Check for updates

Ligand-induced conformational preorganization of loops of *c-MYC* G-quadruplex DNA and its implications in structure-specific drug design†

S. Harikrishna, [ID] * Saikiran Kotaru [ID] and P. I. Pradeepkumar [ID] *

Stabilization of a G-quadruplex (G4) DNA structure in the proto-oncogene *c-MYC* using small molecule ligands has emerged as an attractive strategy for the development of anticancer therapeutics. To understand the subtle structural changes in the G4 structure upon ligand binding, molecular dynamics (MD) simulations of *c-MYC* G4 DNA were carried out in a complex with six different potent ligands: **3AQN**, **6AQN**, **3APN**, **360A**, **Nap-Et**, and **Nap-Pr**. The results show that the ligands **3AQN**, **6AQN**, **3APN**, and **360A** stabilize the G4 structure by making stacking interactions with the top quartet. On the other hand, **Nap-Et** and **Nap-Pr** bind at the groove of the G4 structure. These groove binding ligands make crucial H-bond contacts with the guanines and electrostatic interactions with the phosphate backbone. Two-dimensional dynamic correlation maps unraveled the ligand-induced correlated motions between the guanines in the quartet and a di-nucleotide present in the propeller loop-2 of the G4 structure. Cluster analysis and ONIOM calculations revealed the structural dynamics in the loop of the quadruplex upon ligand binding. Overall, the results from the present study suggest that engineering specific contacts with the propeller loop can be an efficient way to design *c-MYC* G4-specific ligands.

Introduction

G-quadruplexes (G4) are nucleic acid secondary structures that consist of π–π stacked G-tetrads (quartets), which are formed by eight hydrogen bonds using both the Watson–Crick and the Hoogsteen edges of guanines (Fig. 1A).[1] The G4 structures have diverse structural topologies that depend on intervening loop lengths, sequence, and local environment.[2–4] The G-rich sequences, which have the propensity to form G4 structures, are present in the single stranded tandem repeat regions of telomeres,[5] promoter regions,[6] introns,[7] and the untranslated regions of mRNAs.[8] The G-rich sequences present in the promoter regions of several oncogenes such as *c-MYC*,[9] *c-KIT*,[10,11] *VEGF*,[12] and *KRAS*[13] have the potential to form intramolecular parallel G4 structures. The human *c-MYC* gene is a transcription factor that regulates the expression of proteins involved in cell growth and proliferation.[14] Overexpression of *c-MYC* is associated with

cancer.[14,15] Nuclease hypersensitive element III₁ (NHE), which is present in the promoter region of *c-MYC*, controls 90% of the transcription.[12] This element contains a 27nt (−142 to −115 bp) G-rich sequence that can form a G4 structure.[6] This wild-type 27mer was shown to form parallel G4 with different propeller loop isomers as revealed by NMR studies.[16] The structure of a modified 22mer (Fig. 1B), which adopts a single predominant parallel G4 conformer, was determined by NMR (PDB ID: 1XAV), and it has emerged as an excellent model for the *in silico* structure-based drug design using *c-MYC* G4 DNA.[9,17]

Small molecules have been reported to stabilize the *c-MYC* G4 structure and down regulate gene expression.[6] These include TmPyP4,[17] expanded porphyrin,[18] quarfloxin,[19] derivatives of quindoline,[20,21] and metal complexes.[22] The dynamic characteristic features of the G4 loops as well as the ligand-induced loop conformers of the *c-MYC* G4 structure upon interactions with TmPyP4 reveal various conformational changes in the G4 structure upon ligand binding.[23] Along these lines, NMR analysis and molecular dynamics (MD) simulations of **Phen-DC₃** with *c-MYC* G4 (24mer, 1 : 1 ligand : G4 stoichiometry, PDB entry: 2MGN) reveal the correlated movements between the quinolinium moiety and the guanine bases in the top quartet.[24] The solution structure of quindoline with the *c-MYC* (2 : 1 ligand : G4 stoichiometry, PDB ID: 2L7V) provides insights into the structure-based drug design of small molecules, which can specifically target the promoter G-rich elements.[25] In addition, a structure-based

Department of Chemistry, Indian Institute of Technology Bombay,
Mumbai – 400076, India. E-mail: harikrishna.s@iitb.ac.in, pradeep@chem.iitb.ac.in

† Electronic supplementary information (ESI) available: Energy optimized structures of ligands at HF/6-31G* level; stacking distance and angle between ligand and G4 DNA; non-covalent interactions between *c-MYC* G4 and ligands; conformational changes in the groove binding ligands; dynamic cross-correlation map of ligand free *c-MYC* G4 DNA; percentage occupancies of the clusters of loop-2 conformers obtained from MD simulations; SASA values of the *c-MYC* G4 DNA. See DOI: 10.1039/c7mb00175d
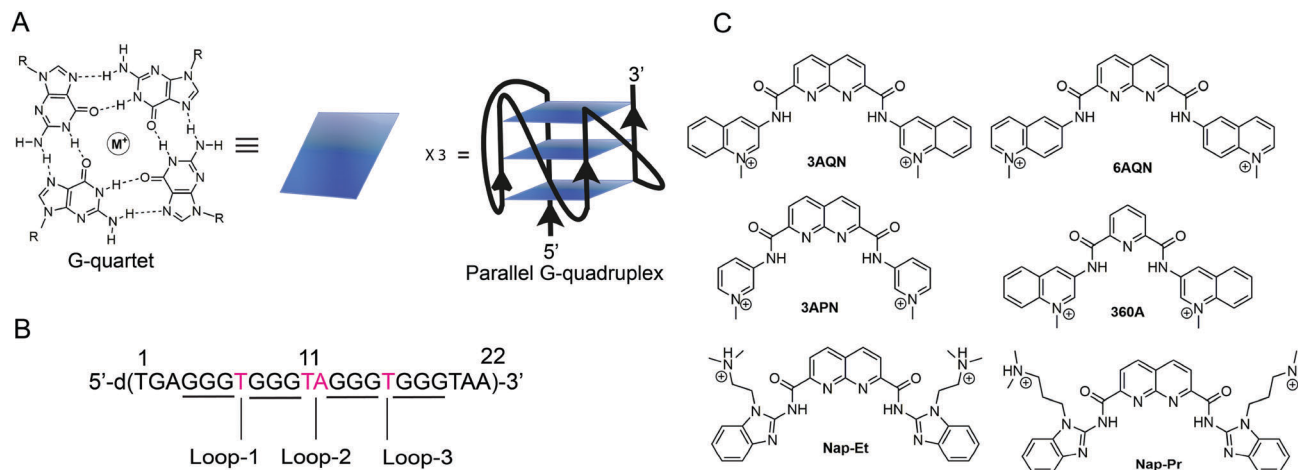
Fig. 1 Schematic representation of *c-MYC* G4 DNA, its sequence, and the chemical structures of the small molecule ligands used in the molecular dynamics (MD) simulation study. (A) Structure of G-quartet showing the eight Hoogsteen H-bonds (dashed lines) between the four guanines. R indicates the deoxyribose sugar. Three such G-quartets stack on each other to form the G4 DNA. The central electronegative oxygen atoms are stabilized by monovalent cations (M^+). (B) Sequence of the 22mer *c-MYC* G4 DNA used for structural studies. (C) Chemical structures of G4 stabilizing small molecule ligands.

virtual screening identified fonsecin B[26] and carbamide[27] natural products as specific stabilizers of the *c-MYC* G4 structure. A fragment-based drug design approach was also used to identify molecules that target the *c-MYC* G4 DNA.[28] Recently, we have reported the specific stabilization of *c-MYC* and *c-KIT* G4 DNAs by indenopyrimidine derivatives,[29] indolylmethyleneindanone scaffolds,[30] and benzimidazole derivatives of napthyridine and phenanthroline anchoring novel benzimidazole-based side chains.[31] These topology-specific ligands are able to recognize the propeller loops along with the G-quartets of the parallel G4 structure as revealed by MD simulations.[29,31]

Although there is a growing list of G4 stabilizing agents, none of the small molecules show high affinity and specificity toward the stabilization of *c-MYC* G4 DNA. To rationally design small molecules that can specifically stabilize a particular G4 topology, it is very important to study the conformational dynamics of the G4 DNA in the absence and presence of ligands.[32–34] Recently reported MD simulation studies of the *c-KIT1* G4 structure highlight the conformational dynamics and internal stability of the loops.[35,36] To gain insights into the subtle conformational changes that occur upon ligand binding at the molecular level, and to explore structure-specific ligand design, comprehensive structural probing of a number of well-known ligands with *c-MYC* G4 structures is warranted.[37,38] We have chosen potential G4 stabilizing agents **3AQN**, **6AQN**, **3APN**, **360A**, **Nap-Et**, and **Nap-Pr** (Fig. 1C) to probe the structural dynamics of G4 DNA.[31,39,40] Among these, **3APN**, **Nap-Et**, and **Nap-Pr** were able to selectively stabilize the parallel promoter G4 DNAs over telomeric and ds DNA.[31,39] Various analyses from MD simulation trajectories have been harnessed to unravel the binding interactions of ligands toward G4 DNA. The two-layered (DFT: MM AMBER) ONIOM (Our N-layered Integrated molecular Orbital and molecular Mechanics) method was used to evaluate the interaction energies between the ligands and G4 DNA.[41] Also, the ONIOM calculations were used to assess the energy of the loop conformers obtained from the MD trajectories. The results

highlight that the ligand binding affects the local and global *c-MYC* G4 structures and imparts conformational flexibility to the propeller loop.

Computational methods

Ligand preparation

The chemical structures of the ligands used in our study are shown in Fig. 1. The energy optimized (HF/6-31G*) structures of **3AQN**, **6AQN**, **3APN**, and **360A** were previously reported by us.[39] The **Nap-Et** and **Nap-Pr** were optimized at the HF/6-31G* level using Gaussian 09. Atom types and bond types were assigned using the antechamber program of AMBER 14.[42] The conformers with the lowest energies were taken for molecular docking.

Molecular docking

The NMR structure of the *c-MYC* G4 DNA was used (PDB entry: 2L7V) as a receptor for performing molecular docking using Autodock 4.2.[43] The ligand molecule present in the PDB structure was removed to dock the ligands of interest. The torsion angles of the ligands were made flexible for docking. The docking area was centered on the Cartesian coordinates at the center of mass of the G4 structure and defined by a grid box, which was large enough to include the whole macromolecule. The grid points of $92 \times 92 \times 92$ with a 0.375 Å spacing were calculated in the grid box for all the ligand atom types using AutoGrid 4.2. The Lamarckian genetic algorithm was used for docking with a maximum number of 25 000 000 energy evaluations with an initial population of 200 randomly placed individuals having a mutation rate of 0.2 along with a maximum number of 27 000 generations. A crossover rate of 0.85 and 300 iterations of local search were used. Finally, 250 independent docking runs were carried out for each ligand. Each docking run consisted of 10 million energy evaluations using the local search method implemented in the genetic algorithm. The output docking

conformations were clustered based on the RMSD between the Cartesian coordinates of the ligand atoms (cutoff = 1.5 Å) and were ranked based on the scoring function. The best docked conformations were selected with the lowest RMS deviations and low binding energies for further MD simulations.

Molecular dynamics

MD simulations were carried out with the SANDER and Particle Mesh Ewald Molecular Dynamics (PMEMD) module in AMBER 14.[44] The Restrained Electrostatic Potential Atomic (RESP) charges of the ligands were calculated using Gaussian at the HF/6-31G*(d) level, and were fitted using the antechamber RESP fitting procedure.[45] Generalized AMBER Force Field (GAFF) parameters were used for the ligand, and parmBSC1 force field parameters were used for the G4 DNA.[46–48] The G4–ligand complexes were loaded in the xleap program in AMBER 14 and potassium ions were added to neutralize the charge of the backbone of DNA. Each complex was immersed in a 10 Å octahedron box from any atom of the solute with TIP3P water molecules. Approximately 6000–6500 water molecules were used to solvate each system. Water molecules and counter ions were energy minimized by 7000 steepest descent energy minimizations, and then by conjugate gradient minimization with a convergence of root mean square gradient around 0.1 kcal mol^{-1} Å. The entire system was then subjected to 2500 steps of steepest descent and conjugate gradient minimization. The system was then heated from 0 K to 298 K with constant pressure using a weak-coupling algorithm. Solvent molecules were relaxed using short MD simulations (250 ps) by imposing 30 kcal mol^{-1} restraint on the solute atoms at a temperature of 100 K. The system was then heated from 100 K to 300 K in 150 ps. This was followed by 5 stages of minimization (2000 steps) and 10 stages of MD simulations (30 ps) carried out with restraints of 5, 4, 3, 2, and 1 kcal mol^{-1} on solute atoms. Finally, unrestrained production MD simulations were performed for 300 ns using the CUDA version of PMEMD[49] in a GPU accelerated version[50,51] of AMBER 14.[42] The Particle Mesh Ewald (PME) method was used for calculating the contributions from the non-bonded interactions with a cutoff of 10 Å. The unrestrained MD simulations were performed in the NPT ensemble. A constant pressure of 1 atmosphere was maintained using a Bendersen weak-coupling barostat in a time constant of 1 ps.[52] The MD simulations temperature (300 K) was maintained by the Bendersen thermostat in a time constant of 4 ps. The same MD simulations protocol was followed for the ligand-free G4 DNA. The temperature was controlled by Langevin dynamics, and was kept constant at 298 K. The electrostatic interactions were taken into account using the PMEMD with a cutoff distance of 10 Å.[50] The SHAKE algorithm was used to constrain all the hydrogen atoms with a 2 fs integration time step. The output trajectory files were saved for every 1 ps for further analysis. The MD trajectories were visualized using UCSF Chimera.[53]

The RMSD of heavy atoms of the ligand, heavy atoms of the dG present in the G-quartet, and heavy atoms of the DNA backbone (P, O3′, O5′, C3′, C4′ and C5′) and the RMSF of each nucleotide (heavy atoms) in the G4 were calculated using the CPPTRAJ module in AMBER 14. Inter atomic distances and H-bond occupancies were also calculated using the CPPTRAJ module in AMBER 14. PyMOL (www.pymol.org) was used to render the figures.

MM-PBSA calculations

The binding-free energies were estimated using the MM-PBSA module implemented in AmberTools. For these calculations, apart from two ions present inside the G-quartet, all the ions and water molecules were removed from the MD trajectories. Free energies were estimated by collecting the structures from the MD trajectories at 20 ps intervals during the last 200 ns of the MD simulations. The binding energy was calculated using the expression $\Delta G_{bind} = G_{(quadruplex:ligand\ complex)} - [G_{(quadruplex)} + G_{(ligand)}]$, and each of the terms was estimated from $\Delta E_{MM} + \Delta G_{SOLV} - T\Delta S$. The ΔE_{MM} was obtained from the sum of ΔE_{INIT} (sum of bond, angle, and dihedral energies), ΔE_{ELEC} (electrostatic interactions), and ΔE_{vdW} (van der Waals interactions). The solvation energy term (ΔG_{SOLV}) was calculated from the sum of polar and nonpolar solvation energy terms using the adaptive Poisson–Boltzmann solver (APBS) program. The polar solvation term was calculated with a solvent dielectric constant of 80 and a solute dielectric constant of 1. The nonpolar solvation term was calculated with a γ value, which was set to 0.0072 kcal^{-1} (mol Å^2) and a β value, which was set to 0. The solvent accessible surface area (Å^2) term to include in the solvation energy term was estimated using the MOLSURF algorithm. The entropy was calculated using the NMODE, which computes the translational, rotational, and vibrational entropies. The coordinates of each snapshot were minimized using the conjugate gradient minimization for 10 000 steps, and the terminating criterion was set to 0.01 kJ mol^{-1} Å^{-1}.

Dynamic cross-correlation matrix

For each of the systems, the ligand from the MD trajectory was removed, and the DNA structures were fit to the initial structure using backbone atoms (P, O3′, O5′, C3′, C4′ and C5′) as the reference. The Dynamic Cross-Correlation Matrix (DCCM) was computed using the matrix correlation function in CPPTRAJ and DCCM tools in the Bio3D package.[54] The atomic fluctuations were calculated using the Pearson correlation coefficient of the covariance matrix. These coefficients are the measure of the linearly correlated motion between the atoms in the nucleotides. DCCM included cross-correlation along the diagonal and off-diagonal cross-correlations. The correlation of the nucleotides in the orthogonal direction was calculated from the linear mutual information (LMI), as implemented in the Bio3D package.[54]

Cluster analyses

Cluster analyses of the MD trajectories were carried out using USCF Chimera and CPPTRAJ module in AmberTools 12. All solute and solvent atoms, except the G4 and ions in the channel, were removed from each of the MD trajectories. For each of the systems, the cluster analysis was carried out using hierarchical agglomerative clustering after calculating the pairwise RMSDs. The 5′ and 3′-end flanking nucleotides of the G4

were not considered. Since the quartet stems were found to be stable, the clustering was based on the loops. Two closely related clusters (loop-2) were merged into one cluster after one round of clustering iteration. For each system, the cluster analyses were iterated until one to ten clusters were remaining. Using the number of frames >10% of the total simulation time, and the common cluster in all the systems with RMSD <0.5 for loop-2, the conformers from the ten clusters were restricted to six. The best representative conformers from the six clusters were extracted to show the noncovalent interactions and orientations.

ONIOM calculations

The structures obtained from the cluster analysis of the MD trajectories were used as initial geometries to calculate the energies of the loops using DFT/MM calculations as implemented in the ONIOM method.[55] Sodium ions were added 2 Å away from each phosphate atom and the net charge of the system was set to zero. The loop-2 (dT11 and dA12) and the guanines (dG9, dG10, and dG15) were set at a high layer (M06-2X-dzvp),[56] and the other part of the DNA was set at the low layer (molecular mechanics, AMBER). The atomic charges of all the atoms were assigned using the AMBER parmbsc1 force field. For the high layer, the charges were determined using the RESP restrained fit by the Kollman protocol.[45] Geometry optimization was carried out using direct inversion in the iterative subspace (GDIIS) method.[57] The effects of solvent (water) were studied using the ONIOM-PCM polarized continuum model.[58] The electrostatic embedding method was employed using a recently reported procedure.[59] The energy of the loop from the G-quartet stem was separated using an earlier method.[60] DNA was solvated by a cubic box of water molecules in xleap. Approximately 100 solvent molecules were retained, and the rest of them were removed using a solute-solvent cut-off distance of 3 Å.

The averaged structures from the MD simulations of the ligand-G4 complexes were used as the initial geometries to calculate the interaction energies in the gas phase and in the aqueous phase (water-PCM model). The ligands were set to a high layer of theory (B3LYP/6-311G(d,p)) along with the interacting nucleotides, and the non-interacting parts of the DNA were kept to the low layer (molecular mechanics, AMBER). The interaction energies were calculated using a reported protocol.[55]

Results and discussion

Molecular docking

Molecular docking employing Autodock 4.2[43] provided the most suitable poses for ligand interactions with the *c-MYC* G4 DNA.[9] For each ligand, 250 docked poses were generated, and the maximum number of poses and energy showed that the binding site of ligands with bisquinolinium and bispyridinium side chains (**3AQN, 6AQN, 3APN,** and **360A**) was at the top (5′-end) of the G-quartet. The binding mode, interactions, and conformations of these ligands were similar to those of the NMR structure[24] reported for the *c-MYC* G4 **Phen-DC₃** complex. Also, the binding sites of the ligands with benzimidazole side

chains (**Nap-Et** and **Nap-Pr**) were found to be at the groove (>200 docked poses) of the G4 structure. To rule out the possible stacking modes of **Nap-Et** and **Nap-Pr**, a smaller docking grid around the top-quartet was employed. However, the binding energy values obtained from docking did not favor such a binding mode. The RESPs for the docked conformations of the ligands were generated at the HF/6-31G*(d) level using Gaussian 09. Consequently, 300 ns of unrestrained MD simulations were carried out on the six G4–ligand complexes. To compare the ligand-induced conformational changes in the *c-MYC* G4 DNA, MD simulations (PDB entry: 1XAV) were carried out for 300 ns in the absence of ligands.

Overall structural stability

To find the structural and conformational stabilities of the complexes, the RMSDs of the DNA backbone, G-quartets, and ligands were calculated. The RMSD graphs of the G-quartets in comparison with the ligand-free G-quartet indicate that all the ligands stabilize the G4 efficiently (Fig. 2). The backbone atoms of the DNA were converged in a maximum of 40–50 ns in all the complexes. Therefore, the subsequent energy calculations were performed using the trajectories obtained after 50 ns. The RMSD graphs show that ligands including **6AQN, Nap-Et** and **Nap-Pr** are flexible and **3AQN, 3APN** and **360A** are rigid when they form a complex with the G4 DNA. To investigate the fluctuation of the nucleotides in the G4, root-mean-square fluctuations (RMSFs) were calculated. The RMSF graphs suggest that the loops and flanking nucleotides in the G4 DNA are less fluctuated in the complexes of **3AQN, 3APN** and **360A** when compared to the ligand-free G4 DNA (Fig. 3). In contrast, for G4 DNA in complex with **6AQN, Nap-Et,** and **Nap-Pr**, the flanking nucleotides in the 3′-end are more fluctuated in comparison with the ligand-free G4 structure. These observations indicate that the rigid ligands such as **3AQN, 3APN,** and **360A** can minimize the dynamic behavior of the flanking nucleotides, which might result in further stabilization.

Binding interaction of the ligands with *c-MYC* G-quadruplex DNA

The binding mode and the key noncovalent interactions of **3AQN, 6AQN, 360A,** and **3APN** with *c-MYC* G4 DNA are shown in Fig. 4 and 5. Initially, to unravel the stability of the G4–ligand interactions, the percentages of the π–π stacking interactions between the ligand and top G-quartet were calculated from the last 250 ns of the 300 ns MD trajectories. In the case of **3APN**, one of the pyridinium side chains was not found to stack on the guanine quartet during ∼61% of the simulation time, while the central core and the other side chains of the ligand stacked well on the G-quartet (Fig. 4E, F and Fig. S2 of the ESI†). However, in the case **3AQN, 6AQN,** and **360A**, the stacking interactions between the G-quartet and the ligands were found to be present >95% of the simulation time. Hence, all these ligands favor the strong π–π stacking interactions with the top G-quartet. The optimal distances between the side chains in the ligand were found to be 7.8 to 9 Å, which were identified based on highly favored stacking interactions between the ligands and the G-quartet (Fig. 4B, D, F, 5B and Fig. S2 of the ESI†). The dG2 in
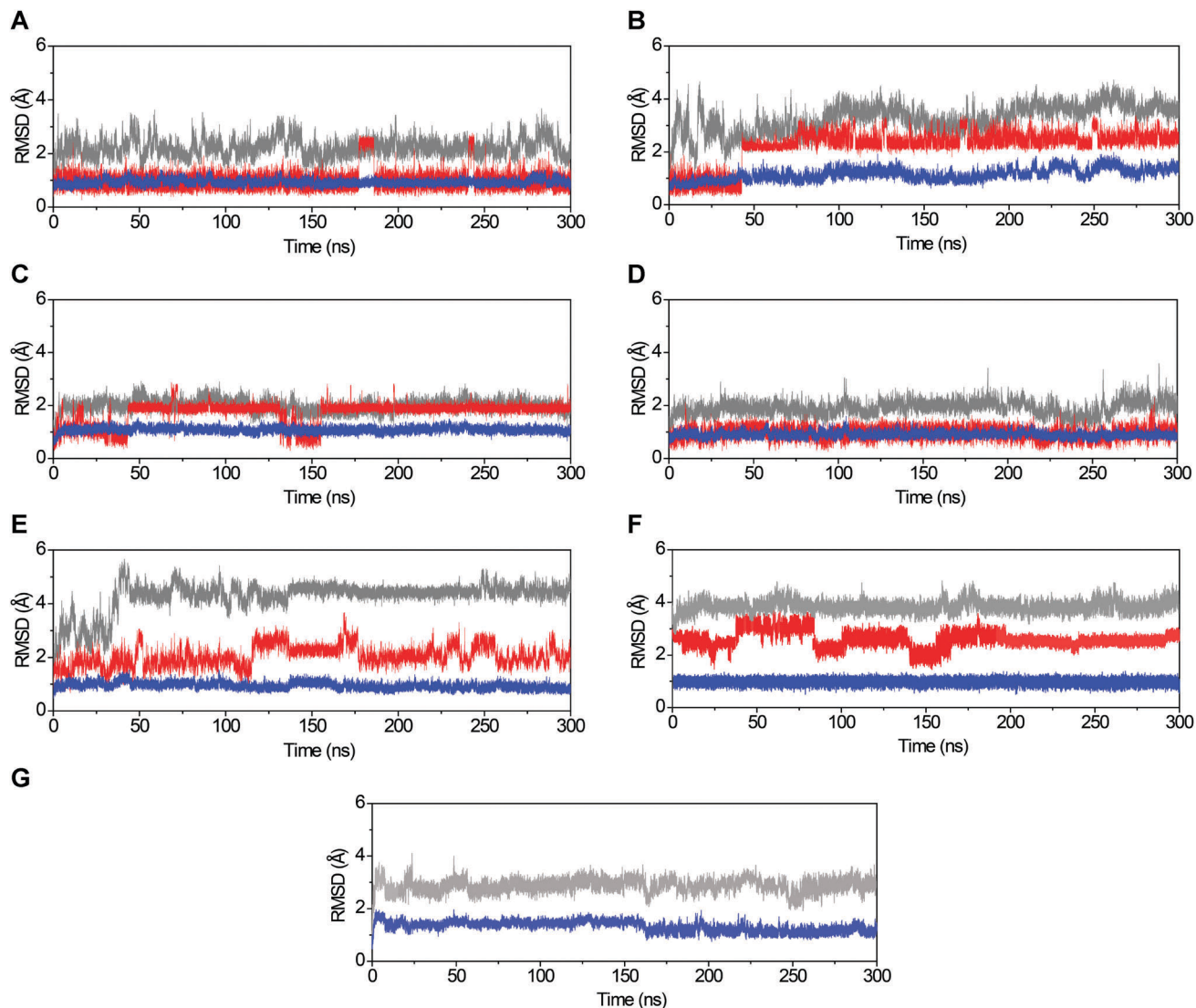
Fig. 2 RMSD graphs of the *c-MYC* G4 DNA and ligand complexes during the 300 ns of MD simulations. The *c-MYC* G4 DNA in complex with the (A) **3AQN**, (B) **6AQN**, (C) **3APN**, (D) **360A**, (E) **Nap-Et**, (F) **Nap-Pr** ligands, and (G) the ligand-free *c-MYC* G4 DNA. The gray, blue, and red lines indicate the RMSDs for the DNA backbone, G-quartet, and ligands, respectively.
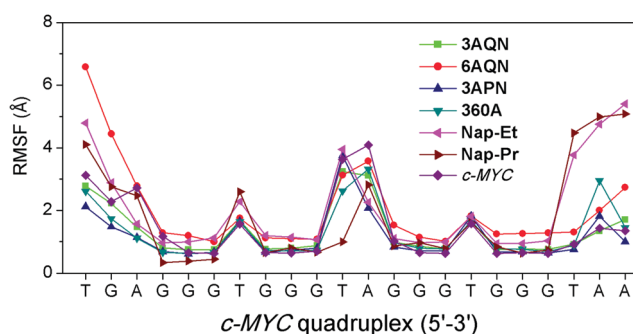

Fig. 3 RMSF plots of each nucleotide (heavy atoms only) in the *c-MYC* G4 DNA and the complexes formed by the six ligands as a function of time.

the 5′-flanking nucleotides of the *c-MYC* G4 stacked on the ligands **3AQN** and **360A**, and such interactions were not observed for **6AQN** and **3APN** (Fig. S3 of the ESI†). However, H-bond interactions

were seen in the complexes of **3AQN** and **3APN** with the 5′-flanking nucleotides (Fig. S3 of the ESI†). We also observed high occupancies of non-covalent interactions between the ligand side chains and the flanking nucleotides, which resulted in less dynamics of the flanking nucleotides. In addition, the intra molecular hydrogen bond between the NH of the side chain and the nitrogen in the ring locked the planar conformation of the **3AQN** (98%), **6AQN** (68%), **3APN** (64%), and **360A** (95%) during the course of the MD simulations (Fig. 4 and 5).

The binding modes and major noncovalent interactions of **Nap-Et**, and **Nap-Pr** with *c-MYC* G4 DNA are shown in Fig. 5, which shows that these ligands bind at the groove of the G4 structure. Very few classes of small molecules are reported to bind at the groove and stabilize G4 DNAs.[61–63] The non-covalent interactions, which favor the groove binding modes with *c-MYC* G4 DNA, were analyzed in detail. The oxygen in the two carbonyl groups and nitrogen atoms in the naphthyridine core
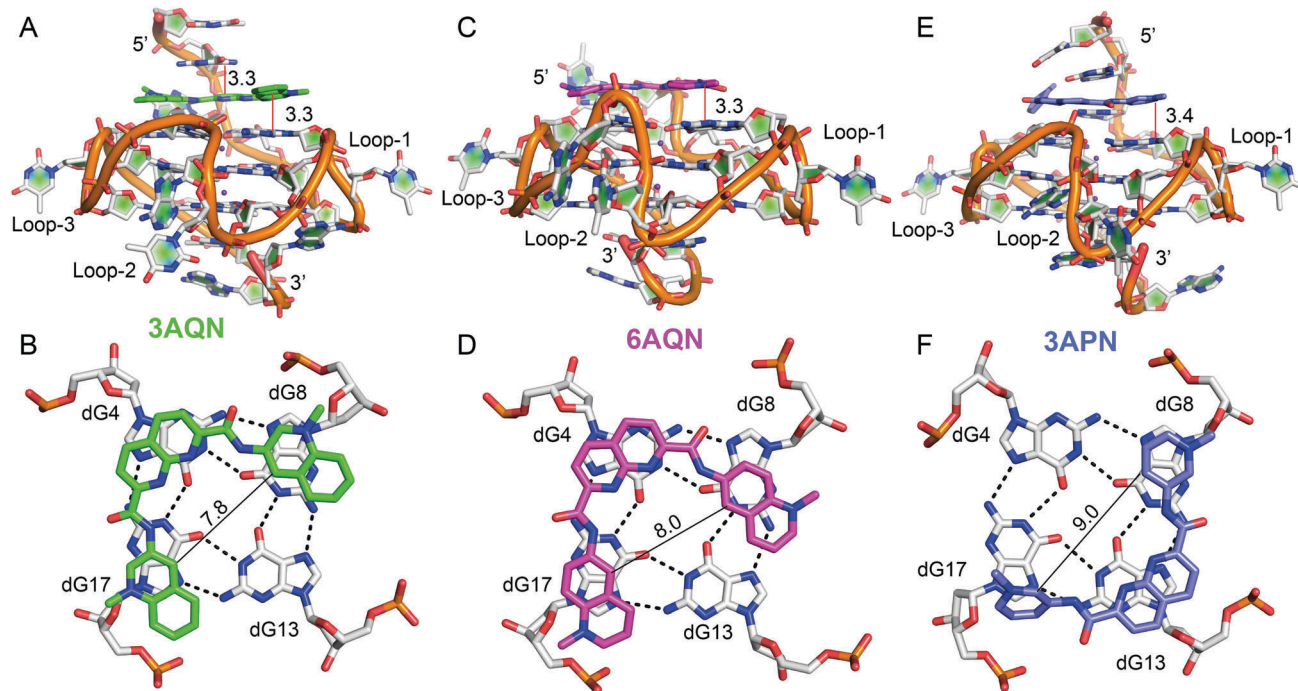
Fig. 4 Best representative structure from the MD simulations (300 ns) of the ligands with the *c-MYC* G4 DNA. Ligands (A) **3AQN**, (C) **6AQN**, and (E) **3APN** stack on the top G-quartet and show their interactions with the 5′-flanking nucleotides. Top quartet (axial view) and ligands (B) **3AQN**, (D) **6AQN** and (F) **3APN**, which form stacking interactions with the G-quartet. The red solid lines in A, C, and E represent the distances between the top quartet and the ligand. The black solid lines in B, D, and F represent the distances between the side chains of the ligands. These distances were measured between two carbon atoms of the side chains as drawn in the figures. The black dotted lines indicate the Hoogsteen hydrogen bonds between the dGs. The backbone of the DNA is shown as a cartoon, the atoms are in stick representation, and the ligands are highlighted in different colors. K$^+$ ions are shown as non-bonded purple spheres. All the distances are mentioned in Å.

of **Nap-Et** and **Nap-Pr** make H-bonds with the NH$_2$ hydrogen of the dG17, dG18, and dG19 present in the G-quartets (Fig. S4A and C of the ESI†). The two positively charged side chains in the ligands mediate electrostatic interactions with the negatively charged backbone of the G-quartet as observed earlier[64] (Fig. S4B and D of ESI†). These electrostatic interactions are present ~79% and ~68% of the simulation time, respectively. These results indicate that the two-carbon linker side chain, as in **Nap-Et**, has the optimal length to facilitate strong electrostatic contacts.

Computation of the average solvent-accessible surface area (SASA) can give direct insights into the binding process of the G4 stabilizing ligands. The difference in the surface accessibility of G4 DNA (ΔSASA) on ligand binding was calculated using the Surf tool following a reported procedure.[65] The SASA of native *c-MYC* G4 DNA is 3274 Å^2 and the SASAs of G4 DNA in the complex with the ligands are shown in Fig. 6 and Table S1 of the ESI.† The ΔSASA for the groove binding ligand is ~70 Å^2 and for the top quartet binding ligands the ΔSASA is ~350 Å^2. This is expected because the ligand stack on the G-quartet maximizes the nonpolar π–π stacking interactions by reorienting the flanking nucleotides. The increase in the averaged SASA values further validates the ligand-induced conformational flexibility in the *c-MYC* G4 DNA upon ligand binding.

Binding-free energies

The energetic parameters driving the interactions between the ligands and G4 DNA were investigated using the MM-PBSA method. The results are shown in Table 1. The ΔGs of the four ligands, which bind at the top of the G-quartet, are almost similar (~−45 kcal mol^{-1}), which can be ascribed to the similarities in the chemical structures and the binding modes of these ligands. The groove binding ligands including **Nap-Et** and **Nap-Pr** have slight differences in the free energies, which could arise from the difference in their electrostatic interactions with the G4 DNA (Table 1). This is indeed reflected in the percentage occupancies of the electrostatic interactions, as mentioned earlier. The individual free energy components reveal that in all the complexes, van der Waals (vdW) contribution to the binding is favored by −57 to −65 kcal mol^{-1}. Also, the electrostatic contribution is strong enough to compensate the contribution from polar solvation (ΔPB$_{SOLV}$) during the complex formation. These electrostatic contributions for the G4–ligand complex formation are in agreement with the results from a recently reported docking simulation protocol.[66] Overall, the electrostatic (ΔE_{ELEC}), van der Waals (ΔE_{vdW}), and nonpolar solvation (ΔPB$_{NP}$) energies contribute favorably to the formation of all the complexes. These results also underscore the importance of positively charged ligands having aromatic units for the binding and stabilization of G4 DNAs.

To accurately examine the contributions from the H-bonding and stacking interactions, inclusion of the polarization effect in the binding energy calculations is required. QM/MM is a computationally inexpensive method, which includes the polarization effects and has been employed to study G4–ligand interactions.[58,67] Along these lines, the binding energies for the
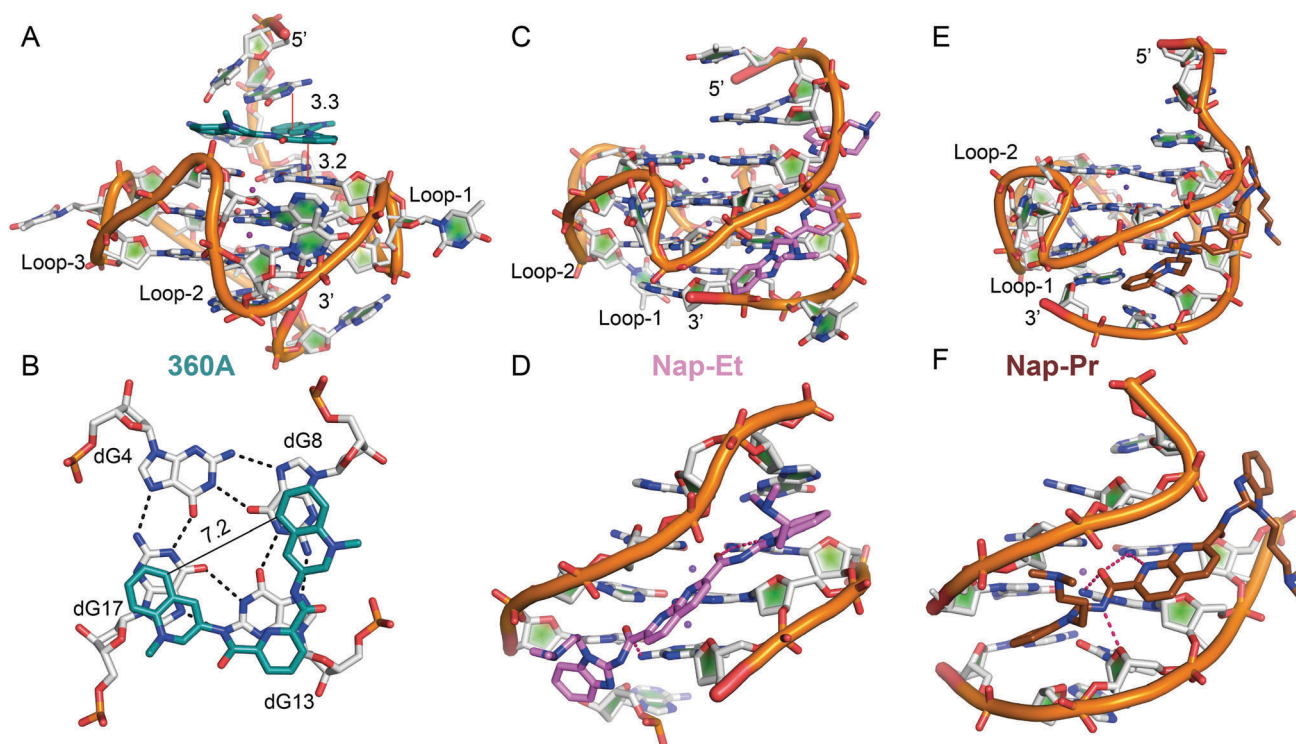
Fig. 5   Best representative structure from the MD simulations (300 ns) of ligands with the *c-MYC* G4 DNA. (A) **360A** stacks on the top G-quartet, (C) **Nap-Et** and (E) **Nap-Pr** bind at the groove of the G4. Top quartet (axial view) of (B) **360A** showing the stacking interaction with the G-quartet and the zoomed view of (D) **Nap-Et** and (F) **Nap-Pr** in the G4 groove. The black solid lines in A and B represent the distances between the nucleotide and the ligand, and the atoms in the side chains of the ligand, respectively. The red dotted lines indicate the H-bond interactions. The hydrogen bond distance between the carbonyl oxygen in the ligand and the amine hydrogen of the guanine base is 2.1 Å. The backbone of the DNA is shown as a cartoon, the atoms are in stick representation, and the ligands are highlighted in different colors. All the distances are mentioned in Å.
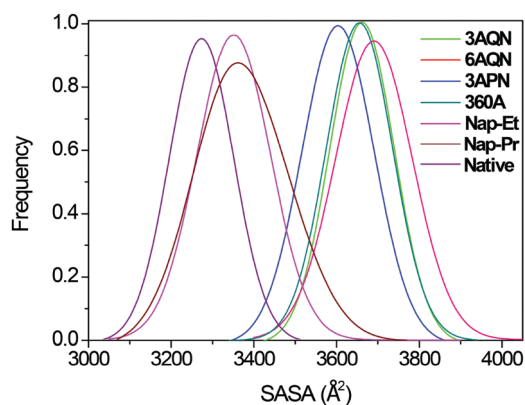
Fig. 6   The normalized frequency of SASA of the ligand-free *c-MYC* G4 DNA and in complex with the ligands during 300 ns of MD simulations. The values are listed in Table S1 of the ESI.†

c-MYC G4–ligand complexes were calculated from the averaged structures emerged from MD simulations using the ONIOM method (Table 2).[41] The ligands and interacting nucleotides (Fig. 4 and 5) were set at a high layer (B3LYP/6-311G(d,p)) and the other part of the DNA was set at a low layer (molecular mechanics, AMBER). The final structures obtained from the ONIOM method and those from the averaged MD simulations do not show any larger variations in RMSD (~1.3 Å). The binding affinity of **3AQN**

with the *c-MYC* G4 DNA is found to be higher than all the other ligand–DNA complexes (Table 2), which is in agreement with the results obtained from the MM-PBSA calculations.

Dynamic cross correlation maps

To explore the dynamic behavior of the nucleotides in the *c-MYC* G4 DNA–ligand complexes, the DCCMs of the atoms in the nucleotides were calculated from the 300 ns MD simulation trajectories. The motion of dG4 in the top quartet of the **3AQN** complex was found to be correlated with the motion of the loop-2 (dT11 and dA12), which is shown as red colored regions in Fig. 7A. Also, the motion of dG14 was found to be correlated with dG17, dG18, and dG19 in this complex. The correlated motions were also observed between loop-2 and the top-quartet (dG17) for the **6AQN** complex (Fig. 7B). For the **3APN** complex, loop-2 motion is correlated with the dG4 and dG8 (Fig. 7C). The **360A** complex showed correlated motions between loop-2 and the top quartet (dG8 and dG17) (Fig. 7D). Contrastingly, for the **Nap-Et** complex, anti-correlated motions are observed between loop-2 and the top quartet (dG4 and dG17), which are shown as the green colored regions in Fig. 7E. Also, correlated motions were observed between the 5′ and 3′-flanking nucleotides of G4 with the **Nap-Et** and **Nap-Pr** complexes (Fig. 7E and F). In the case of the **Nap-Pr** complex, anti-correlated motion was observed between loop-2 and dG4 (Fig. 7F). It should be noted

Table 1 Binding-free energy components of c-MYC G4 DNA–ligand complexes calculated from 300 ns of MD simulations. The molecular mechanical energy calculations were calculated using MM-PBSA and entropy calculations carried out using nmode in AMBER 14 at 298 K. All the values are reported in kcal mol⁻¹

| | c-MYC G4 DNA (PDB entry: 2L7V) | | | | | |
	3AQN	6AQN	3APN	360A	Nap-Et	Nap-Pr
ΔE_{ELEC}	-891.9 ± 17.1	-855.6 ± 21.1	-897.6 ± 13.7	-872.6 ± 16.4	-893.8 ± 23.1	-881.5 ± 20.1
ΔE_{vdW}	-65.7 ± 2.6	-62.6 ± 2.9	-57.8 ± 2.7	-60.0 ± 3.0	-62.0 ± 3.9	-61.4 ± 3.1
$\Delta E_{MM}(\Delta E_{ELEC} + \Delta E_{vdW})$	-957.6 ± 18.2	-918.3 ± 22.4	-949.3 ± 14.5	-932.7 ± 18.2	-955.9 ± 25.3	-943.0 ± 24.2
ΔPB_{NP}	-6.3 ± 0.2	-6.41 ± 0.2	-5.7 ± 0.2	-6.2 ± 0.2	-6.5 ± 0.2	-6.4 ± 0.2
ΔPB_{CAL}	903.1 ± 17.0	871.9 ± 21.6	900.3 ± 13.4	881.8 ± 16.4	912.3 ± 21.9	905.8 ± 19.6
$\Delta PB_{SOLV}(\Delta PB_{NP} + \Delta PB_{CAL})$	896.8 ± 16.9	856.9 ± 20.3	894.6 ± 13.3	875.5 ± 16.2	905.8 ± 21.8	899.4 ± 21.2
$\Delta H_{PB}(\Delta E_{MM} + \Delta PB_{SOLV})$	-60.9 ± 2.0	-61.4 ± 2.3	-54.8 ± 2.6	-57.1 ± 2.1	-50.1 ± 3.1	-43.6 ± 2.1
$T\Delta S$	-12.3 ± 1.6	-14.1 ± 1.7	-14.0 ± 1.1	-15.1 ± 1.3	-15.6 ± 1.4	-14.1 ± 1.1
$\Delta G_{bind}(\Delta H_{PB} - T\Delta S)$	-48.5 ± 2.1	-47.3 ± 2.3	-40.7 ± 2.2	-42.0 ± 2.1	-34.5 ± 2.8	-30.4 ± 2.6

Table 2 Interaction energies of ligand – c-MYC G4 DNA complexes obtained using the ONIOM method in the gas and aqueous phase at 298 K. All the values are reported in kcal mol⁻¹

	Gas phase		Aqueous phase	
Ligand	Complex	Interacting region/QM	Complex	Interacting region/QM
3AQN	-15.6	-10.7	-9.9	-7.5
6AQN	-13.2	-8.1	-9.0	-7.1
3APN	-10.1	-7.6	-8.0	-5.1
360A	-10.5	-7.6	-7.7	-4.9
Nap-Et	-8.3	-6.2	-5.7	-4.4
Nap-Pr	-7.6	-6.1	-5.6	-4.3

that no such correlated motions were observed for the ligand-free G4 DNA (Fig. S6 of the ESI†). Overall, it is evident from these results that the conformational dynamics of loop-2 can be induced by the ligand binding either on the top quartet or at the groove. Thus, engineering-specific ligand interactions with the loop-2 of c-MYC G4 can be the efficient way to design target-specific therapeutic agents targeting c-MYC G4 DNA.

Cluster analysis and ONIOM calculations

To target loop-2 of G4 DNA, it is necessary to identify the conformational dynamics of the loops during the course of simulations. Thus, cluster analysis was used to identify the major conformational ensembles of loop-2, and for this MD simulation trajectories were used. Each G4–ligand complex contributed 150 000 structures, from which loop-2 conformers were analyzed using the RMSD cutoff of 0.5 Å, and the structures with fewer occupancies were discarded. Also, the cluster analysis of the ligand-free c-MYC G4 DNA was carried out to compare the conformational changes upon ligand binding. The results showed that there are six major conformers adopted by loop-2 (Fig. 8). In the first ensemble, the exocyclic amine group of dA12 makes electrostatic contact with the non-bridging oxygen atom in the phosphate backbone (dG15) (Fig. 8A). Also, the exocyclic amine group of dG10 forms a H-bond with the nitrogen atom in the ring (dA12). To gain further insights, the percentage occupancies of H-bonds and electrostatic interactions during the course MD simulations were calculated. For the top quartet binding ligands, these interactions were found to be present ∼35% of the simulation time, which were found to be only ∼12% in the case of the groove binding ligands.

In the second conformational ensemble, the acceptor nitrogen atoms in the adenine ring make a H-bond interaction with the exocyclic amine of the dG9 and dG10 as shown in Fig. 8B. The third conformational ensemble represented in Fig. 8C reveals that nucleobase dT11 in loop-2 projects toward the G-quartet and stacks well on dA12. This stacking interaction in turn facilitates the stabilization of the electrostatic interaction between amine in the adenine (dA12), and the negatively charged phosphate backbone of dG15. In the case of the fourth conformational ensemble (Fig. 8D), the amine in the adenine (dA12) and the negatively charged phosphate backbone of dT11 make electrostatic contacts. This results in the formation of a ring-like loop conformation during MD simulations (Fig. 8D). In the fifth conformational ensemble (Fig. 8E), the dT11 and dA12 of loop-2 project outward from the G-quartet and stack well on each other. The sixth conformational ensemble (Fig. 8F) in which loop-2 is directed toward the G-quartet, and dA12 forms H-bonds with dG9 and dG10. In addition, dT11 stacks on dA12.

The percentage occupancies of these conformers indicate that all the ligands studied can induce these loop conformers (Fig. S7 of the ESI†). Also, aside from the conformer shown in 8D, none of the other conformers was observed for the ligand-free c-MYC G4 DNA structure (Fig. S7 of the ESI†). These results suggest that these ligand-induced conformations of the G4 loops can be used for virtual screening, which may help to identify specific ligands, targeting the c-MYC G4 structure.

The six loop-2 conformers that emerged from the cluster analysis (Fig. 8A–F) were evaluated for deciphering their stabilities using DFT (M06-2X-dzvp) and molecular mechanics (parmbsc1 force field) as implemented in the ONIOM method (Table 3). The results indicate that the energy-optimized geometries obtained from these calculations and those from the MD simulations are only slightly deviated (RMSD ∼ 1.1 Å). The higher stability of the conformer shown in Fig. 8C can be attributed to the presence of three non-covalent interactions. Overall, the results from the ONIOM calculations support the presence of loop conformations observed in the cluster analysis.

Conclusions

Even though a plethora of G4 stabilizing ligands are reported in the literature, there are only very few ligands that show
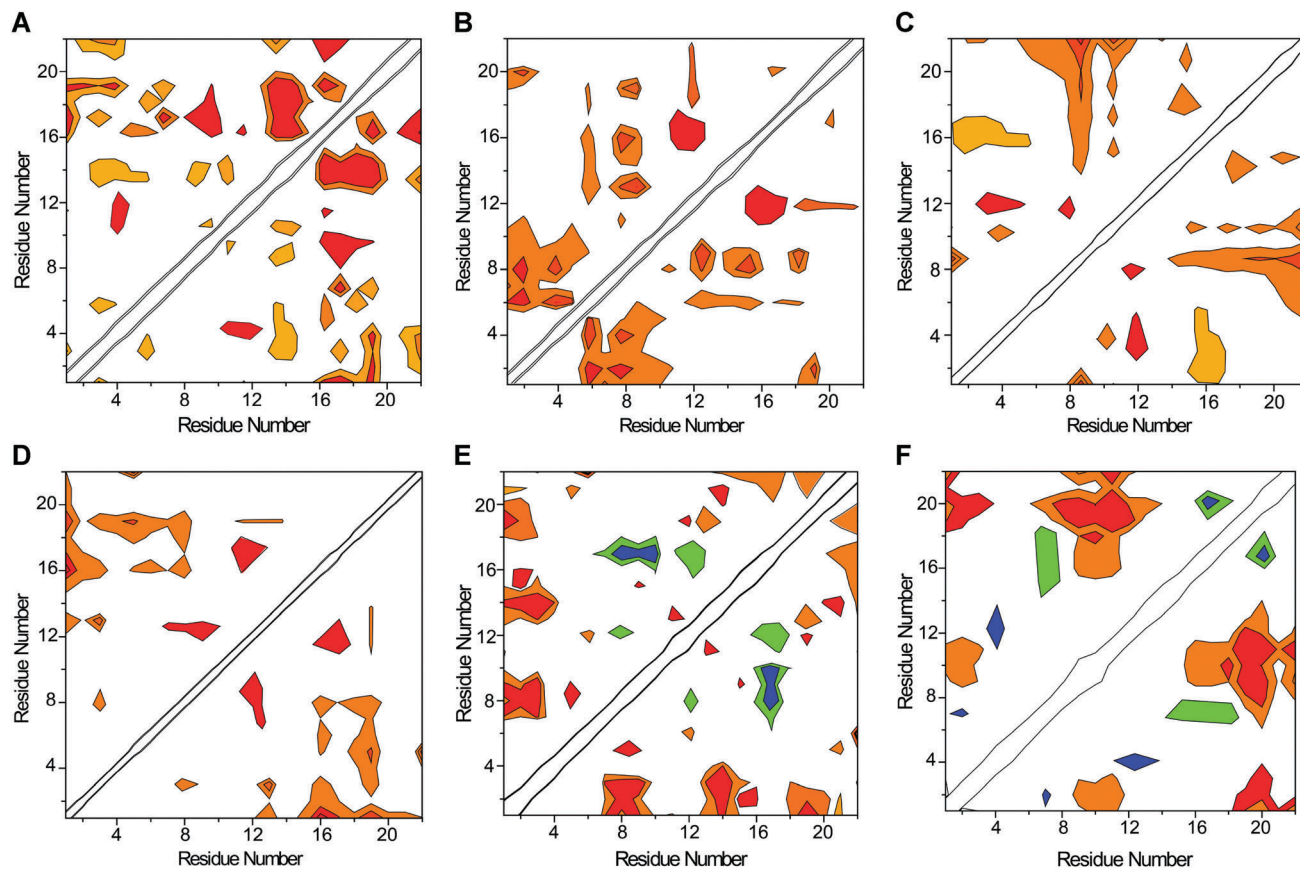
Fig. 7 Dynamic cross-correlation map of *c-MYC* G4 DNA during 300 ns of the MD simulations. DCCMs of *c-MYC* G4 DNA in complex with (A) **3AQN**, (B) **6AQN**, (C) **3APN**, (D) **360A**, (E) **Nap-Et**, and (F) **Nap-Pr**. Correlations between 0.75 and 0.98 and anticorrelations between −0.98 and −0.75 were considered to plot the graphs. Red (0.80 to 0.98); orange (0.78 to 0.88); cyan (0.78 to 0.75) blue (−0.98 to −0.88); green (−0.88 to −0.75). The correlated motion between the G-quartet is not shown for clarity.
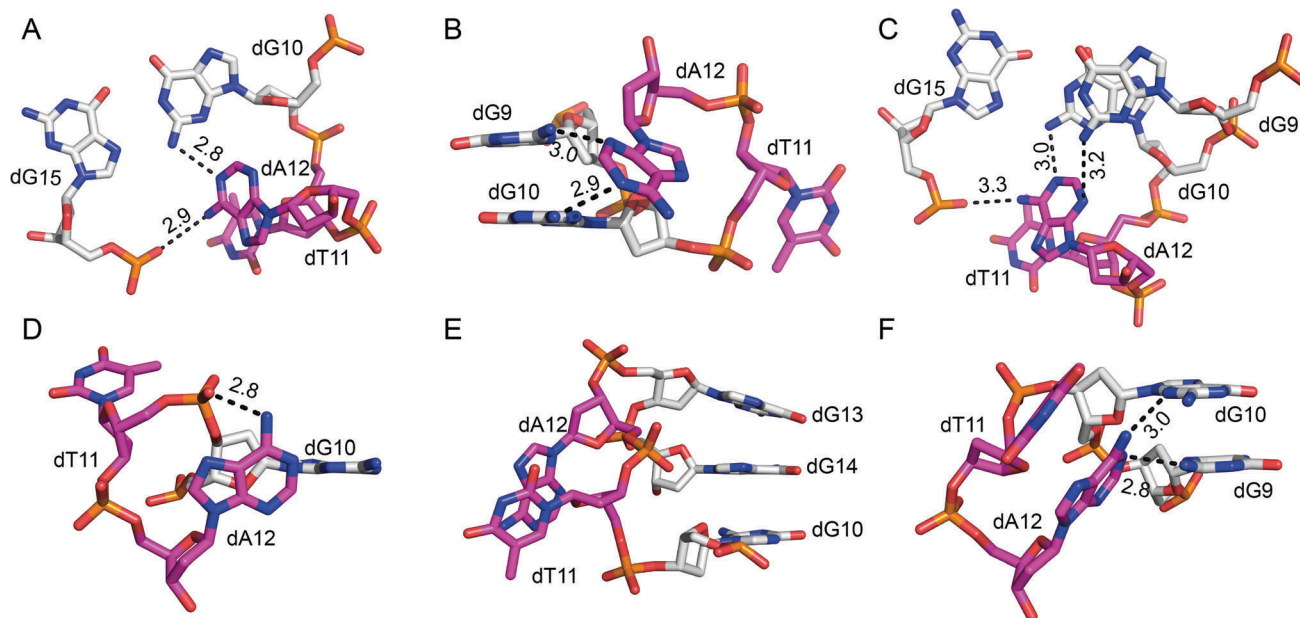
Fig. 8 Representative structures of the six major conformational ensembles (A–F) of loop-2 of *c-MYC* G4 DNA identified from the MD simulation of G4–ligand complexes.

Table 3 Relative energies of loop-2 conformers of *c-MYC* G4 DNA (Fig. 8A–F) obtained from MD simulation cluster analysis with respect to the ligand-free NMR structure using the ONIOM method. All the values are reported in kcal mol^{-1}

Conformer	Gas phase		Aqueous phase	
	Loop-2	DNA	Loop-2	DNA
A	−13.52	−21.27	−9.25	−17.61
B	−12.86	−20.95	−8.68	−16.64
C	−16.54	−24.83	−11.93	−14.84
D	−7.21	−14.22	−5.21	−12.57
E	−8.25	−18.35	−6.71	−15.62
F	−7.81	−12.07	−5.10	−11.75

specificity toward promoter G4 DNAs with parallel topologies.[29–31] It should be noted that none of these ligands is highly specific toward a particular promoter G4 structure such as *c-MYC*. To gain insights into the structure-based ligand design, MD simulations along with ONIOM calculations were performed on 6 recently reported G4 stabilizing ligands in complex with *c-MYC* G4 DNA. The MD simulations revealed that **3AQN**, **6AQN**, **3APN**, and **360A** bind at the top quartet, whereas **Nap-Et** and **Nap-Pr** bind at the groove of the G4. The results also show that loop-2 of the G4 structure adopts 6 different conformers upon ligand binding. There is a growing evidence that G4 structures undergo conformational changes over a wide range of timescales, and these structural fluctuations are important for the design of structure-specific ligands.[33,68,69] We propose that the loop conformers reported in the present study are the most accessible for binding interactions with the side chains of the ligands, and therefore, they can be harnessed for the design of novel ligands. However, a loop-specific ligand design along with a fragment-based ligand screening strategy may be required to discover the structure-specific G4 ligands.

Acknowledgements

References

1 S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res.*, 2006, **34**, 5402–5415.

2 A. N. Lane, J. B. Chaires, R. D. Gray and J. O. Trent, *Nucleic Acids Res.*, 2008, **36**, 5482–5515.

3 A. Guédin, J. Gros, P. Alberti and J.-L. Mergny, *Nucleic Acids Res.*, 2010, **38**, 7858–7868.

4 M. Marušič and J. Plavec, *Angew. Chem., Int. Ed.*, 2015, **54**, 11716–11719.

5 S. Neidle and G. Parkinson, *Nat. Rev. Drug Discovery*, 2002, **1**, 383–393.

6 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discovery*, 2011, **10**, 261–275.

7 J. Eddy and N. Maizels, *Nucleic Acids Res.*, 2008, **36**, 1321–1333.

8 G. W. Collie and G. N. Parkinson, *Chem. Soc. Rev.*, 2011, **40**, 5867–5892.

9 A. Ambrus, D. Chen, J. Dai, R. A. Jones and D. Yang, *Biochemistry*, 2005, **44**, 2048–2058.

10 S. Rankin, A. P. Reszka, J. Huppert, M. Zloh, G. N. Parkinson, A. K. Todd, S. Ladame, S. Balasubramanian and S. Neidle, *J. Am. Chem. Soc.*, 2005, **127**, 10584–10589.

11 A. T. Phan, V. Kuryavyi, S. Burge, S. Neidle and D. J. Patel, *J. Am. Chem. Soc.*, 2007, **129**, 4386–4392.

12 D. Sun, W.-J. Liu, K. Guo, J. J. Rusche, S. Ebbinghaus, V. Gokhale and L. H. Hurley, *Mol. Cancer Ther.*, 2008, **7**, 880–889.

13 S. Cogoi and L. E. Xodo, *Nucleic Acids Res.*, 2006, **34**, 2536–2549.

14 K. B. Marcu, S. A. Bossone and A. J. Patel, *Annu. Rev. Biochem.*, 1992, **61**, 809–858.

15 S. Pelengaris, B. Rudolph and T. Littlewood, *Curr. Opin. Genet. Dev.*, 2000, **10**, 100–105.

16 A. T. Phan, Y. S. Modi and D. J. Patel, *J. Am. Chem. Soc.*, 2004, **126**, 8710–8716.

17 A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 11593–11598.

18 J. Seenisamy, S. Bashyam, V. Gokhale, H. Vankayalapati, D. Sun, A. Siddiqui-Jain, N. Streiner, K. Shin-ya, E. White, W. D. Wilson and L. H. Hurley, *J. Am. Chem. Soc.*, 2005, **127**, 2944–2959.

19 D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. B. Ho, C. Proffitt, K. Trent, J. P. Whitten, J. K. C. Lim, D. Von Hoff, K. Anderes and W. G. Rice, *Cancer Res.*, 2009, **69**, 7653–7661.

20 T.-M. Ou, Y.-J. Lu, C. Zhang, Z.-S. Huang, X.-D. Wang, J.-H. Tan, Y. Chen, D.-L. Ma, K.-Y. Wong, J. C.-O. Tang, A. S.-C. Chan and L.-Q. Gu, *J. Med. Chem.*, 2007, **50**, 1465–1474.

21 A. Rangan, O. Y. Fedoroff and L. H. Hurley, *J. Biol. Chem.*, 2001, **276**, 4640–4646.

22 M. Trajkovski, E. Morel, F. Hamon, S. Bombard, M.-P. Teulade-Fichou and J. Plavec, *Chem. – Eur. J.*, 2015, **21**, 7798–7807.

23 J. Seenisamy, E. M. Rezler, T. J. Powell, D. Tye, V. Gokhale, C. S. Joshi, A. Siddiqui-Jain and L. H. Hurley, *J. Am. Chem. Soc.*, 2004, **126**, 8702–8709.

24 W. J. Chung, B. Heddi, F. Hamon, M.-P. Teulade-Fichou and A. T. Phan, *Angew. Chem., Int. Ed.*, 2014, **53**, 999–1002.

25 J. Dai, M. Carver, L. H. Hurley and D. Yang, *J. Am. Chem. Soc.*, 2011, **133**, 17673–17680.

26 H.-M. Lee, D. S.-H. Chan, F. Yang, H.-Y. Lam, S.-C. Yan, C.-M. Che, D.-L. Ma and C.-H. Leung, *Chem. Commun.*, 2010, **46**, 4680–4682.

27 D.-L. Ma, D. S.-H. Chan, W.-C. Fu, H.-Z. He, H. Yang, S.-C. Yan and C.-H. Leung, *PLoS One*, 2012, **7**, e43278.

28 H. R. Nasiri, N. M. Bell, K. I. E. McLuckie, J. Husby, C. Abell, S. Neidle and S. Balasubramanian, *Chem. Commun.*, 2014, **50**, 1704–1707.

29 K. V. Diveshkumar, S. Sakrikar, S. Harikrishna, V. Dhamodharan and P. I. Pradeepkumar, *ChemMedChem*, 2014, **9**, 2754–2765.

30 K. V. Diveshkumar, S. Sakrikar, F. Rosu, S. Harikrishna, V. Gabelica and P. I. Pradeepkumar, *Biochemistry*, 2016, **55**, 3571–3585.

31 V. Dhamodharan, S. Harikrishna, A. C. Bhasikuttan and P. I. Pradeepkumar, *ACS Chem. Biol.*, 2015, **10**, 821–833.

32 N. H. Campbell, M. Patel, A. B. Tofa, R. Ghosh, G. N. Parkinson and S. Neidle, *Biochemistry*, 2009, **48**, 1675–1680.

33 G. W. Collie, N. H. Campbell and S. Neidle, *Nucleic Acids Res.*, 2015, 4785–4799.

34 M. Debnath, S. Ghosh, D. Panda, I. Bessi, H. Schwalbe, K. Bhattacharyya and J. Dash, *Chem. Sci.*, 2016, **7**, 3279–3285.

35 K. G. Moghaddam and S. M. Hashemianzadeh, *RSC Adv.*, 2015, **5**, 76642–76650.

36 B. Islam, P. Stadlbauer, M. Krepl, J. Koca, S. Neidle, S. Haider and J. Sponer, *Nucleic Acids Res.*, 2015, **43**, 8673–8693.

37 A. Arora and S. Maiti, *J. Phys. Chem. B*, 2008, **112**, 8151–8159.

38 A.-J. Guan, E.-X. Zhang, J.-F. Xiang, Q. Li, Q.-F. Yang, L. Li, Y.-L. Tang and M.-X. Wang, *J. Phys. Chem. B*, 2011, **115**, 12584–12590.

39 V. Dhamodharan, S. Harikrishna, C. Jagadeeswaran, K. Halder and P. I. Pradeepkumar, *J. Org. Chem.*, 2012, **77**, 229–242.

40 C. Granotier, G. Pennarun, L. Riou, F. Hoffschir, L. R. Gauthier, A. De Cian, D. Gomez, E. Mandine, J.-F. Riou, J.-L. Mergny, P. Mailliet, B. Dutrillaux and F. D. Boussin, *Nucleic Acids Res.*, 2005, **33**, 4182–4190.

41 S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, *THEOCHEM*, 1999, **461–462**, 1–21.

42 D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, S. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER 14*, University of Califorai, San Francisco, 2014.

43 G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785–2791.

44 R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**, 198–210.

45 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Chem. Phys.*, 1993, **97**, 10269–10280.

46 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

47 M. Krepl, M. Zgarbová, P. Stadlbauer, M. Otyepka, P. Banáš, J. Koča, T. E. Cheatham, P. Jurečka and J. Šponer, *J. Chem. Theory Comput.*, 2012, **8**, 2506–2520.

48 I. Ivani, P. D. Dans, A. Noy, A. Perez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goni, A. Balaceanu, G. Portella, F. Battistini, J. L. Gelpi, C. Gonzalez, M. Vendruscolo, C. A. Laughton, S. A. Harris, D. A. Case and M. Orozco, *Nat. Methods*, 2016, **13**, 55–58.

49 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.

50 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2013, **9**, 3878–3888.

51 S. Le Grand, A. W. Götz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374–380.

52 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

53 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.

54 B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon and L. S. D. Caves, *Bioinformatics*, 2006, **22**, 2695–2696.

55 T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery, K. Morokuma and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 815–826.

56 Y. Zhao and D. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

57 P. Császár and P. Pulay, *J. Mol. Struct.*, 1984, **114**, 31–34.

58 K. Gkionis, H. Kruse and J. Šponer, *J. Chem. Theory Comput.*, 2016, **12**, 2000–2016.

59 T. Vreven, K. Morokuma, Ö. Farkas, H. B. Schlegel and M. J. Frisch, *J. Comput. Chem.*, 2003, **24**, 760–769.

60 E. Fadrná, N. A. Špačková, R. Štefl, J. Koča, T. E. Cheatham and J. Šponer, *Biophys. J.*, 2004, **87**, 227–242.

61 S. Cosconati, L. Marinelli, R. Trotta, A. Virno, S. De Tito, R. Romagnoli, B. Pagano, V. Limongelli, C. Giancola, P. G. Baraldi, L. Mayol, E. Novellino and A. Randazzo, *J. Am. Chem. Soc.*, 2010, **132**, 6425–6433.

62 F. S. Di Leva, P. Zizza, C. Cingolani, C. D'Angelo, B. Pagano, J. Amato, E. Salvati, C. Sissi, O. Pinato, L. Marinelli, A. Cavalli, S. Cosconati, E. Novellino, A. Randazzo and A. Biroccio, *J. Med. Chem.*, 2013, **56**, 9646–9654.

63 S. Cosconati, L. Marinelli, R. Trotta, A. Virno, L. Mayol, E. Novellino, A. J. Olson and A. Randazzo, *J. Am. Chem. Soc.*, 2009, **131**, 16336–16337.

64 A. Spinello, G. Barone and J. Grunenberg, *Phys. Chem. Chem. Phys.*, 2016, **18**, 2871–2877.

65 S. M. Haider, I. Autiero and S. Neidle, *Biochimie*, 2011, **93**, 1275–1729.

66 G. Prato, S. Silvent, S. Saka, M. Lamberto and D. Kosenkov, *J. Phys. Chem. B*, 2015, **119**, 3335–3347.

67 J. Šponer, A. Mládek, N. Špačková, X. Cang, T. E. Cheatham and S. Grimme, *J. Am. Chem. Soc.*, 2013, **135**, 9785–9796.

68 R. Rodriguez, G. D. Pantoş, D. P. N. Gonçalves, J. K. M. Sanders and S. Balasubramanian, *Angew. Chem.*, 2007, **119**, 5501–5503.

69 A. Marchand, A. Granzhan, K. Iida, Y. Tsushima, Y. Ma, K. Nagasawa, M.-P. Teulade-Fichou and V. Gabelica, *J. Am. Chem. Soc.*, 2015, **137**, 750–756.

Supporting Information

Ligand Induced Conformational Preorganization of Loops of *c-MYC* G-Quadruplex DNA and its Implications in Structure Specific Drug Design
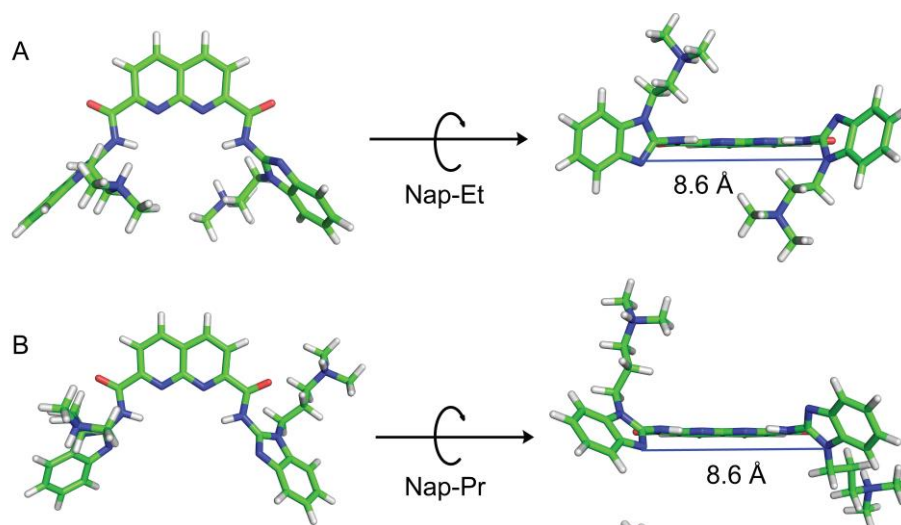
S. Harikrishna*, Saikiran Kotaru and P. I. Pradeepkumar*

Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400076

*Email: harikrishna.s@iitb.ac.in or pradeep@chem.iitb.ac.in

TABLE OF CONTENTS

**Energy optimized structures of ligands at HF/6-31G* level**

Figure S1. Energy optimized structures of ligands using HF/6-31G* theory level in Gaussian 09. (A) **Nap-Et** and (B) **Nap-Pr** optimized structures. Atoms are shown in stick representation. The solid blue lines between two benzimidazole rings specify the distance between benzimidazole side chains.

Stacking interaction between ligands and top-quartet of the G4
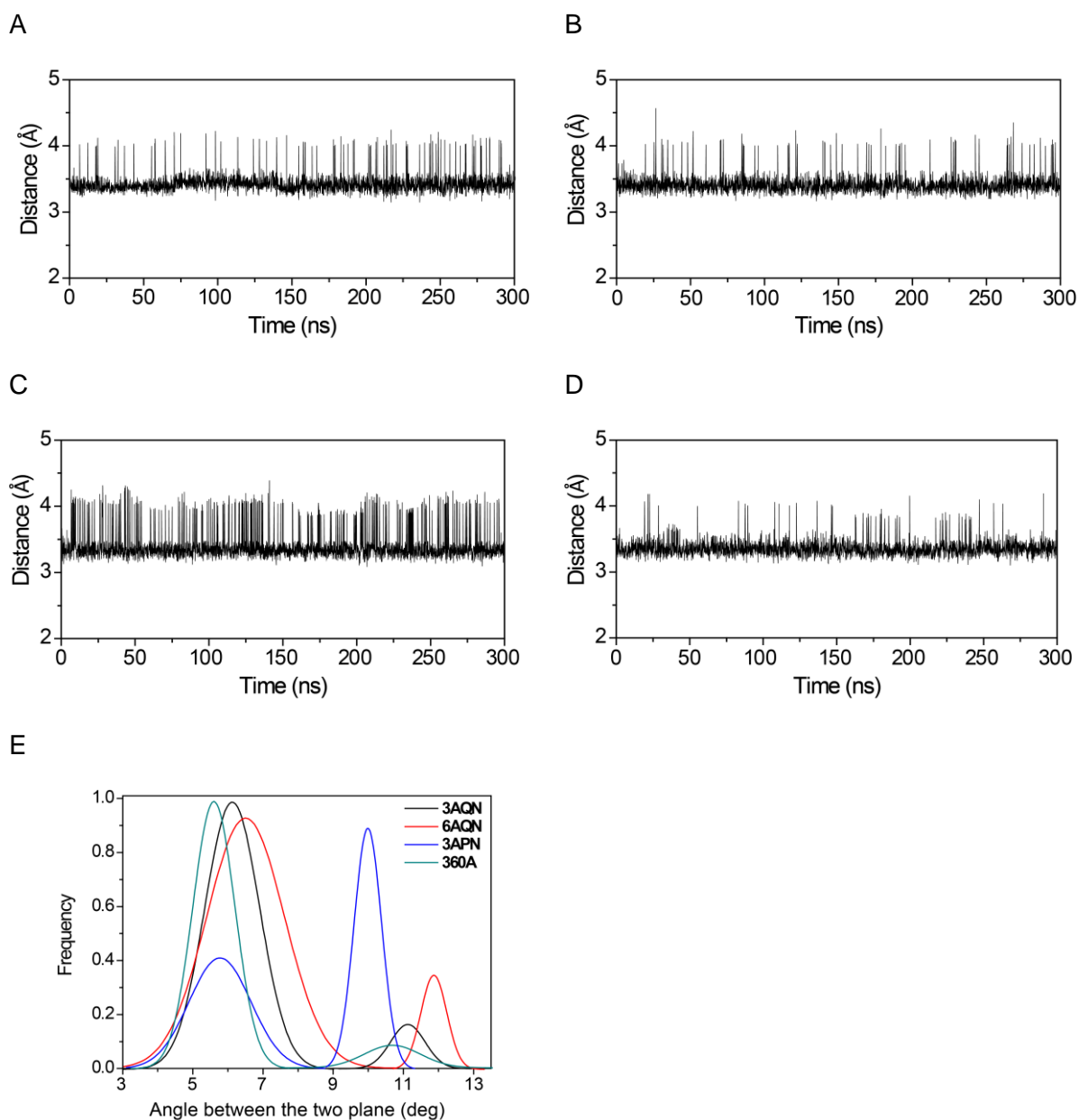
Figure S2. Stacking distance and angle between the plane of the aromatic moiety in the ligand and plane of the G-quartet during the course of MD simulations. Distance between the plane of (A) 3AQN and top quartet, (B) 6AQN and top quartet, (C) 3APN and top quartet (D) 360A and top quartet and (E) angle between the plane of ligand and quartet. These calculations were performed using PLUMED plugin in the UCSF Chimera.
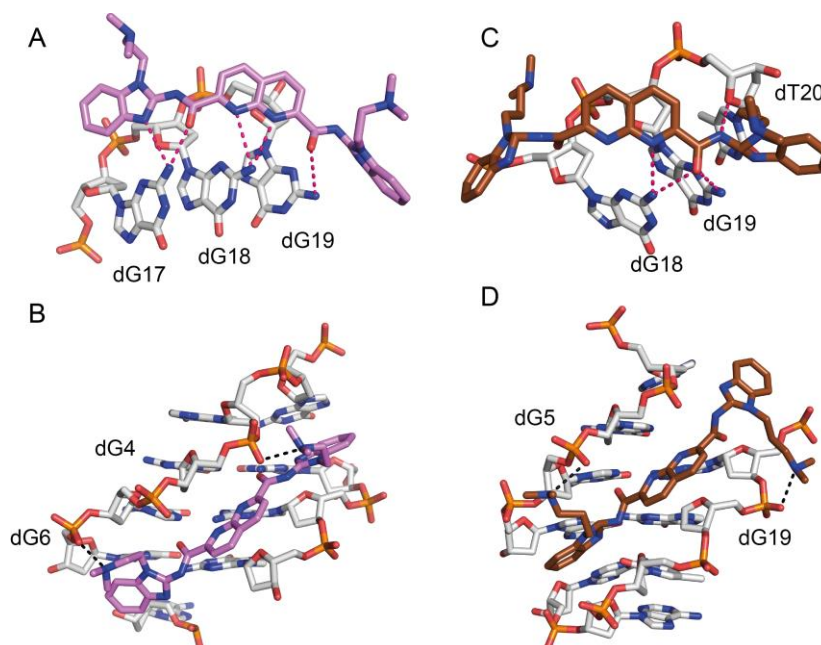
## Non-covalent interactions between 5'-flanking nucleotides and ligands
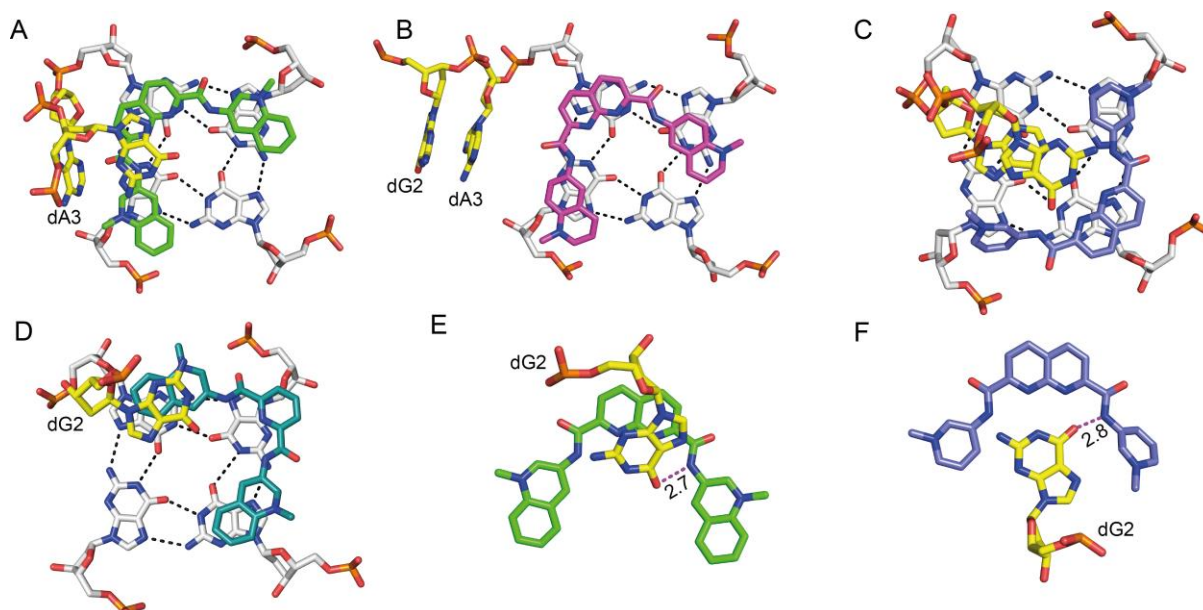
Figure S3. Non-covalent interactions between top quartet binding ligands and the 5'-flanking nucleotides. (A) The dG2 in the 5'-flanking nucleotide stacks on the **3AQN**. (B) Both dG2 and dA3 in the 5'-flanking nucleotides flipped out of the G-quartet surface and are not stacking on the **6AQN** (C) The 5'-flanking nucleotide dA2 stacks on the G-quartet and the dA3 stacks on the dG2 nucleotide in the **3APN** complex. (D) The dG2 nucleotide stacks on the **360A**. (E) The O^6 in the dG2 nucleotide make H-bond interaction with the NH group of the side chain in **3AQN**. (F) The O^6 in the dG2 nucleotide make H-bond interaction with the NH group of the side chain in **3APN**. All the ligands shown here stack on the top quartet of the G4 DNA. All the distances were mentioned in Å.

**Non-covalent interactions between DNA and groove binding ligands**

Figure S4. Non-covalent interactions between the G4 DNA and the groove binding ligands. (A) H-bond interactions between guanines including dG17, dG18 and dG19 in the G-quartet and **Nap-Et**, the distances of these H-bonds were between 2.7 and 3.1 Å, and the occupancies of these H-bonds are found to be >65 % of the total simulation time. (B) Electrostatic interactions between the positively charged side chain in **Nap-Et**, and the negatively charged phosphate backbone of dG4 and dG6 in the G-quartet. The distances of the two electrostatic contacts were between 2.5 and 3.3 Å. (C) H-bond interactions between guanines including dG18 and dG19 in the G-quartet, dT20 in the 3'-flanking nucleotide and **Nap-Pr**. The distances of these H-bonds were between 2.8 and 3.1 Å, and the occupancy of these H-bonds are found to be >62 % of the total simulation time. (D) Electrostatic interactions between the positively charged side chain in **Nap-Pr** and the negatively charged phosphate backbone of dG5 and dG19 in the G-quartet. The distances of the two electrostatic interactions are between 2.7 and 3.4 Å.

Dynamic cross-correlation map of ligand free *c-MYC* G4 DNA
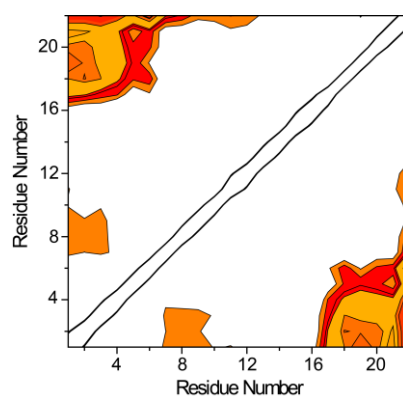
Figure S5. Dynamic cross-correlation map (DCCM) of ligand free *c-MYC* G4 DNA during 300 ns of MD simulations. Correlation between 0.75 to 0.98 and anticorrelation between −0.98 to −0.75 were considered to plot the graph. Red (0.80 to 0.98); Orange (0.78 to 0.88). The correlated motion between the G-quartet were discarded for clarity.

Percentage occupancies of the clusters of loop-2 conformers from MD simulations
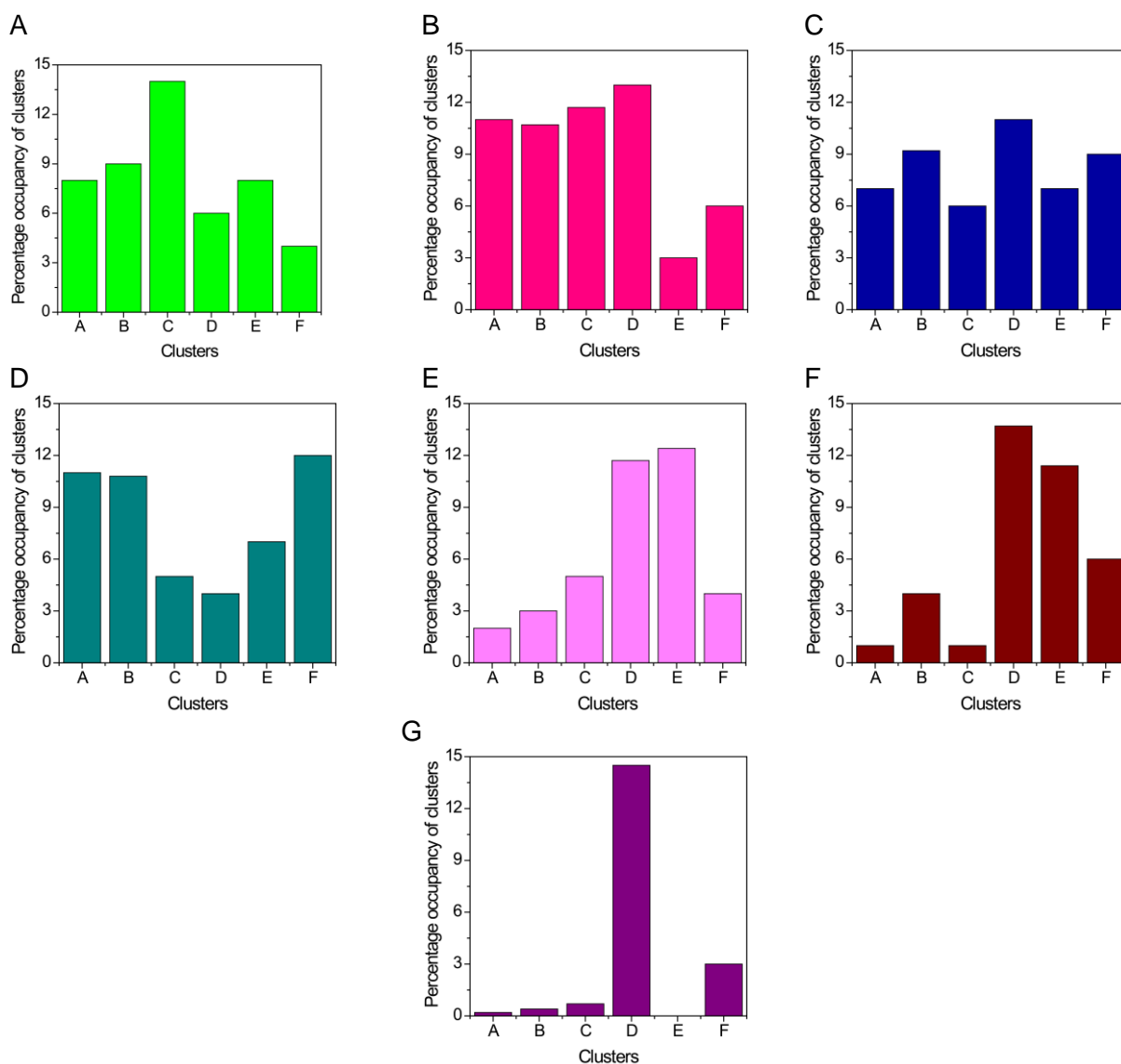
Figure S6. Percentage occupancies of the clusters (loop2) from the 300 ns MD simulations trajectories. Six conformational ensembles identified from the cluster analysis of the MD simulations in complex with six different ligands including (A) **3AQN**, (B) **6AQN**, (C) **3APN**, (D) **Nap-Et** (E) **Nap-Pr** and the (F) ligand free G4 DNA. The representative structures of the conformers A-F were shown in Figure 8 (Main text).

SASA values of the *c-MYC* G4 DNA in complex with ligands

c-MYC G4 DNA -Ligand	SASA (Å^2)	ΔSASA (Å^2)
3AQN	3671	397
6AQN	3715	441
3APN	3607	333
360A	3678	404
Nap-Et	3342	68
Nap-Pr	3346	72

Table S1. The solvent accessible surface area values of *c-MYC* G4 DNA in complex with ligands used. The SASA of native *c-MYC* G4 DNA is 3274 Å^2. The ΔSASA is calculated as the difference between the values of native *c-MYC* G4 DNA and ligand bound *c-MYC* G4 DNA complexes after 300 ns of MD simulations. SASA values were calculated using SURF tool in AMBER 14.