadruplex folding studies, conditions that stabilise the GQ structures.

G-quadruplexes are stabilised by ions inserted in their central channel, between neighbouring tetrads. However, an accurate representation of these ion-base interactions, with the eight-fold coordination of the ion, is difficult. All-atom models use a point charge and an isotropic Lennard-Jones term. More accurate parametrisations for ions with such coordination geometries (in particular magnesium) have been proposed,[32] but reproducing other properties remains difficult.[33] The melting temperatures of the basket forms (structures 3 and 4) and the 3+1 form 2 (structure 2) are compared in our study, while it is known experimentally that they are stabilized by Na^+ or K^+ ions, respectively.

The hydrogen-bonding and stacking terms involve the last three particles of each base and take into account their planarity, relative distances, and orientations. An extensive description of the current model and its parametrisation for both DNA and RNA can be found in Ref. 30. Crucially for the present work, the model allows for both canonical and non-canonical base pairings and contains a reparametrisation of the anti/syn equilibrium for the glycosidic χ torsion, an important element in G-quadruplex structures. To the best of our knowledge, only one previous study attempted to model GQs at a coarse-grained level.[34]

Other models suitable for the study of nucleic acids have been proposed recently,[35–38] though to the best of our knowledge none have been applied to study conformational changes between quadruplex structures.

B. Preparation of starting structures

Several experimental structures are known for the human telomeric sequences forming DNA G-quadruplexes, differing in the sequence and organisation of loops, and by the ions present in the central channel. To study the relative stability, we created structures sharing the same sequence, based on the experimental data available in the nucleic acid database (NDB). We used the following six NDB structures as starting points in this work (Fig. 1): (1) 2JSM[39] 3+1 form 1; (2) 2JSL[39] 3+1 form 2; (3) 143D[40] basket form 1; (4) 2KF8[41] basket form 2; (5) 2KM3[42] chair structure, which comprises a GCGC quartet and forms an extra GC base pair (while the consensus sequence that we use cannot form those GC pairings, we still included this structure for its unique loop topology); (6) 1KF1[43] propeller structure, the only crystal structure in this list; its overall shape is markedly different from the NMR structures, and the loop organisation allows it to form contacts with neighbouring G-quadruplexes in the crystal. Energy minimisation was then used to bring each structure to a nearby local minimum. In each case, the local minimum is structurally very similar to the original structure.

Our aim in the present contribution is to understand the landscape associated with these six experimental motifs, and in particular, to predict the pathways and mechanisms that connect them. Other structures with alternative loop topologies, and syn/trans orientations of the χ torsions in the guanine

bases, are possible and indeed have been reported.[44] While a more thorough investigation of the dynamics of the G-quadruplex including more potentially stable states is planned, the large barriers separating the various topologies make an exhaustive search complicated. However, these alternative structures lie sufficiently far away in configuration space, and are separated by large enough barriers, that their inclusion would not significantly affect the present results for the subset of experimental motifs considered.

Lastly, we generated a representation of the unfolded state to assist in establishing possible folding pathways. While defining the unfolded state can be difficult,[45] we decided to simply generate a structure as fully unstructured as possible. We therefore performed a pulling molecular dynamics (MD) simulation,[46] until no secondary structure remained, and the backbone was stretched. The structure with the highest end-to-end distance was selected as an unfolded structure. This high energy structure is not expected to be representative of the unfolded state but was employed as a reference in the landscape to which we connected the low-lying experimental structures. The unfolding pathways from the pulling simulations were not used due to the possible bias towards shorter pathways due to pulling. Lower energy routes between the key structures were always located away from within the initial connected database, as described in Sec. II C.

Unfolded structures were also generated using molecular dynamics in the NPT ensemble, using largely the same protocol used in previous studies performed with HiRE-RNA.[30] The time propagation was performed with the velocity Verlet algorithm with an integration time step of 4 fs and a Langevin thermostat at 300 K. A pulling simulation was conducted using a distance based potential between the initial O5′ and final C4′ atoms of the structure. The pulling potential adopted was $(r - r_0)^2$ if $r < 4$ and $4(2|r - r_0| - 4)$ if $r \geq 4$ (see the supplementary material for derivation), with r being the distance described above and r_0 being a value larger than the DNA contour length.

C. Exploring the energy landscape: Discrete path sampling

The discrete path sampling (DPS) procedure[47,48] was employed to explore the energy landscape of the DNA G-quadruplex. This approach has been successfully applied to investigate the energy landscapes for a diverse range of atomic and molecular systems[49–53] and has proved to be particularly efficient in exploring the landscapes featuring broken ergodicity. DPS exploits geometry optimisation techniques to provide a coarse-grained description of the underlying energy landscape in terms of minima and transition states. These databases of stationary points constitute a kinetic transition network,[7,54] which can be used to analyse the global thermodynamics and kinetics.

The connectivity of stationary points in the database is described in terms of discrete paths. In a discrete path, successive minima between endpoints of interest are connected by intervening transition states. Starting from an unconnected pair of minima, candidate transition state structures (and intervening minima) are identified by the doubly-nudged[55] elastic band[56] method. Once identified, the minima are refined using a modified version of the limited memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm,[57] and the transition states are accurately refined by the hybrid eigenvector-following method.[58,59] The geometry optimisations and transition state searches were carried out using the OPTIM program.[60] Geometry optimisations were deemed to be converged when the root mean square gradient fell below 10^{-5} kcal (mol^{-1}Å^{-1}).

We first attempted to find complete discrete paths between all the low-lying local minima, corresponding to the starting structures, in a pairwise fashion. If the endpoints are well separated in configuration space, numerous intervening minima and transition states may be found in each cycle, before the original endpoints are connected. We employed a missing connection algorithm[61] to build a priority list of connection attempts based on appropriate edge weights and distributed transition state searches on different compute nodes using the PATHSAMPLE program.[62] These initial paths are unlikely to be kinetically relevant because higher energy transition states often appear in the first interpolation between more distant minima, so they were systematically refined using various schemes available within PATHSAMPLE. Procedures to shorten key pathways and find lower barriers[63,64] were used sequentially. We also attempted to improve the connectivity of the database by seeding single-ended transition searches from unconnected minima. This step was followed by double-ended searches between pairs of minima that were close in configuration space (less than 1 Å root mean square deviation), but unconnected. Finally, the UNTRAP scheme,[63] which is based on the ratio of the potential energy barrier to the potential energy difference between pairs of local minima, was used to remove artificial frustration from the database caused by undersampling. The conformational transitions between the different quadruplex polymorphs were visualised in terms of the pathways that make the largest contribution to the steady-state rate constants, where the dynamics between adjacent minima is treated as harmonic. These pathways are the "fastest paths" and were extracted from the stationary point databases using Dijkstra's shortest path algorithm[65] using an appropriate edge weight.[66]

D. Analysis of global thermodynamics and kinetics

Free energies were estimated from the stationary point databases using the harmonic superposition approach (HSA).[67,68] In this approach, the total density of states and the canonical partition function are written as a sum of contributions from the catchment basin of each local minimum. The potential well around each local minimum is assumed to be harmonic, and the vibrational density of states for each minimum is calculated from the normal mode frequencies. The heat capacities reported in this work have also been estimated using the HSA. The normal mode frequencies were obtained using numerical second derivatives for the Hessian matrix. Previous work has demonstrated that the HSA can provide a fairly accurate estimate of heat capacities, particularly in the low-temperature regime,[69] with systematic shifts caused by the neglect of well anharmonicity and land-

scape entropy (the energy density of local minima) at higher temperatures.

The rate constants corresponding to the overall conformational transitions between the different quadruplex polymorphs were estimated at 298 K using the new graph transformation (NGT) method,[70] in conjunction with a self-consistent regrouping scheme[71] based on free energy barriers. The regrouping procedure recursively lumps together structures separated by free energy barriers below a certain threshold into one macrostate. This approach is particularly attractive, as it exploits the separation of time scales between interbasin transitions and intrabasin relaxations[72] and alleviates possible bias arising due to the original choice of endpoints.

E. Visualisation using disconnectivity graphs

Disconnectivity graphs[8,9,73,74] were used to visualise the potential and free energy landscapes. In contrast to approaches based on low-dimensional projections of the energy landscape onto selected order parameters, disconnectivity graphs faithfully represent the potential or free energy barriers.[75,76] To construct a disconnectivity graph, the energy landscape is segregated at a regular series of energy thresholds into disjoint sets of minima, known as superbasins.[8] Minima within each superbasin are mutually accessible via transition states lying below the threshold, whereas interbasin transitions involve higher transition states. The vertical scale is potential or free energy, and the branches terminate at the values for individual local minima. They merge together at the lowest energy threshold where the minima can interconvert.

The potential and free energy disconnectivity graphs were coloured according to the number of hydrogen bonds. Detection of hydrogen bonds was performed using the corresponding energy term from the HiRE-RNA forcefield: a hydrogen bond was diagnosed if the energy contribution was lower than −0.4 kcal/mol, which is about 40% of the maximum strength of the weakest hydrogen bonds in the model. Empirically, this cutoff appears to be capable of detecting the weaker but stable hydrogen bonds, while producing a rather small number of false positives.

III. RESULTS

The discrete path sampling (DPS) procedure[47,48] (Sec. II) was employed here to explore the energy landscape of the DNA G-quadruplex. Starting from an initial set of known G-quadruplex conformations (shown in Fig. 1), we created a database of minima and first-order transition states connecting those minima, eventually linking all those conformations by pathways formed by alternating minima and transition states. Several schemes designed to discover the new minima and transition states were employed (Sec. II C), until the database was deemed converged. The resulting potential energy disconnectivity graph[8,9] (Sec. II E) is shown in Fig. 2. Several distinct regions are clearly visible, with one or more experimental structures present in each one. The propeller structure (obtained by crystallography) lies at the bottom of its own funnel, while the five NMR structures appear at higher energies in their respective basins.

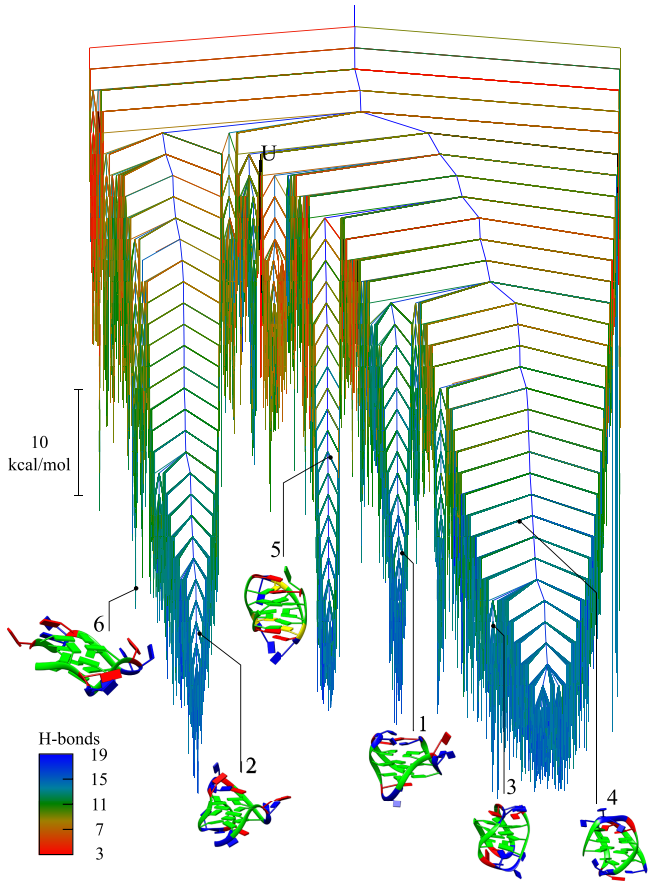

FIG. 2. Disconnectivity graph for the potential energy landscape of the G-quadruplex. The energy scale is in kcal/mol, and the colour is based on the number of hydrogen bonds. The numbering scheme is the same as that in Sec. II B, with U designating the initial unfolded structure used in the connection runs.

While the graph in Fig. 2 provides a general overview of the potential energy landscape, an analysis of the corresponding free energy landscape can be particularly useful to highlight information arising from the ensemble of structures explored. In the free energy disconnectivity graph[73,74] (Sec. II D) shown in Fig. 3, each of the experimental structures lies at (or close to) the bottom of their respective funnel. This observation is particularly interesting for the NMR structures that appear at higher energies in the potential energy landscape. From this graph, it is clear that the interconversion between the various structures will take place on very different time scales due to the diverse range of barrier heights.

The appearance of the free energy disconnectivity graphs presented in Fig. 3 depends on the regrouping threshold used in the calculation.[71] It is instructive to compare free energy disconnectivity graphs using different thresholds, which can be related to an observation time scale:[72] structures separated by a smaller barrier will interconvert faster, and such transitions will require higher temporal resolution in experiments. We emphasize that the graphs simply provide a helpful visualisation; when quantitative results are presented for thermodynamic or kinetic properties, they are always obtained directly from the underlying kinetic transition network including the associated vibrational densities of states.

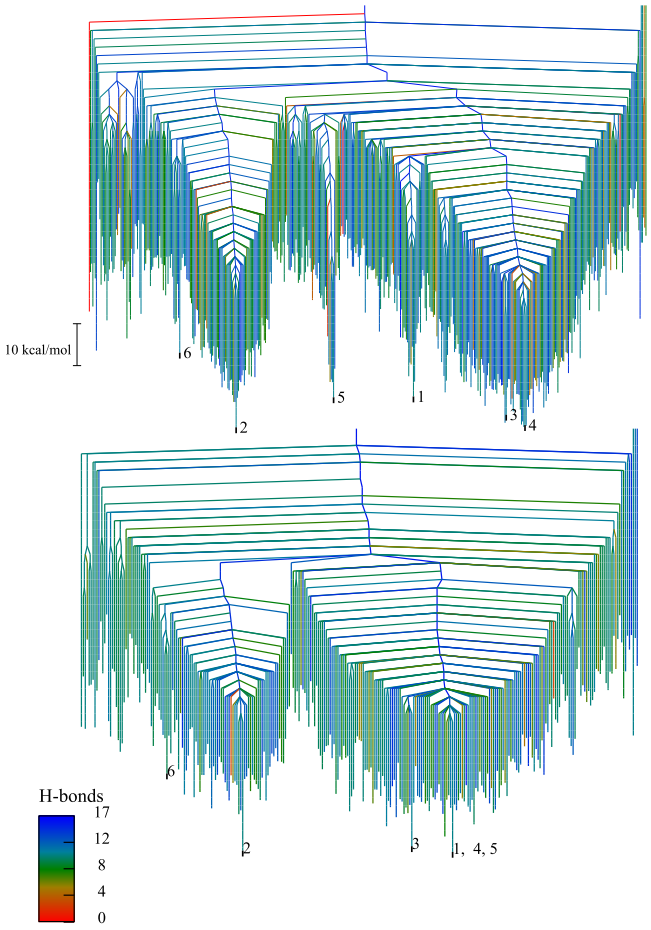

FIG. 3. Disconnectivity graphs of the free energy landscape for the G-quadruplex. The energy scale is in kcal/mol, and the colour is based on the number of hydrogen bonds. The numbering scheme is the same as that in Sec. II B. The top graph was created with a regrouping threshold of 25 kcal/mol and the bottom graph was created with a 28 kcal/mol threshold.

In Fig. 3, we present free energy disconnectivity graphs obtained using regrouping thresholds of 25 and 28 kcal/mol. The 25 and 28 kcal/mol thresholds roughly correspond to conversion time scales of seconds to hours,[72] which is the usual time scale needed for the conversion from one G-quadruplex structure to another. The free energy landscapes exhibit little change for regrouping thresholds below

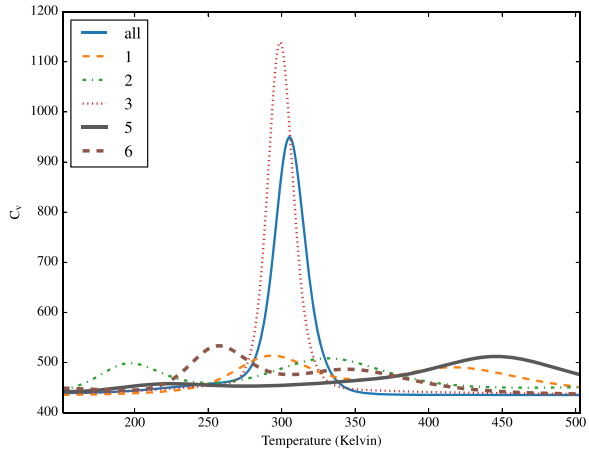

FIG. 4. Heat capacities calculated using the harmonic superposition approach. The contributions of each region were evaluated separately to evaluate the heat capacity and compared with the total calculated from the complete database. Structures 3 and 4 are part of the same region and are grouped together.

25 kcal/mol and appear very similar to the potential energy landscape: only local structures are grouped together. However, as we move to higher thresholds, some features of the landscape (corresponding to the main G-quadruplex structures) start to merge and become part of the same free energy basins, until eventually all structures are grouped in a single basin. At a 28 kcal/mol threshold, only two main features remain. This result may explain why only certain structures appear in time-resolved NMR studies of G-quadruplexes:[15] if the observation time scale is long enough, the structures will interconvert too fast to be observed individually. For example, at longer time scales, we would expect the hybrid 2 structure to be indistinguishable from the two basket conformations.

We also show the heat capacity, C_V, calculated using the harmonic superposition approach (HSA) in Fig. 4.[67,68] In addition to the heat capacity calculated from all the sampled conformations, we also show C_V evaluated independently for each identifiable region. These separate melting curves, while not reflecting the overall equilibrium behavior of the sampled points, may provide curves closer to those observed experimentally when the experimental conditions favor a

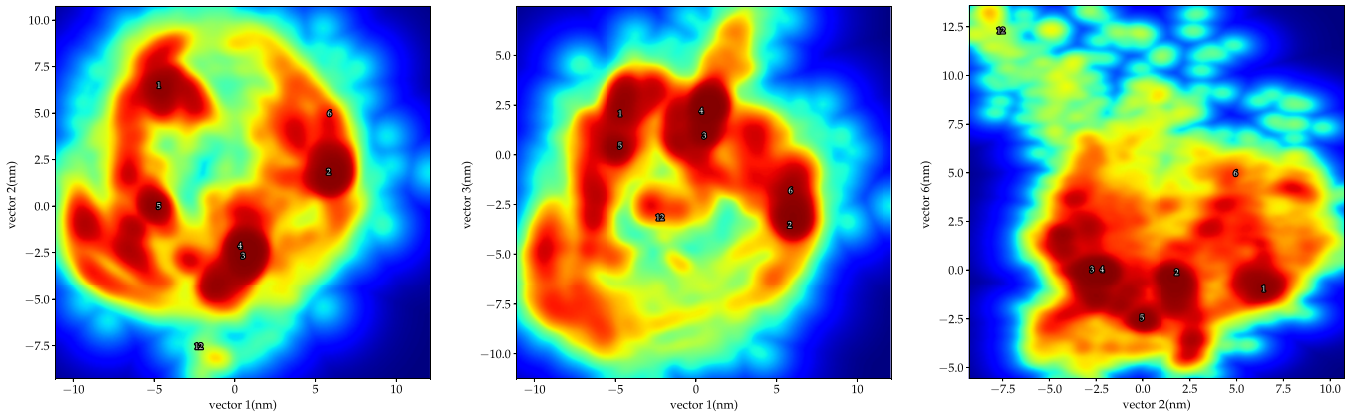

FIG. 5. Two-dimensional free energy projections of the landscape along several PCA eigenvectors (in nanometres). Different paths connecting the main funnels are apparent in the three projections, illustrating the complex underlying connectivity of the landscape.

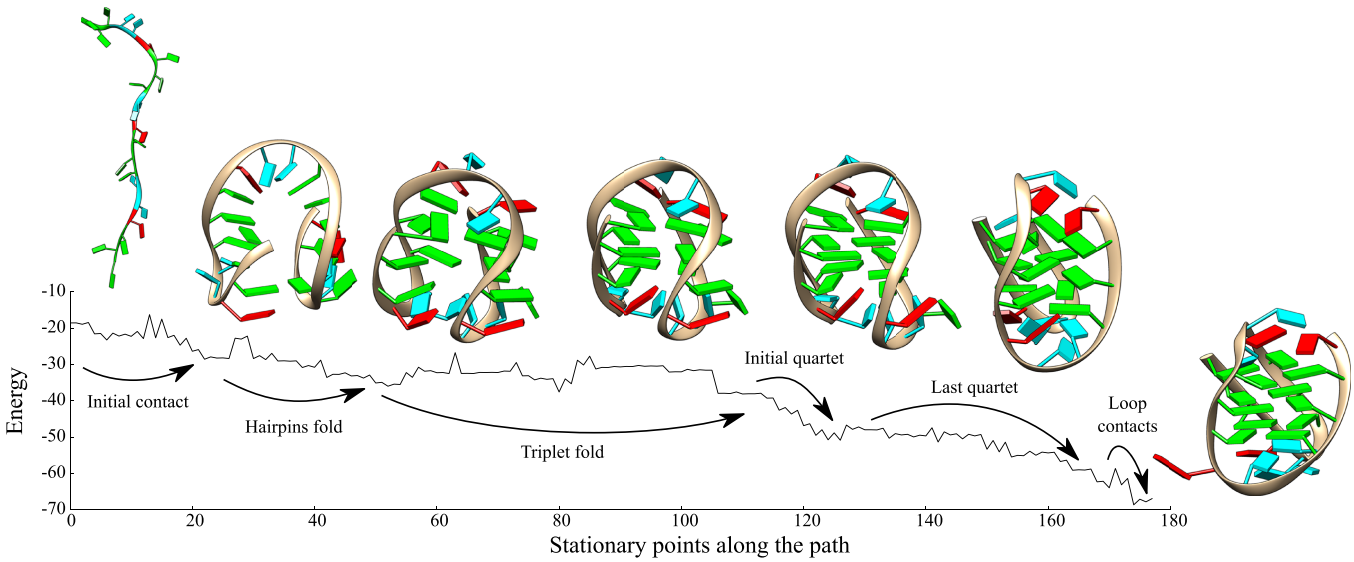

FIG. 6. Selected structures on the fastest pathway from the unfolded state to the type 2 basket.

particular quadruplex structure, with a corresponding shift in the population of each possible structure.

The regions were separated either using an energy threshold below the lowest energy transition state connecting two basins or by a graph cutting approach to obtain disjoint subsets of the database, with virtually identical results. The apparent two-state behavior of the system as a function of temperature, as seen in the total heat capacity, actually masks a more complex situation, with each separate region producing distinct melting curves. The melting temperatures for structures 2 and 3/4 are roughly in line with the experimental values of around 323 K.[41]

To better understand this complex landscape, we performed a principal component analysis (PCA) of our data. The PCA was performed on the aggregate set of all the minima sampled, using their Cartesian coordinates (Fig. 5). These graphs exhibit multiple paths between each funnel, and their organisation is more complex than can be represented by a single 2D plot.

We can obtain insight into the folding mechanism by extracting specific pathways. Figure 6 illustrates the pathway making the largest contribution to the rate constant[47,48] for the conversion from the unfolded state to the basket 2 conformation. The mechanism involves numerous elementary steps, and we present a few snapshots to show the relatively straightforward and hierarchical nature of this quadruplex folding pathway (a complete movie is available in the supplementary material). After an initial contact between bases appears in the extended state, the formation of two hairpin-like structures between G-bases is observed; extra contacts are established, leading to the formation of triplets, followed by quartet formations. The last step to a fully folded structure is an improvement of the contacts formed by bases from the loops. This mechanism is in line with previous suggestions from both theory and experiment,[15,26] although we tend to observe hairpins forming first at the 5′ and 3′ ends of the chain, rather than in the center. Complete coordinates for the pathways are available in the supplementary material.

IV. DISCUSSION

It is interesting to note that the crystal structure lies at the bottom of its own funnel in the potential energy landscape (Fig. 2), unlike the NMR structures. All these experimentally determined configurations correspond to basin bottoms for the free energy landscape. This result can likely be explained by two observations: crystal structures are resolved at relatively low temperatures, and the constraints placed by the crystalline environment result in a lower entropy. Both effects favour a lower potential energy minimum. The X-ray structure appears in a rather narrow funnel. From our analysis of the landscape, with the melting peak of this structure in Fig. 4 lying well below 300 K, this structure should have a low equilibrium population in a solvated environment at 300 K. The environment in a crystalline structure, with comparatively low hydration, and the presence of copies forming multiple contacts for each molecule (in particular stacking interactions) may lead to stabilisation of a configuration that is somewhat different from the conformations favoured in solution.

The presence of NMR structures at or close to the bottom of free energy basins in Fig. 3 is interesting for several reasons. First, it provides a direct theoretical validation of the NMR structures at relevant temperatures. In the potential energy landscape, the minima corresponding to the different NMR structures are not amongst the lowest in their basins, but they are amongst the lowest free energy states.

The second point of interest is that the force field employed stabilises experimentally known structures, a good indication that it can reproduce ensemble properties of a solvated nucleic acid. The stabilisation involves the landscape entropy[77,78] through the free energy regrouping scheme, arising naturally from the multiple potential energy minima.

The direct connection between the potential energy and free energy landscapes provides a route to a more detailed analysis of structures and dynamics and, in particular, the interpretation of NMR data based directly on the known structures and their connectivity.

The overall topology of the landscape, and the relative stability of the alternative structures, provides insight into several issues. For example, it has been reported that the basket type 2 form is more stable than the basket 1, despite having only two G tetrads.[41] It can be seen in Fig. 2 that while the basket 2 structure itself is higher in the potential energy landscape, it is part of a larger basin. The basket 1 form, on the other hand, is part of a narrow side basin, and while its potential energy is lower, as expected, it is not as thermodynamically stable at relevant temperatures (see Fig. 3).

The present work shows how the discrete path sampling approach,[47,48] coupled with a coarse-grained representation of the system, allows for a systematic exploration of the low-lying parts of the landscape for the selection of experimentally observed structures, even when they are separated by high barriers. The kinetics calculated from our database tend to be slower than those observed experimentally. In particular, the free energy thresholds used in the regrouped free energy graphs (Fig. 3) correspond to longer time scales than experiment. It may be possible to use these results to guide reparametrisation of the potential to reproduce barriers and kinetics more accurately. These dynamical effects depend upon parts of configuration space beyond the equilibrium geometries that are usually employed in fitting.

The agreement between theory and experiment achieved in the present study, and the additional insight we obtain, especially in terms of pathways, suggests a number of opportunities for further work. This multifunnel free energy landscape is likely to be conserved for DNA G-quadruplex sequences in physiological conditions, though the energy barriers and the interconversion paths may change, for example, with changes in the concentration of ionic species. We will explore how changes in the electrostatic description of the environment affect the different structures and the landscape, in particular, by adding details to the representation of tightly bound ions. A systematic exploration of all possible quadruplex organisations, based on their topologies and the *anti*/*syn* conformations of the guanines, should now be feasible. We also plan to investigate alternative quadruplex forming sequences, such as promoter sequences and telomere variants, and the various sequences that were used to obtain the experimental data. Finally, we will analyse the formation of higher-order structures, by assembling several G-quadruplexes, using a local rigid body approach.[79,80] The present framework should allow us to study the propensity of all the major G-quadruplex structures to form higher-order structures, such as G-wires, cholesteric phases, and other possible supramolecular arrangements.

SUPPLEMENTARY MATERIAL

See supplementary material for additional information concerning figure creation, choice of consensus sequence, appearance of experimental structures in the landscape, additional free energy disconnectivity graphs, PCA projections. Parameters of the HiRE-v3 potential are also specified, namely, fixed parameters, local interactions, excluded volume, electrostatics, stacking, and base-pairing. The pulling potential is also defined.

ACKNOWLEDGMENTS

D.C. acknowledges the Cambridge Commonwealth, European and International Trusts for Ph.D. funding. T.C. acknowledges funding from EPSRC Grant No. EP/I001352/1. S.P. and P.D. acknowledge support from "DYNAMO" ANR-11-LABX-0011 and PSL (Paris Sciences et Lettres). J.S. was supported by Czech Science Foundation Grant No. 16-13721S. D.J.W. acknowledges funding from EPSRC Grant No. EP/N035003/1.

T.C. and D.C. generated the data, under the supervision of D.J.W. T.C., J.S., S.P., and P.D. contributed to the development of the nucleic acid model used in the study. T.C. analysed the data and created the figures, with contributions from D.C. and D.J.W. D.J.W. created and maintains the GMIN, OPTIM, and PATHSAMPLE programs for exploring energy landscapes. All authors contributed to writing the manuscript.

[1] T. E. Cheatham III and D. A. Case, Biopolymers **99**, 969 (2013).
[2] J. R. Bothe, E. N. Nikolova, C. D. Eichhorn, J. Chugh, A. L. Hansen, and H. M. Al-Hashimi, Nat. Methods **8**, 919 (2011).
[3] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
[4] F. Rao and A. Caflisch, J. Mol. Biol. **342**, 299 (2004).
[5] F. Noé and S. Fischer, Curr. Opin. Struct. Biol. **18**, 154 (2008).
[6] D. Prada-Gracia, J. Gómez-Gardeñes, P. Echenique, and F. Fernando, PLoS Comput. Biol. **5**, e1000415 (2009).
[7] D. J. Wales, Curr. Opin. Struct. Biol. **20**, 3 (2010).
[8] O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).
[9] D. J. Wales, M. A. Miller, and T. R. Walsh, Nature **394**, 758 (1998).
[10] D. J. Wales, Philos. Trans. R. Soc., A **363**, 357 (2005).
[11] J. P. K. Doye, M. A. Miller, and D. J. Wales, J. Chem. Phys. **110**, 6896 (1999).
[12] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **111**, 11070 (1999).
[13] C. Hyeon, J. Lee, J. Yoon, S. Hohng, and D. Thirumalai, Nat. Chem. **4**, 907 (2012); e-print arXiv:1211.0662.
[14] S. V. Solomatin, M. Greenfeld, S. Chu, and D. Herschlag, Nature **463**, 681 (2010).
[15] I. Bessi, H. R. A. Jonker, C. Richter, and H. Schwalbe, Angew. Chem., Int. Ed. **54**, 8444 (2015).
[16] A. Marchand and V. Gabelica, Nucleic Acids Res. **44**, 10999 (2016).
[17] M. Aznauryan, S. Søndergaard, S. L. Noer, B. Schiøtt, and V. Birkedal, Nucleic Acids Res. **44**, 11024 (2016).
[18] V. K. de Souza and D. J. Wales, J. Chem. Phys. **129**, 164507 (2008).
[19] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **21**, 167 (1995).
[20] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[21] Y. Chebaro, A. J. Ballard, D. Chakraborty, and D. J. Wales, Sci. Rep. **5**, 10386 (2015).
[22] D. U. Ferreiro, J. A. Hegler, E. A. Komives, and P. G. Wolynes, Proc. Natl. Acad. Sci. U. S. A. **108**, 3499 (2011).
[23] S. Neidle, FEBS J. **277**, 1118 (2010).
[24] N. B. Leontis and E. Westhof, RNA **7**, 499 (2001).
[25] W. Li, X.-M. Hou, P.-Y. Wang, X.-G. Xi, and M. Li, J. Am. Chem. Soc. **135**, 6423 (2013).
[26] P. Stadlbauer, P. Kührová, P. Banáš, J. Koča, G. Bussi, L. Trantírek, M. Otyepka, and J. Šponer, Nucleic Acids Res. **43**, 9626 (2015).
[27] S. Balasubramanian, L. H. Hurley, and S. Neidle, Nat. Rev. Drug Discovery **10**, 261 (2011).
[28] K. Reddy, B. Zamiri, S. Y. R. Stanley, R. B. Macgregor Jr., and C. E. Pearson, J. Biol. Chem. **288**, 9860 (2013).
[29] J. Husby, A. K. Todd, J. A. Platts, and S. Neidle, Biopolymers **99**, 989 (2013).
[30] T. Cragnolini, Y. Laurin, P. Derreumaux, and S. Pasquali, J. Chem. Theory Comput. **11**, 3510 (2015).
[31] T. Cragnolini, P. Derreumaux, and S. Pasquali, J. Phys.: Condens. Matter **27**, 233102 (2015).
[32] O. Allnér, L. Nilsson, and A. Villa, J. Chem. Theory Comput. **8**, 1493 (2012).

[33]J. C. Bowman, T. K. Lenz, N. V. Hud, and L. D. Williams, Curr. Opin. Struct. Biol. **22**, 262 (2012).
[34]M. C. Linak, R. Tourdot, and K. D. Dorfman, J. Chem. Phys. **135**, 205102 (2011).
[35]C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. U. S. A. **102**, 6789 (2005); e-print arXiv:1512.00567.
[36]S. Cao and S.-J. J. Chen, J. Phys. Chem. B **115**, 4216 (2011).
[37]P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. K. Doye, and A. A. Louis, J. Chem. Phys. **137**, 135101 (2012).
[38]M. Rebič, F. Mocci, A. Laaksonen, and J. Uličný, J. Phys. Chem. B **119**, 105 (2015).
[39]A. T. Phan, V. Kuryavyi, K. N. Luu, and D. J. Patel, Nucleic Acids Res. **35**, 6517 (2007).
[40]Y. Wang and D. J. Patel, Structure **1**, 263 (1993).
[41]K. W. Lim, S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. J. Patel, K. N. Luu, and A. T. Phan, J. Am. Chem. Soc. **131**, 4301 (2009).
[42]K. W. Lim, P. Alberti, A. Guédin, L. Lacroix, J.-F. Riou, N. J. Royle, J.-L. Mergny, and A. T. Phan, Nucleic Acids Res. **37**, 6239 (2009).
[43]G. N. Parkinson, M. P. H. Lee, and S. Neidle, Nature **417**, 876 (2002).
[44]K. W. Lim, V. C. M. Ng, N. Maítin-Pintado, B. Heddi, and A. T. Phan, Nucleic Acids Res. **41**, 10556 (2013).
[45]R. Narayanan, L. Zhu, Y. Velmurugu, J. Roca, S. V. Kuznetsov, G. Prehna, L. J. Lapidus, and A. Ansari, J. Am. Chem. Soc. **134**, 18952 (2012).
[46]F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y. Chebaro, J.-F. St-Pierre, M. Kalimeri, A. Barducci, Y. Laurin, A. Tek, M. Baaden, P. H. Nguyen, and P. Derreumaux, Chem. Soc. Rev. **43**, 4871 (2014).
[47]D. J. Wales, Mol. Phys. **100**, 3285 (2002).
[48]D. J. Wales, Mol. Phys. **102**, 891 (2004).
[49]J. D. Farrell, C. Lines, J. J. Shepherd, D. Chakrabarti, M. A. Miller, and D. J. Wales, Soft Matter **9**, 5407 (2013).
[50]J. W. R. Morgan and D. J. Wales, Nanoscale **6**, 10717 (2014).
[51]J. M. Carr and D. J. Wales, Phys. Chem. Chem. Phys. **11**, 3341 (2009).
[52]B. Strodel, J. W. L. Lee, C. S. Whittleston, and D. J. Wales, J. Am. Chem. Soc. **132**, 13300 (2010).
[53]D. Chakraborty, R. Collepardo-Guevara, and D. J. Wales, J. Am. Chem. Soc. **136**, 18052 (2014).
[54]D. Gfeller, P. De Los Rios, A. Caflisch, and F. Rao, Proc. Natl. Acad. Sci. U. S. A. **105**, 6 (2007).
[55]S. A. Trygubenko and D. J. Wales, J. Chem. Phys. **120**, 2082 (2004).
[56]G. Henkelman and H. Jónsson, J. Chem. Phys. **111**, 7010 (1999).
[57]D. Liu and J. Nocedal, Math. Program. **45**, 503 (1989).
[58]L. J. Munro and D. J. Wales, Phys. Rev. B **59**, 3969 (1999).
[59]Y. Zheng, P. Xiao, and G. Henkelman, J. Chem. Phys. **140**, 044115 (2014).
[60]D. J. Wales, "Optim: A program for geometry optimisation and pathway calculations," http://www-wales.ch.cam.ac.uk/software.html.
[61]J. M. Carr, S. A. Trygubenko, and D. J. Wales, J. Chem. Phys. **122**, 234903 (2005).
[62]D. J. Wales, "Pathsample: A program for generating connected stationary point databases and extracting global kinetics," http://www-wales.ch.cam.ac.uk/software.html.
[63]B. Strodel, C. S. Whittleston, and D. J. Wales, J. Am. Chem. Soc. **129**, 16005 (2007).
[64]J. M. Carr and D. J. Wales, J. Chem. Phys. **123**, 234901 (2005).
[65]E. W. Dijkstra, Numerische Math. **1**, 269 (1959).
[66]D. A. Evans and D. J. Wales, J. Chem. Phys. **121**, 1080 (2004).
[67]F. H. Stillinger and T. A. Weber, J. Chem. Phys. **80**, 2742 (1984).
[68]B. Strodel and D. J. Wales, Chem. Phys. Lett. **466**, 105 (2008).
[69]V. A. Sharapov, D. Meluzzi, and V. A. Mandelshtam, Phys. Rev. Lett. **98**, 105701 (2007).
[70]D. J. Wales, J. Chem. Phys. **130**, 204111 (2009).
[71]J. M. Carr and D. J. Wales, J. Phys. Chem. B **112**, 8760 (2008).
[72]D. J. Wales and P. Salamon, Proc. Natl. Acad. Sci. U. S. A. **111**, 617 (2014).
[73]D. A. Evans and D. J. Wales, J. Chem. Phys. **118**, 3891 (2003).
[74]S. V. Krivov and M. Karplus, J. Chem. Phys. **117**, 10894 (2002).
[75]S. V. Krivov and M. Karplus, Proc. Natl. Acad. Sci. U. S. A. **101**, 14766 (2004).
[76]D. J. Wales, J. Chem. Phys. **142**, 130901 (2015).
[77]G. Meng, N. Arkus, M. P. Brenner, and V. N. Manoharan, Science **327**, 560 (2010).
[78]D. J. Wales, ChemPhysChem **11**, 2491 (2010).
[79]H. Kusumaatmaja, C. S. Whittleston, and D. J. Wales, J. Chem. Theory Comput. **8**, 5159 (2012).
[80]V. Rühle, H. Kusumaatmaja, D. Chakrabarti, and D. J. Wales, J. Chem. Theory Comput. **9**, 4026 (2013).