Supplementary information

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, UK

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

(Dated: 21 July 2017)

^{a)}dw34@cam.ac.uk

I. FIGURE CREATION

Figures 1 and 2 were generated with the disconnectionDPS software.

Figure 3 was generated with the matplotlib library, and the specific heat was calculated with the pele software.

Figure 4 was generated with the matplotlib library, and the PCA calculation was performed with the gromacs software suite.

Figures 5 and 6 were created using inkscape. The data from figure 5 was created using PATHSAMPLE.

The structures in figures 1, 5, and 6 were generated with Chimera, using a full-atom reconstruction method detailed in Cragnolini et al., 2013.

Pathway movies were generated using chimera and ffmpeg.

II. CONSENSUS SEQUENCE

The original sequences of the PDB structures we used are as follows:

>2JSM, chain A/1-23 >2JSL, chain A/1-23 >143D, chain A/1-22 >2KF8, chain A/1-22 >2KM3, chain A/1-22 >1KF1, chain A/1-22 TAGGGTTAGGGTTAGGGTTAGGG. TAGGGTTAGGGTTAGGGTTAGGG. .AGGGTTAGGGTTAGGGTTAGGG. .GGGTTAGGGTTAGGGTTAGGGT. AGGGCTAGGGCTAGGGCTAGGG. .AGGGTTAGGGTTAGGGTTAGGG. .AGGGTTAGGGTTAGGGTTAGGG.

. AGGGTTAGGGTTAGGGTTAGGG.

143D and 1KF1 have exactly the same sequence as the consensus. 2JSM and 2JSL were modified by removing the initial thymine. For 2KF8, we removed the final thymine, and added an initial adenine. Finally, we needed to mutate both cytosines into thymine for 2KM3. The chosen sequence is the closest possible to the experimentally known sequences, and matches the most commonly found telomere repeat motif, AGGGTT. We have a quadruple repetition of this motif, minus the two final thymines. Those thymines are expected to be a flexible linker to another quadruplex structure, and are not present in the available structures.

III. ONLINE DATA ACCESS

The data for this system is accessible at: https://www.repository.cam.ac.uk/handle/ 1810/254628.

The paths between folded and unfolded structures, and the main structures, are available at: http://www-wales.ch.cam.ac.uk/CCD.html.

IV. EXPERIMENTAL STRUCTURES IN THE LANDSCAPE

The experimental structures are all close to a local minimum, usually with a small conformational change (on the order of 2 Å). The distance between these structures is on the order of 10 Å, corresponding to significant structural differences for this short sequence, and compact experimental structures.

The following RMSDs are observed between the experimental structures and the corresponding local minima:

TABLE I: Table of RMSDs between native structures and corresponding local minimum

(in Å).

The distance between those structures is on the order of 10 Å, corresponding to considerable structural differences when considering the small sequence length, and compact experimental structures.

V. FREE ENERGY DISCONNECTIVITY GRAPHS

Figure 1 illustrates disconnectivity graphs for different free energy regrouping thresholds. As expected, features corresponding to separate funnels for each of the stable G-quadruplex structure merge as the free energy regrouping threshold is raised. The higher thresholds can be related to the longer timescales required to cross barriers separating different funnels.

VI. PCA PROJECTIONS

Figure 2 shows PCA projections along some other eigenvectors.

VII. HIRE-V3 PARAMETERS

The parameters of the model can be classified in two categories, geometric parameters and energetic parameters, or energy scales. The fist class defines all equilibrium values and shapes of the potentials and they have a direct influence on the local geometries of the molecule. They have been determined through the statistical analysis of 200 RNA structures in the NDB. These parameters are set once and for all and they are not subject to an optimisation procedure.

Energetic parameters give the relative weights of the various interactions and have a global impact on the molecular behaviour, leading the simulation toward low energy states. These parameters are subject to optimisation to determine the linear combination of the interactions that best distinguishes between native experimental structures and decoys.

A. Fixed parameters of the model

1. Local interactions

All parameters for local interaction potentials are given in tables II a, b, and c.

2. Excluded volume

The stiffness of the exponential potential describing the excluded volume interaction is $\kappa = 4$. It is chosen so that the potential is sufficiently steep to at distances corresponding

(a) 10 kcal/mol

(b) 20 kcal/mol

(c) 22 kcal/mol

(d) 25 kcal/mol

(e) 26 kcal/mol

(f) 27 kcal/mol

(g) 28 kcal/mol

(h) 30⁵kcal/mol

(i) 40 kcal/mol

(a) Table of all be	onds	(b)	(b) Table of all angles					
definition	coupling	equilibrium	definition	coupling	equilibrium				
	$\operatorname{constant}$	length		constant	angle				
	$(\rm kcal/mol/Å^2)$	(Å)		$(\rm kcal/mol)$	(deg)				
R4-P	20	3.850	R4-R1-A1	70.	123.6				
R4-R1	200	2.344	R4-R1-U1	70.	132.5				
R1-G1	200	2.622	R4-R1-G1	70.	123.5				
R1-A1	200	2.633	R4-R1-C1	70.	131.1				
R1-U1	200	3.062	R1-A1-A2	120.	116.7				
R1-G1	200	2.622	R1-G1-G2	120.	111.2				
R1-C1	200	3.004	P -05-C5	70.	122.9				
G1-G2	200	2.45	O5-C5-R4	70.	110.6				
A1-A2	200	2.18	C5-R4-P	70.	98.0				
C5-R4	200	1.52	R4-P -O5	50.	110.0				
P -O5	200	1.593	C5-R4-R1	70.	135.5				
O5-C5	200	1.430	R1-R4-P	100.	98.0				

TABLE II: Tables of local parameters.

(c) Table of all dihedrals

definition	coupling	equilibrium
	constant	angle
	$(\rm kcal/mol)$	(deg)
R4-R1-G1-G2	20.0	-30.0
R4-R1-A1-A2	20.0	-30.0
C5-R4-P -O5	6.0	35.0
C5-R4-R1-A1	10.0	-155.0
C5-R4-R1-G1	10.0	-150.0
C5-R4-R1-C1	10.0	-150.0
C5-R4-R1-U1	10.0	-145.0
P -O5-C5-R4	15.0	-2.0
R4-P -O5-C5	7.0	80.0
O5-C5-R4-P	6 _{2.0}	-100.0
O5-C5-R4-R1	2.0	140.0
P -R4-R1-G1	15.0	100.0
P -R4-R1-A1	15.0	100.0

(a)

(b)

(c)

(d) (e)	(f)
---------	-----

FIG. 2: Projections along additional PCA eigenvectors.

to interpenetration of the particles, and goes rapidly to zero at larger distances.

The distance parameters d_0 (Å) regulate the interpenetration distance between beads. B1 and B2 indicate particles of a base independently of the species.

d_0	Р	C5	O5	CA	CY	Β1	B2
Р	4.0	4.0	4.0	4.0	4.0	4.0	4.0
C5		3.6	3.6	4.0	4.0	3.6	3.6
O5			3.6	4.0	4.0	3.6	3.6
CA				4.0	4.0	4.0	4.0
CY					4.0	4.0	4.0
B1						3.2	3.2
B2							3.2

TABLE III: Table of excluded volume distance parameters (in Å).

3. Electrostatics

The only variable parameter in the electrostatic potential is the Debye length, which can be adjusted to match different ionic conditions. In the current study we set $\lambda_D = 2$ Å.

4. Stacking

The parameters of the stacking potential are the equilibrium distance between the geometric centres of the planes defined by the bases r_{st} and the width of the Gaussian regulating the distance dependence of the potential σ_{st} . For this study we have set $r_{st} = 4.2$ Å and $\sigma_{st} = 1.0$ Å². The value of r_{st} is derived from a statistical analysis of the distance between stacked bases, while the value of σ_{st} is chosen so that the potential is sufficiently narrow to be active only for distances close to the r_{st} . It was also chosen from an analysis of existing structures in the NDB.

5. Base-Pairing

The base base-pairing potential depends on two generic parameters regulating the width of the Gaussian functions for the planarity term δ and for the distance dependence of the hydrogen bond term (ξ), and on parameters specific for each base pair. In the present study $\delta = 1.5$ Å and $\xi = 0.8$ Å².

The 28 possible base pairs depend on an equilibrium distance ρ_0 (Å), two angles α_1 and α_2

(rad) and a coupling constant reflecting the number of hydrogen bonds formed in the pairing k_{hb} . To reflect the higher propensity of nucleic acids to form Watson-Crick pairs (due to the interplay between base-pairings and backbone arrangements) that we might miss with our model, base-pairs occurring on the Watson-Crick side of both bases have an interaction strength multiplied by the adjustable factor wc. For all simulation results presented here wc = 1.3.

6. New Parameters

Using a training set chosen as described in the main text, we ran the genetic algorithm from an initial vector taken from version 2 of HiRE-RNA, which was therefore already tested to give good results, even if the parameters had been selected by hand. We obtained 221/320 as best global score for the new parameter, so that for the structures in the training set 2/3 of the decoys have energies in absolute values greater than the native structure. The new parameters were then tested in long MD simulations (from the native structure) on three molecules not included in the training set and compared with simulations run with the initial, non-optimised, parameter set. The simulations with the optimised parameters show a more pronounced stability both in terms of RMSD and in terms of preserved native base-pairs.

The energy scales resulting from the optimisation procedure described in the main text are given in table V.

B. Pulling potential

The pulling potential used to obtain extended structure is linearized at larger distance, in order not to introduce very large forces in the system. For a potential $U(x) = x^2$, we obtain linearisation at a point p by $U_l(x) = pU'(x) - U(p) = 2p2x - p^2 = p(2x - p)$ In this study we take p = 4, and $U_l(x) = 8x - 16$. This formulation ensures the potential has smooth first derivatives (forces) at x = 4.

TABLE IV: Table of base pairing parameters.

	G-	G		G-A					G	-C		G-U			
ρ	α_1	α_2	k_{hb}												
6.24	5.91	4.34	2	5.07	5.81	6.09	2wc	4.80	5.67	5.31	3wc	5.20	5.67	5.05	2wc
7.33	0.21	4.41	1	6.26	1.66	4.34	2	7.40	4.42	6.03	1	7.05	0.29	4.57	1
6.24	4.34	5.91	2	7.02	5.07	5.20	1								
7.33	4.41	0.21	1	7.54	1.00	3.94	1								

	A-	G			A-	A			A-	-C		A-U			
ρ	α_1	α_2	k_{hb}												
5.07	6.09	5.81	2wc	5.43	5.69	5.69	2wc	4.80	5.74	5.19	wc	4.96	6.06	5.37	2wc
6.26	4.34	1.66	2	6.44	4.15	1.41	1	5.60	5.54	4.96	1	6.43	3.98	5.09	2
7.02	5.20	5.07	1	6.86	4.12	4.12	2	6.94	5.21	4.62	1				
7.54	3.94	1.00	1	6.44	1.41	4.15	1								

C-G C-A						A	A C-C					C-U			
ρ	α_1	α_2	k_{hb}	ρ	α_1	α_2	k_{hb}	ρ	α_1	α_2	k_{hb}	ρ	α_1	α_2	k_{hb}
4.80	5.31	5.67	3wc	4.80	5.19	5.74	wc	4.91	5.36	5.38	2	5.20	5.32	5.32	2wc
7.40	6.03	4.42	1	5.60	4.96	5.54	1					5.62	4.89	5.78	1
				6.94	4.62	5.21	1					7.48	6.02	0.43	1

U-G U-A						A	A U-C				U-U				
ρ	α_1	α_2	k_{hb}	ρ	α_1	α_2	k_{hb}	ρ	α_1	α_2	k_{hb}	ρ	α_1	α_2	k_{hb}
5.20	5.05	5.67	2wc	4.96	5.37	6.06	2wc	5.20	5.32	5.32	2wc	5.39	4.85	5.75	1
7.05	4.57	0.29	1	6.43	5.09	3.98	2	5.62	5.78	4.89	1	6.18	5.70	0.75	1
								7.48	0.43	6.02	1	5.39	5.75	4.85	1

Scale	Interaction	Value
	Bond	0.4
€a	Angle	0.3
eu Ed	Dihedral	0.1
Eau	Excluded volume	1.0
Eel	Electrostatics	1.0
eet et	Stacking	2.0
Enl	Planarity	2.0
ϵ_{pl}	Hvdrogen bond	2.4
-110		

TABLE V: Table of energy scales.