Supporting Information

for

Destabilization of Quadruplex DNA by 8-Aminoguanine

Jaime López de la Osa, Carlos González*, Raimundo Gargallo, Manuel Rueda, Elena Cubero, Modesto Orozco*, Anna Aviñó, and Ramón Eritja*

Methods for the theoretical calculations.

We analyze theoretically the impact of guanine→8-aminoguanine mutation on tetraplex structure and stability using the thrombin-binding DNA aptamer model of the anti-parallel quadruplex d(GGTTGGTGTTGGT). The experimental structure of the tetraplex was used as starting point to build two models: i) a complete system containing the central tetraplex and the connecting loop, and ii) an ideal system defined by considering only the central tetraplex. Both systems were hydrated by 2700 (tetraplex+loops) or 2200 (tetraplex) water molecules, a central ion (Na⁺ or K⁺) was placed in the centre of the tetrad, and extra Na⁺ were placed around the structure to guarantee electroneutrality. The two systems were optimized, thermalized (300 K) and equilibrated for 0.4 ns using our standard equilibration protocol.[1] The resulting systems were then analyzed during 5 ns of isothermal-isobaric molecular dynamics simulations using PBC-PME conditions, SHAKE (integration step 2 fs) and the AMBER(P99)-TIP3P force-field[2] with parameters for 8AG taken from our previous work.[3] The structures after 5 ns were used to create two models (with and without loops) of tetraplexes containing 8-aminoguanine
at the second position. These models were equilibrated for 2 additional ns of MD simulations.

The equilibrated structures of the four tetraplexes were used as starting points for thermodynamic integration calculations (MD/TI). The reversible work necessary for mutation of guanine into 8-aminoguananine was determined by slowly interchanging the two purines at position 2 of the aptamer. Following our standard procedure\[^3\] simulation was repeated starting from the equilibrated structure containing guanine and 8-aminoguanine, using double wide sampling technique. Furthermore, to even increase the statistical quality of estimates each mutation was repeated using 420 or 820 ps trajectories and 21 or 41 windows (divided in two 10 sub-windows). Overall, each individual free energy estimate is obtained by averaging on 12 independent values. Following standard thermodynamic cycles the impact of \( \text{G} \rightarrow \text{8AG} \) substitution on tetraplex stability was determined by subtracting the free energy associated to the mutation in tetraplex to that on a single strands (\( \text{d(GG)} \)) for the pure tetraplex and \( \text{d(GGTT)} \) for the tetraplex+ loop structure). The single stranded oligonucleotides were created using a protocol similar to that of tetraplexes and equilibrated for 1 ns before simulation.

**Methods for analysis of multivariate data recorded along melting experiments.**

Information about the thermodynamic behaviour of every system studied was also obtained from multivariate analysis of data recorded along melting experiments. This information consisted of the number of conformations, their concentration profiles and the pure spectrum for each one of them.

The spectroscopy data recorded along the melting experiments were analyzed with the soft-modelling MCR-ALS procedure\[^4\]. This procedure was applied to calculate the concentration profiles and the pure spectra of the spectroscopically active species present in the system from the decomposition of the experimental data matrix \( \mathbf{D} \) according to equation:

\[
\mathbf{D} = \mathbf{C} \times \mathbf{S}^T + \mathbf{E} \quad (S1)
\]

where \( \mathbf{C} \) and \( \mathbf{S}^T \) are, respectively, the data matrices containing the concentration profiles and the pure spectra for each one of the species or conformations present in the experi-
ment. \( E \) contains the residual noise not explained by the proposed species or conformations in \( C \) and \( S^T \).

The MCR-ALS procedure applied in this work consisted of the following steps:

1. Data arrangement: For an experiment simultaneously monitored by circular dichroism and molecular absorption, the recorded spectra recorded were collected in a table or matrix \( [D_{CD}; D_{abs}] \). The dimensions of this matrix were \( N_r \) rows x \( N_m \) columns, where \( N_r \) were the spectra recorded at successive temperature values and \( N_m \) was the number of wavelengths measured in CD and molecular absorption spectra.

2. Determination of the number of acid-base species or conformations, \( N \): The number of spectroscopically active chemical species or conformations \( N \) was estimated by applying several methods like Singular Value Decomposition (SVD), Evolving Factor Analysis (EFA) or SIMPLISMA.\(^5\)

3. ALS optimization: The ALS optimization procedure is an iterative method used to solve the equation 1 for the proposed number of species or conformations \( N \). This iterative process is started either with an initial estimation of the concentration profiles \( C \) or with an initial estimation of the pure spectra \( S^T \) for each one of the \( N \) species or conformations proposed. An initial estimation of \( C \) can be obtained from EFA. Alternatively, it is possible to obtain an initial estimation of \( S^T \) from SIMPLISMA.

The iterative optimization advances in two steps, the order of which depends on the initial estimation (\( C \) or \( S^T \)) used. Hence, when the ALS optimization starts with an initial estimation of the \( C \) data matrix the calculation is carried out as follows:

a) In the first step, \( S^T \) is calculated from the solution of the least squares optimization problem:

\[
\min ||D^* - CS^T|| \tag{S2}
\]

where \( D^* \) is the SVD reproduced data matrix for the \( N \) species considered, and \( C \) is the initial estimation from EFA. The constraint of non negativity in the spectral responses is applied along this optimization step, i.e., the calculated spectra in \( S^T \) are constrained to be positive or zero for molecular absorption data. This constraint is not applied for CD data.
b) In the second step, \( C \) is recalculated by solving the least squares optimization problem:

\[
\text{min} \| \mathbf{D}^* - \mathbf{CST} \| \quad (S3)
\]

where \( \mathbf{ST} \) are the pure spectra calculated in the previous step. In this case the concentration values in \( \mathbf{C} \) are constrained not only to be positive but also to give unimodal concentration profiles, i.e., only one maximum is allowed in each of the \( N \) concentration profiles. Since the total concentration of oligonucleotide is constant throughout the experiment, a closure constraint is applied for the concentration profiles in this step, i.e., the sum of the different species concentrations at each temperature value are forced to be a known constant value.

c) Steps a and b are repeated alternatively until the data matrix \( \mathbf{D}^* \) is well explained within experimental error.

Steps a and b are swapped when the ALS optimization starts with an initial estimation of the pure spectra \( \mathbf{ST} \).

All MCR calculations were performed using in-house MATLAB routines (Codes and tutorials are freely available at the URL: www.ub.es/gesq/mcr/mcr.htm).

**Oligonucleotide synthesis.**

Oligonucleotides sequences A: 5'-GGTTGGTGTGGTTGG-3', and B: 5'-GNTTGGTGTGGTTGG-3' (being \( N \)= 8-aminoguanine) were prepared on an automatic Applied Biosystems 392 DNA synthesizer. Oligonucleotides were prepared on 1 \( \mu \)mol scale using commercially available chemicals and following standard protocols. The phosphoramidite of 8-aminoguanine was prepared as described elsewhere. After the assembly of the sequences, oligonucleotide-supports were treated with 32% aqueous ammonia at 55 °C for 16 h except for the oligonucleotide bearing 8-aminoguanine. In this case, a 0.1 M 2-mercaptoethanol solution in 32% aqueous ammonia was used and the treatment was extended to 24 h at 55 °C. Ammonia solutions were concentrated to dryness and the products were purified by reversed-phase HPLC. HPLC conditions: HPLC solutions were as follows. Solvent A: 5% acetonitrile in 100 mM triethylammonium acetate pH 6.5 and solvent B: 70% acetonitrile in 100 mM triethylammonium acetate pH 6.5. Columns:
PRP-1 (Hamilton), 250 x 10 mm. Flow rate: 3 mL/min. A 30 min linear gradient from 10-80% B (DMT on) or a 30 min linear gradient from 0-50% B (DMT off). Oligonucleotides were converted to the sodium salt by passing through a Dowex 50Wx4 (Na⁺ form) followed by desalting over a Sephadex G25 (NAP-25) column eluted with water. Yield. Sequence A: 80 O.D.₂₆₀ units, sequence B: 52 O.D.₂₆₀ units.

References.

**Table S1.** Change in stability free energy (in kcal/mol) of the tetraplex system (pure and containing loops; see methods) upon G→8AG mutation. A positive sign means that the presence of 8AG destabilizes the tetraplex. Standard errors in the simulations are shown in parenthesis.

<table>
<thead>
<tr>
<th>Quadruplex</th>
<th>Mutation</th>
<th>Central Ion</th>
<th>ΔΔG [kcal mol⁻¹]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetraplex</td>
<td>G→8AG</td>
<td>Na⁺</td>
<td>1.14 (0.1)</td>
</tr>
<tr>
<td>Tetraplex+loop</td>
<td>G→8AG</td>
<td>Na⁺</td>
<td>1.94 (0.3)</td>
</tr>
<tr>
<td>Tetraplex</td>
<td>G→8AG</td>
<td>K⁺</td>
<td>1.18 (0.1)</td>
</tr>
<tr>
<td>Tetraplex+loop</td>
<td>G→8AG</td>
<td>K⁺</td>
<td>1.55 (0.3)</td>
</tr>
</tbody>
</table>

**Table S2.** Major components of the interaction energy (in kcal mol⁻¹) for the first tetrad (that where the mutation is introduced) of the d(GNTTGGTTGGTGGTTGG) tetraplex. Values and standard deviations are obtained by averaging data from the last ns of the respective trajectories.

<table>
<thead>
<tr>
<th>K⁺ ion</th>
<th>Na⁺ ion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G 8AG</td>
</tr>
<tr>
<td>Stacking</td>
<td>-107.50±5.1</td>
</tr>
<tr>
<td></td>
<td>-105.16±5.0</td>
</tr>
<tr>
<td>H-Bonding</td>
<td>-21.29±3.2</td>
</tr>
<tr>
<td></td>
<td>-21.16±3.5</td>
</tr>
<tr>
<td>Nucleobase-</td>
<td>-139.44±6.1</td>
</tr>
<tr>
<td>ion</td>
<td>-139.98±5.9</td>
</tr>
<tr>
<td>Total</td>
<td>-268.23±8.6</td>
</tr>
<tr>
<td></td>
<td>-266.30±8.5</td>
</tr>
</tbody>
</table>
Figure S1. Melting of sequence 5’-GGTTGGTGTGGTTGG-3’ monitored by CD and molecular absorption spectroscopies, [D\text{CD}, D\text{abs}]. A: Experimental data; B: resolved concentration profile, C; C: resolved pure CD spectra, S\text{CD}; D: resolved pure molecular absorption spectra, S\text{abs}.

Continuous line: initial quadruplex conformation, dotted line: final disordered conformation.
Figure S2. Melting of sequence 5’-GNTTGGTGTGGTTGG-3’ monitored by CD and molecular absorption spectroscopies, \([D_{CD}, D_{abs}]\). A: Experimental data; B: resolved concentration profile, \(C\); C: resolved pure CD spectra, \(S_{CD}^T\); D: resolved pure molecular absorption spectra, \(S_{abs}^T\). Continuous line: initial quadruplex conformation, dotted line: final disordered conformation.
Figure S3: CD melting curves of the 15-bases long thrombin aptamer 5'-GGTTGGTGTGGTGTTGG-3' (dark circles) and the 8-aminoguanine modified sequence 5'-GNTTGGTGTGGTTGG-3' (being N= 8-aminoguanine) (open circles) (5 μM KCl, pH=7).

Table S3: Thermodynamic parameters for quadruplex formation determined by van't Hoff analysis of CD melting curves.

<table>
<thead>
<tr>
<th></th>
<th>$T_m$</th>
<th>$\Delta G_{298}$</th>
<th>$\Delta H$</th>
<th>$\Delta S^0$[a]</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-GGTTGGTGTGGTGTTGG-3’</td>
<td>46</td>
<td>-2.7±0.2</td>
<td>-40±1</td>
<td>-126±1</td>
</tr>
<tr>
<td>5'-GNTTGGTGTGGTGTTGG-3’</td>
<td>39</td>
<td>-1.6±0.2</td>
<td>-35±1</td>
<td>-111±1</td>
</tr>
</tbody>
</table>

[a] The destabilization induced by the G→8AG substitution is enthalpic in nature. However, the modified oligonucleotide is entropically favored. This effect may be due to the larger flexibility of the modified quadruplex compared to the unmodified one. This larger flexibility is suggested by the broader signals observed in the NMR spectra of the modified quadruplex.