

G-Quadruplex Recognition by DARPIs through Epitope/Paratope Analogy**

Tom Miclot,^{*,[a, b]} Emmanuelle Bignon,^[b] Alessio Terenzi,^[a] Stéphanie Grandemange,^[c] Giampaolo Barone,^{*,[a]} and Antonio Monari^{*,[b, d]}

Abstract: We investigated the mechanisms leading to the specific recognition of Guanine Quadruplex (G4) by DARPIs peptides, which can lead to the design of G4 s specific sensors. To this end we carried out all-atom molecular dynamic simulations to unravel the interactions between specific nucleic acids, including human-telomeric (h-telo), Bcl-2, and *c-Myc*, with different peptides, forming a DARPin/G4 complex. By comparing the sequences of DARPin with that of a peptide known for its high affinity for *c-Myc*, we show that

the recognition cannot be ascribed to sequence similarity but, instead, depends on the complementarity between the three-dimensional arrangement of the molecular fragments involved: the α -helix/loops domain of DARPin and the G4 backbone. Our results reveal that DARPIs tertiary structure presents a charged hollow region in which G4 can be hosted, thus the more complementary the structural shapes, the more stable the interaction.

Introduction

In addition to the well-known double helical arrangement, the important biological role of non-canonical nucleic acid is nowadays widely recognized. Among the different non-canonical DNA or RNA structures, guanine quadruplexes (G4 s) are highly studied and characterized.^[1–3] From a chemical point of view, G4 s are formed in guanine-rich nucleic acids, whose nucleobases develop primarily cooperative Hoogsteen-type hydrogen bonds. Because of the specificity of this interaction, DNA (or RNA) is then organized in a series of stacked quartets, whose macromolecular arrangement is further stabilized by a

metal cation occupying the central channel. Three main topologies can be adopted by non-canonical G4 structures depending on the 5'-3' orientation of the strands forming the G4 s backbone: parallel, antiparallel and hybrid (Figure 1). In the parallel conformation, all the strands are oriented in the same way, in the antiparallel conformation the strands are inversely oriented two by two, while in the hybrid conformation only one strand is inversely oriented with respect to the other three.^[4]

G4 s have been identified in cellular nucleic acids and have been associated to the control of key biological functions. As a matter of fact, G4 s are involved in gene regulations,^[5,6] in

[a] T. Miclot, Prof. A. Terenzi, Prof. G. Barone
Department of Biological, Chemical and Pharmaceutical Sciences and Technologies
Università degli Studi di Palermo
Viale delle Scienze, 90128 Palermo (Italy)
E-mail: tom.miclot@unipa.it
giampaolo.barone@unipa.it

[b] T. Miclot, Dr. E. Bignon, Prof. A. Monari
Université de Lorraine and CNRS LPCT UMR 7019
54000 Nancy (France)

[c] Prof. S. Grandemange
Université de Lorraine and CNRS CRAN UMR 7039
54000 Nancy (France)

[d] Prof. A. Monari
Université Paris Cité and CNRS, ITODYS
75006 Paris (France)
E-mail: antonio.monari@u-paris.fr

[**] A previous version of this manuscript has been deposited on a preprint server (<https://doi.org/10.1101/2022.06.13.495947>).

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/chem.202201824>

© 2022 The Authors. Chemistry - A European Journal published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

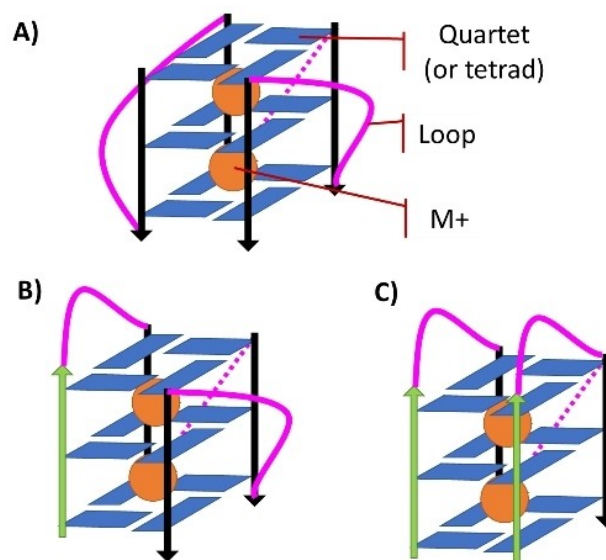


Figure 1. Schematic representation of the different topologies of G4. A) The parallel form in which all strands are parallel to each other, B) the hybrid form in which one strand is oriented antiparallel to all the others, and C) the antiparallel form in which the strands are antiparallel two by two.

neurodegenerative diseases,^[7,8] in the induction of DNA damages and in oncogenesis.^[9–12] Furthermore, they have been recognized to play a role in the regulation of cellular cycles and in the regulation of post-translational modification in proteins.^[12–15] G4 s have also been identified in both DNA and RNA viral genomes, including SARS-CoV-2,^[16–18] where they may exert vital functions in regulating viral infection cycles.^[19–22] Obviously, all these processes can only take place through a molecular machinery involving proteins selectively recognizing specific DNA or RNA G4 s. Among them we may cite ATP-dependent DNA/RNA helicase DHX36,^[23] G-rich sequence factor 1,^[24] or fragile X mental retardation protein (FMRP).^[25] Thus, the development of artificial or biomimetic specific G4 binders, capable to recognize either DNA or RNA, is highly valuable. Furthermore, such ligands may be exploited either in a therapeutic context or for the rapid identification and localization of G4 s in cells or cellular compartments.^[26–29] Protein engineering has also led to the development of antibodies presenting specificity and selectivity toward G4 s. In this case, the recognition of G4 s proceeds through the epitope/paratope mechanism, in which the G4 acts as the epitope of an antibody.^[30] However, the design of antibodies is definitively not straightforward, and their use is typically limited to the identification of the subcellular localization of G4 s.^[30–33] Smaller peptides specifically recognizing G4 s and even discriminating between different G4 types have also been proposed. This is typically the case of DARPins,^[34] a class of synthetic proteins derived from the modification of natural ankyrins and mostly known as chaperone agents in crystallography.^[35] In addition, DARPins have also been used as cellular markers in biological imaging and for therapeutic purposes.^[36,37]

Understanding the factors underlying the specific recognition of G4 s by DARPins can facilitate the design of sensors able to discriminate the G4 s subcellular localization and their specific sequences. In this contribution, we model the DARPIn/G4 interaction and, thus, unravel the specific recognition modes by combining molecular docking and long-scale all atom molecular dynamics (MD) simulations. We focus on 2E4 DARPIn, which is specific for the G4 present in the *c-Myc* oncogene promoter, and 2G10, which has a slight specificity for different G4 s.^[34] As for the nucleic acid counterpart, we restrict our study to the human telomeric G4, as well as the G4 s in the *c-Myc*^[34] and *Bcl-2* promoters.

Results and Discussion

MD simulations shed light on the structural details underlying the specific DARPins/G4 interaction. Indeed, by sampling the conformational space through different initial interaction positions, it is possible to analyze whether the interaction is conserved, the binding of the G4 affects the flexibility, the nucleic acid rearranges to reach a more stable pose, or if the proposed DARPIn/DNA complex is not stable and separates. In our case most of the G4 s/DARPIn complexes are persistent and stable all along the MD and the peptides interact with the G4 s through regions composed of large loops and helices, which

overlap well with the recognized canonical interaction zones of the DARPins.

The only exceptions can be highlighted for 2E4/*c-Myc* which in one of the poses leads to a very labile and mobile interaction between G4 and the protein as confirmed by clustering yielding two dominant structures, representing 41.71 % and 30.56 % of the trajectory, respectively. On the other hand, 2E4/*h-Telo* (64.33 % of the trajectory) and 2G10/*c-Myc* (75.10 %) yield dominant clusters that interacts only through the loop ends of DARPins and one or two nucleotides of the flexible G4 s loops. (All the clustered structures can be found in the Supporting Information).

Residue-scale analysis of the G4/DARPIn complexes

Before exposing the structural details of the G4 s/DARPIn complexes at the atomistic scale, it is interesting to consider the interaction at a residue-level scale, and in particular classify the different interaction patterns in terms of the number of involved nucleic acid or protein residues. Figure 2 shows all the residues which remains within a cutoff of 3 Å from either the protein or the nucleic acid with a frequency at least equal to 50 % of the simulation time. On average six nucleic acid residues of *h-Telo* and eight amino acids of 2G10 can be identified. However, 2G10 interacts persistently through only five amino acids with *c-Myc* and *Bcl-2* which in turn only bring a maximum of two or three nucleotides into persistent contact with the protein. Conversely, the 2E4/*h-Telo* interaction appears to be driven by three nucleotides and six amino acids. For 2E4, the interaction gathers eight amino acids with both *c-Myc* and *Bcl-2*, yet a different number of nucleic acid residues is involved, that is, four for *Bcl-2* and seven for *c-Myc*.

This first analysis, at the residue level, already draws a general picture of the specific recognition of G4 by DARPins, confirming that the protein/nucleic acid recognition is favored by a high number of interacting residues. However, it needs to be completed identifying the exact nature of the interacting residues, their specific frequency, and the specific structural features.

Three different modes of interaction leading to the G4/DARPIn complexes

To improve the global scale analysis presented in the previous subsection we should identify a region of the G4 s that is selectively recognized by DARPins. As a matter of fact, *h-Telo* does not show any specific interacting region or hotspot with 2E4 or 2G10 (see Figure 3A). This could confirm that the non-specific recognition of *h-Telo* is due to the absence of a well-defined target region on this nucleotide. In contrast, more pronounced specific interaction regions may be recognized in the two others G4 s. Indeed, it can be seen in Figure 3(B) that two interaction areas clearly stand out for *Bcl-2* interacting with either 2EA or 2G10, that is, the one including residues dC5 to dG8 and the one involving residues dG20 to dG22. Since the

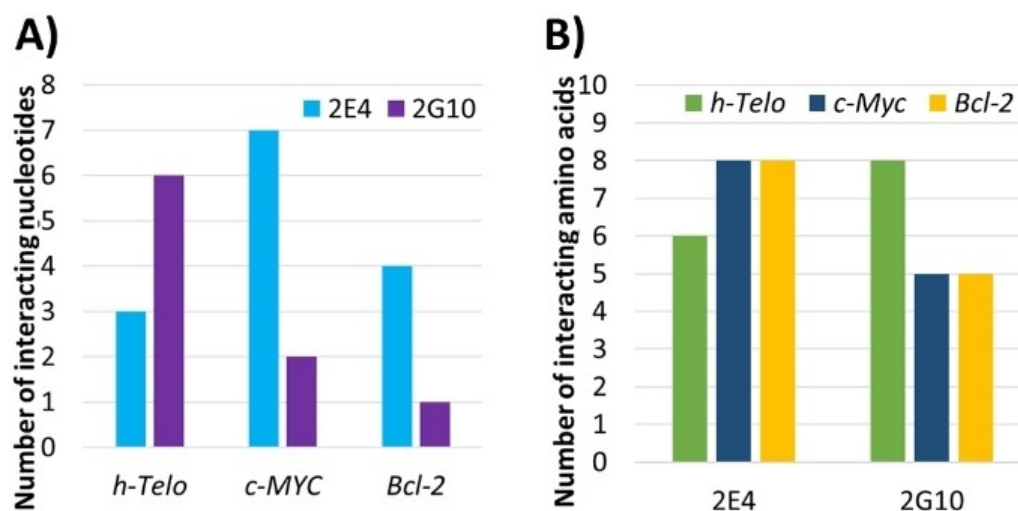


Figure 2. Number of A) nucleic acids and B) amino acids residues involved in the protein/DNA interaction and found at a frequency of at least 50% of the simulation time.

same nucleic acid regions are evidenced for both DARPins, the interaction mode can be classified as structurally similar in each complex. Thus, no specific recognition of *Bcl-2* by 2E4 or 2G10 can be inferred, since such specific recognition should involve interaction areas that must differ between two different DARPins. This is, indeed, the case for the G4 present in the *c-Myc* promoter. Figure 3(C) clearly shows regions of very pronounced contact and different for each of the DARPins. For the interaction with 2G10, the hot spot includes residues dG15 to dT20, although the contact frequencies are still quite spread across the whole G4. On the other hand, 2E4 highlights two very strong and localized contact points. The most important one concerns residues dG6 to dA12, while the second one corresponds to the last two residues of G4, dA21 and dA22. Thus, the specific recognition of *c-Myc* by 2E4 could be achieved either through the recognition of its sequence, or through a specific structural motif. Our analysis indicates three possible scenarios: 1) a rather general interaction that does not involve any specific G4 region or sequence (*h-telo*); 2) a non-specific interaction involving particular G4 s regions, which are however recognized by all the proteins (*Bcl-2*); 3) a specific interaction driven by a few nucleotides having very high contact frequencies with specific DARPin (*c-Myc*).

Identification of a putative selective DARPin interaction area

Repeating the same analysis while focusing on the protein counterpart we identify the amino acids mostly involved in the recognition of the non-canonical DNA structure. Similarly, to what has already been observed for G4 s, selectivity should correlate with few specific amino acids having high contact frequencies with G4 s. Conversely, for non-selective recognition a more scattered distribution of the interaction frequencies should be observed.

The distribution of the interaction contacts of 2G10 with the three G4 s (Figure 4A) shows three distinct peaks. The first one corresponds to residues N34 and I35, the second one gathers residues R67, W68, R70, K78 and W79, while the last one comprises residues K100 and K101. While these localized protein areas certainly correspond to a strong interaction with G4, they appear rather non-specific since they are present for all the three G4 s. However, the interaction with *h-Telo* is also driven by amino acids whose contact frequency was low or zero for the other G4S. This case concerns mainly residues H107, L108, I111, R112, K133, F134, K136, and I141. However, caution should be taken to avoid overinterpretation of this result, since 2G10 is not showing any specificity for *h-Telo*.^[34] The contact frequency for 2E4 (Figure 4B) shows the emergence of even more defined trends presenting stronger and more localized maxima. In particular, we can mention residues K5, E9, and R12 as well as the regions spanning residues R34, W35, and M46, and residues H67, W68 and R70. However, only a relatively small difference in the interaction patterns between the three G4 s can be highlighted. In particular in the case of *c-Myc* residues Y45, R70, L75, S78, R79, and G80 develop persistent contacts, and hence could be regarded as potential hot-spots for the selectivity of 2E4 towards this G4.

Although, the 2G10/G4 complex involves a larger number of residues developing more persistent contacts than 2EG/G4, this should not be necessarily correlated to a higher affinity towards G4. Indeed, few residues developing stable and persistent interactions may be regarded as more favorable than an extended weakly interacting region. Furthermore, 2G10 is larger than 2E4, thus the higher number of contacting residues may be also regarded as an obvious statistical effect.

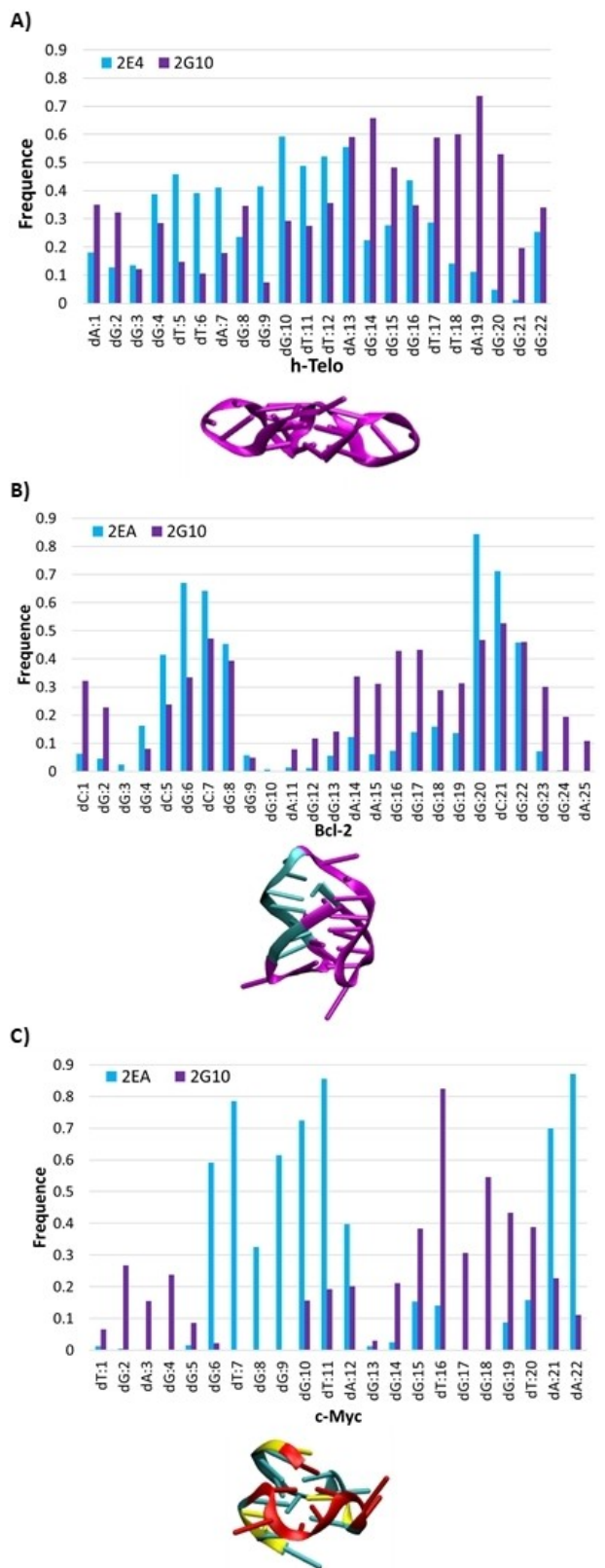


Figure 3. Frequency and location, in magenta, of nucleotide of A) *h-Telo*, B) *Bcl-2* and C) *c-Myc* involved in a 3 Å cutoff from the protein. Note that the color code has been changed for *c-Myc*, because the mode of interaction between the two proteins involves different nucleotides. Thus, the nucleotides most frequently involved with 2E4 are in red, while those interacting preferentially with 2G10 are in yellow.

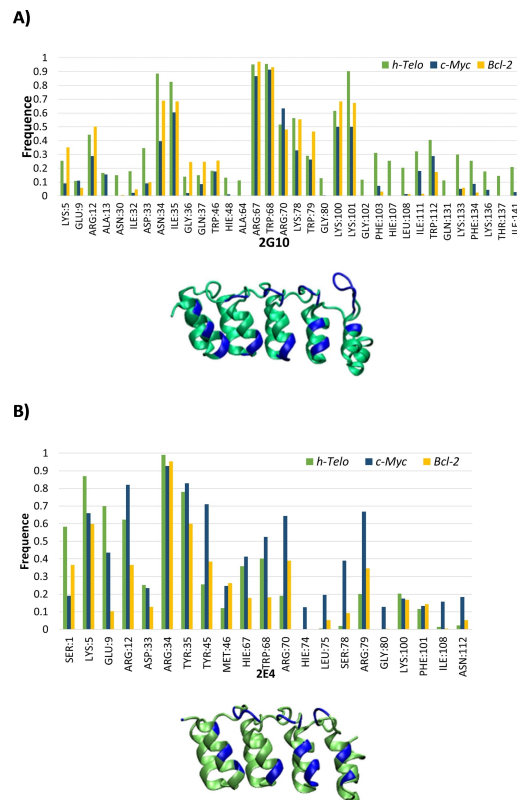


Figure 4. Frequency and location, in blue, of amino acids of A) 2G10 and of B) 2E4 involved in a 3 Å cutoff of G-quadruplex.

Alignment between 2E4 and a *c-Myc*-specific peptide reveals no sequence similarity

Several examples of peptides able to selectively bind G4s are reported in the literature. Usually, they are derived from the DHX36 helicase whose α -helix provides the binding interface with the nucleic acid.^[38,39] Notably, Minard et al.^[40] designed a specific peptide, DM102 (PGHLKGRRIWLWYASKQGQKNK), which is able to preferentially recognize G4 in the *c-Myc* promoter. Since 2E4 is also specific to *c-Myc*, it is legitimate to ask whether there is a sequence similarity between DM102 and 2E4. It is also important to note that the artificial peptide DM102 has a hydrocarbon staple ($i, i+7$) between residues R8 and S15, which enforces a defined α -helix. Hence, in addition to sequence similarity, one must also look for stable α -helical secondary structure, which is, indeed, a structural motif frequently present in DARPin. Two algorithms were used: Clustal Omega and M-coffee (see Figure S37). Here it is important to specify that Clustal Omega is an individual method of alignment, while M-coffee is an algorithm that combines results from several individual methods.^[41]

Interestingly, the two algorithms show divergent results. Clustal Omega previews similarity mainly concerning the 2E4 region spanning M46 to V62. The representation of the 3D structure of this region (see Figure S38) reveals a helical arrangement, which could partially support the 2E4 selectivity conditions. However, this region is also common to all DARPin,

except for residue 58 (residue 70 following Scholz et al. notation),^[34] which is embedded in the similarity region, and residues 45 and 46, which border it. Furthermore, the mutated residues at position 46 and 58 are structurally very distant, suggesting a low quality of the alignment. Conversely, M-Coffee alignment highlights three subunits. Two of them have no significance, the first being located at the previously invalidated region, and the third pertaining to the N-terminal region common to all DARPins. Instead, the second subunit is aligned with the R70-R79 region of DM102, as visually represented in Figure 5 by the transparent shaded area. This observation is also coherent with our MD simulations which indicate an increase of the DARPIn/*c-Myc* contact frequency for the residues belonging to this region. Furthermore, from a structural point of view, the R70-R79 region is organized in α -helix motif and is located towards the canonical recognition zone of DARPins. Yet, this region is highly conserved among the DARPins designed by Scholz et al.^[34] and only the residues bordering the helix, that is, R70, S78, and R79, have been mutated. Indeed, when M-Coffee alignment between DM102 and 2G10 the same DARPIn region is evidenced (see Figure S39).

Thus, the search for a conserved sequence between DM102 and 2E4, does not unambiguously justify the selectivity of the DARPIn. Going a step further, this could suggest that the recognition of *c-Myc*'s by 2E4 does not necessarily involve sequence similarity between DM102 and 2E4. This is also supported by the fact that the mutations of the wild type sequence as performed by Scholz et al.^[34] are mainly concentrated on the peripheral protein loops. Consequently, we decided to focus on structural features which should add up to the rather modest sequence effects and, ultimately, drive the selectivity.

2E4 recognizes a particular structural motif of *c-Myc*

From our MD simulations two most important factors should be considered when analyzing the local structural arrangements of the DARPIn/G4 contact region. First, the DARPIn canonical interaction zone is not consistently interacting throughout the whole MD simulation. Instead, as highlighted in Figure 4(B), other amino acids either located in α -helices or in peripheral loops develop more persistent interactions.

Furthermore, the analysis of the interaction networks shows that a DARPins/G4 complex is mainly stabilized by electrostatic

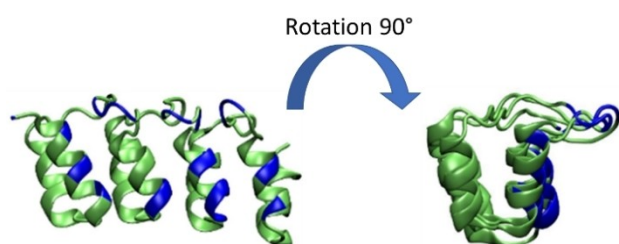


Figure 5. The preferential interaction region of the 2E4 protein is located at the α -helix and the loops and evidenced in blue.

interaction between positively charged amino acids and the negatively charged backbone of the nucleic acid. The paper by Scholz et al.^[35] clearly excludes any interaction between DARPins and canonical double strand DNA. This fact also confirms that the recognition of the nucleic acid should involve important and specific structural motifs, as confirmed by our study. In addition, π -cation interactions are also present mainly when the extended conjugated system of a tetrad faces the DARPIn. Even if this interaction appears persistent along the MD simulation it should be confirmed by using quantum chemistry-based modeling, or even hybrid quantum/classical approaches, to avoid any spurious force field artifact and precisely calculate energy interaction terms. However, such a study, even if highly interesting would be out of the scope of the present contribution. Finally, DARPIn associates with G4 through its canonical interaction zone involving the α -helix, but also via interactions mediated by the peripheral loops. The interaction with the loop is most pronounced, but not unique, in the case of *h-Telo*, which in the course of the MD simulation departs from the initial docking pose and slides over the DARPIn surface until an interaction between its quartet and the peripheral loops is established at around 150 ns (Figure 6). Interestingly, the electrostatic interactions involving the G4 backbone take place mainly through the G4 external loops rather than the tetrad core. However, this conformation appears as scarcely stable, and as a matter of fact the G4 oscillates and reverts to a more classic interaction mode involving one of its accessible quartets. These observations are also consistent with the frequency distribution reported in Figure 3 and explain the specific behavior of *h-Telo*, which due to its high mobility spans different interaction poses and develops rather non-specific contacts with a high number of 2E4 and 2G10 residues.

Thus, the interaction mode involving a quartet is not leading to a specific recognition mode. Hence, interactions between *c-Myc* or *Bcl-2* and DARPIn which would be driven by the G4 s quartets (Figure 7) will most probably be trapped in a non-specific recognition and cannot be used to infer on the specific recognition. On the contrary, specificity may be established when DARPins interact mainly with the nucleic acid backbone. The behavior of *Bcl-2*'s, which interacts in a similar non-specific way with 2E4, confirms nicely this statement. Indeed, despite different initial conditions, the G4 again positions itself exposing a quartet to the 2E4 DARPIn interaction region. On the contrary, the interaction with 2G10 leads to the exposure of the nucleic acid backbone to the contact region of DARPIn and hence, to a selective recognition. As a matter of fact, these results are also coherent with the contact frequency analysis showing that *Bcl-2* interaction with 2G10 is mostly driven by highly conserved and persistent amino acids.

2E4 recognizes a peculiar structural motif of *c-Myc*

c-Myc is the G4 more consistently promoting an interaction via its backbone (Figure 8). This, in turn, could also point to a greater specificity of its recognition, although *c-Myc*'s is able to interact via its backbone with both 2E4 and 2G10. Thus, to

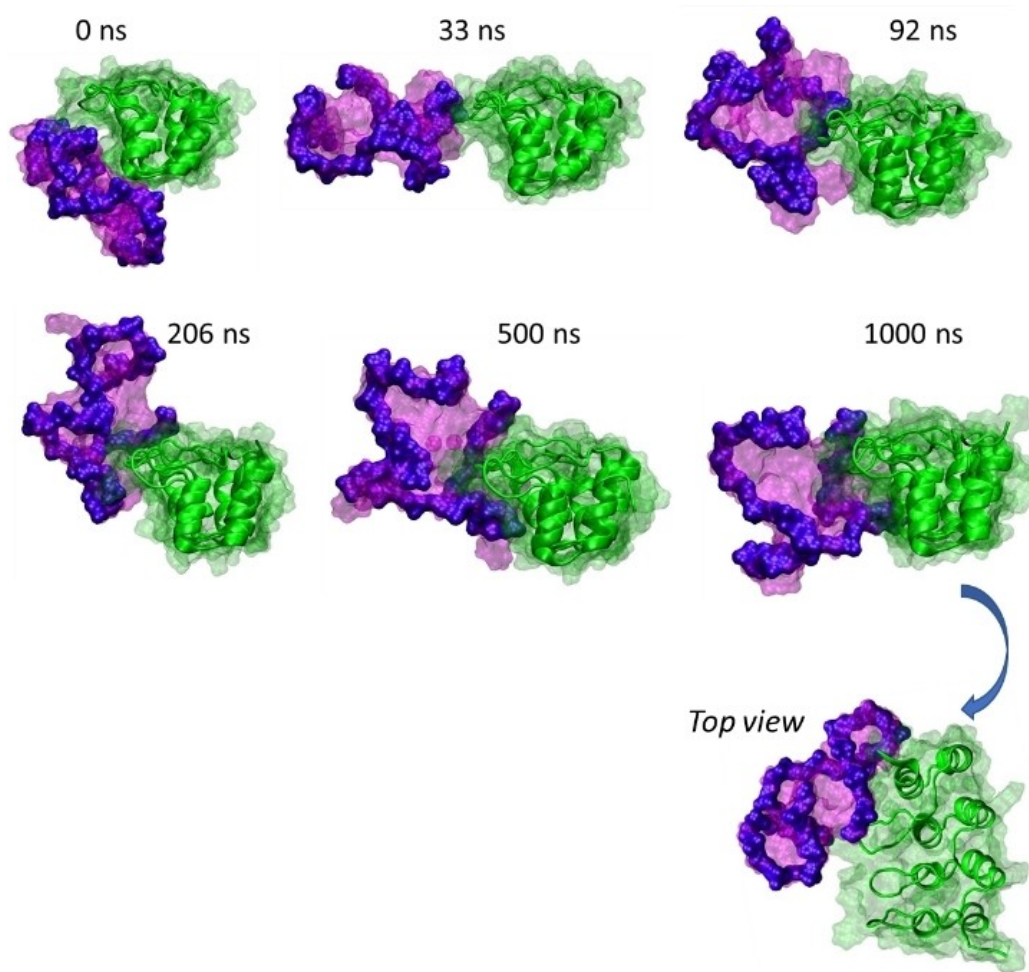


Figure 6. Screenshots of the dynamics of a 2E4/*h-Telo* complex. The *h-Telo* quartet is initially oriented towards the α -helix. But within the first 150 ns, the G-quadruplex rotates and then reorients to interact with the protein loops. During the dynamics, the movements of the G-quadruplex do not affect the protein/G-quartet loops interaction.

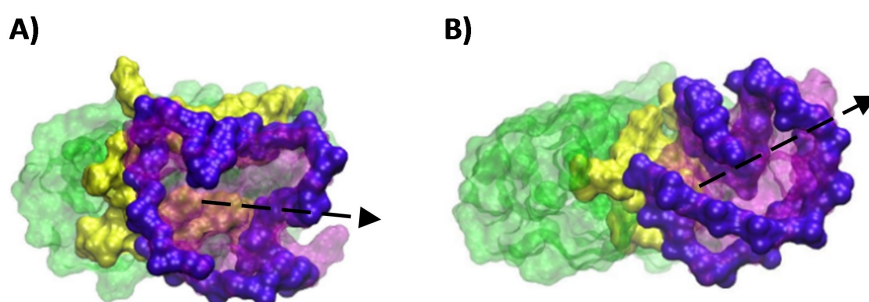


Figure 7. Non-specific 'nucleotide face' interaction found in cluster of the A) 2E4/*h-Telo* pose 1–1 (77.23% of the MDs) and B) 2E4/*Bcl-2* pose 1–1 (79.01% of the MDs) complexes.

further justify the selectivity of 2E4 a structural motif specifically recognized by this DARPin should be identified.

The main factor that could lead to a recognized structural motif includes the presence of a backbone folding involving the nucleotides most frequently in contact with the DARPin. This

feature can be easily assessed by clustering the MD simulation while checking the maintenance of the interaction patterns in the most populated clusters. By highlighting the E24 highest frequency contact nucleotides, i.e., G6 to A12, A21, and A22, we see that they are involved in the interaction with the protein for

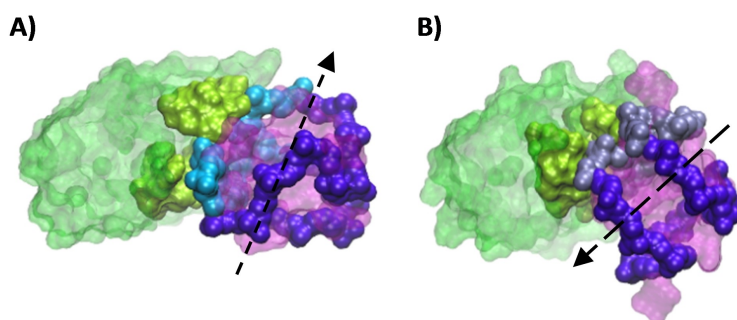


Figure 8. *c-Myc* G-quadruplex interacts preferentially through its backbone A) in the dynamics resulting from the most stable docking pose and B) in the dynamics resulting from a less favorable pose in which the G-quadruplex reorients itself to interact in a pose like the most stable pose found by docking. The previously identified high frequency nucleotides are colored A) in cyan and B) in steel blue, respectively.

the two most important clusters (Figures 8 and 9A, B). However, the two clusters differ by a rotation of about 180° of the G4 on the protein surface (pose 1–1: 78.47% of the MDs and pose 6–4: 79.55% of the MDs), as shown in Figure 10. Nonetheless, a well conserved structural motif is evidenced, determined by the folding of the G4 backbone into a U-shaped loop with an extruded nucleotide, further completed by a horizontal extension to the right, and overlaid by a dangling segment (Figure 9). Interestingly, all the structural characteristics are well evidenced in the most populated cluster, while in the secondary structures their identification remains more elusive. Indeed, if the linear extension remains evident, as well as the extruded nucleotide, the U-shaped loop and the appendix are more scarcely visible.

The amino acids located in the interaction site are also conserved between the two most populated clusters (Figure 10). Residues Y45, R70, and R79 organizes around the loop at an average distance of 3 Å, while Y35 is oriented towards the appendix and R34 points towards the linear extension. At a slightly higher distance of around 5 Å, M46 is interacting with the U-shaped loop, W68 is oriented towards the appendix, while R12 and D33 flank the linear extension. In addition, S78

stays close to the extruded nucleotide, probably assuring a further stabilization.

By superposing the two most populated clusters of the 2E4/*c-Myc* complexes a similar positioning of the G4 on the protein site is also observed, which is again consistent with the recognized backbone-based structural motif (Figure 10D). Finally, the increase of the contact frequency observed in the analysis of the MD simulations correlates well with the specific recognition of *c-Myc* by 2E4. Indeed, the region presenting the highest increase in the contact frequency corresponds to the residues recognizing the U-shaped loop and the extruded nucleotide. This, together with the similarity of the interaction pattern found in the superposition of the two G4 s poses, further validates the hypothesis that the selective recognition of *c-Myc* by 2E4 is driven by the structural motif we have identified. Because of this structural-based recognition, and the often-invoked analogy between DARPin and antibodies, it is tempting to characterize this interaction pattern as an epitope/paratope recognition. Here the paratope-like element being the 2E4 interaction site, and the epitope-like region the G4 structural motif identified for *c-Myc*.

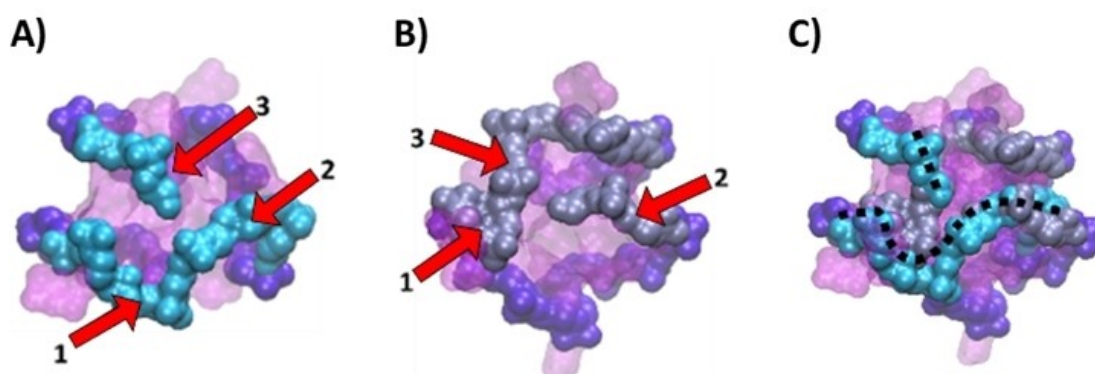


Figure 9. Face view of the *c-Myc* G-quadruplex interacting with 2E4 A) in the dynamics resulting from the most stable docking pose (pose 1–1) and B) in the dynamics resulting from a less favorable pose (pose 6–4): a rotation of 180 degrees with respect to A). Three elements can be identified on each: 1) a U-shaped-loop with an extruded nucleotide, 2) a linear extension and 3) an appendage. The superposition of the two interaction poses C), identify the recognized motif in dotted line.

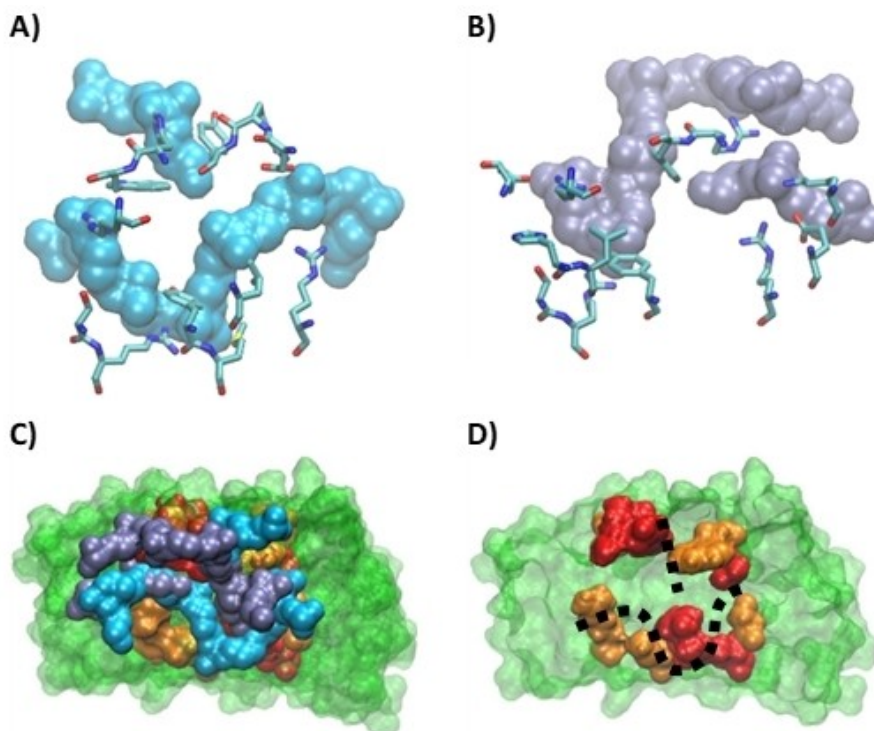


Figure 10. Protein residue involved in the interaction with the *c-Myc* structural motif A) in the dynamics resulting from the most stable docking pose (pose 1–1) and B) in the dynamics resulting from a less favorable pose (pose 6–4). C) Superposition of the complexes of the two clusters. D) Interaction site of 2E4 with *c-Myc*; residues at 3 Å are in red and those at 5 Å are in orange.

Conclusion

Our results highlight two modes of interaction for DARPins/G4 complexes. The first one is a non-specific recognition that is established when G4 interacts through its guanine tetrad, or through peripheral nucleotides π -stacked with the tetrads. The second binding mode is driven by the specific recognition of the conformation of the G4 backbone and leads to a DARPIn/G4 paratope/epitope like recognition. This specific mode, which we have identified for the 2E4/*c-Myc* complex is based on a peculiar folding motif of the G4 backbone and presents a U-shaped loop with a linear extension and an overhanging short appendix. Consequently, a large extension of the U-shaped loop, also including extruded nucleotides should enhance selective recognition of the G4 s. Conversely, the identification of backbone-based recognition motifs could also improve the rational design of DARPins. Indeed, the quest for selective G4 ligands has a tremendous significance, especially in the proposition of specific anticancer or antiviral agents. Our results, and the first identification of paratope/epitope specific structural recognition may lead to significant development in the design of potentially therapeutics peptides targeting specific G4 arrangements.

Experimental Section

Structure of G4 and reconstruction of DARPins: The structure of the *h-Telo* G4 was retrieved from PDB data bank 1KF1,^[42] as well as that of the *c-Myc* (1XAV)^[43] and *Bcl-2* (6ZX7).^[44] The sequences of 2E4 and 2G10 DARPins were obtained from the Supporting Information of Ref. [34] and their structure reconstructed with the SWISS-MODEL server.^[45] 2E4 were reconstructed based on high similarity with the PDB entry 2CH4^[46] and 2G10 was reconstructed on the basis of the 1SVX structure^[47] similarity.

Sequence alignment. DM102 peptide and DARPins sequences were aligned using the Clustal Omega on EBI server^[48] and the M-coffee server,^[49] using their default parameters.

Docking and selection of initial structures: The reconstructed DARPins and G4 were loaded onto the HADDOCK server^[50] to perform protein/nucleic acid docking while searching the entire protein and the entire G4 structure and using the standard HADDOCK parameters. Three poses were selected from the docking results (Figure 11), always including the most favorable one. The selection was based on the relative position of G4 with respect to the DARPIn. The three poses correspond to an interaction with the G4 backbone, an interaction with the tetrads facing the nucleotides and an interaction developed in a peripheral region of the DARPIn. This choice allowed to assure a significant sampling of a complex conformational space, also including rather unfavorable interaction areas, such as the one corresponding to the peripheral binding.

Molecular dynamics simulations: MD simulations has been performed for 2E4 and 2G10 interacting with *c-Myc*, *Bcl-2* and *h-Telo*. Three poses for each complex have been used as starting conditions. Each system was calculated in two independent replicates of 1 μ s each, thus a total of 36 simulations of DARPIn/G4

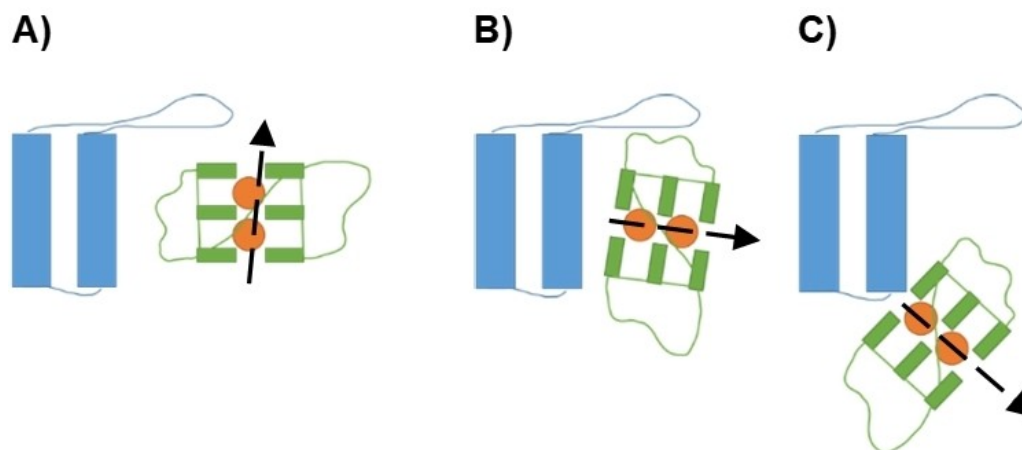


Figure 11. Schematic representation of the three types of poses selected from the docking. There is interaction type A) backbone, B) nucleotide face and C) protein bottom zone.

complexes have been performed. In addition, simulations of the free G4s and DARPin have also been obtained as a control. All simulations have been run using the NAMD software^[51,52] with the Amber parm99 force field^[53] including the bsc1 corrections^[54] for nucleic acids. A truncated octahedral box of TIP3P^[55] water was used to solvate the systems, using periodic boundary condition (PBC). All the calculations were performed in the isothermal and isobaric (NPT) ensemble at a temperature of 300 K and a pressure of 1 atm. A minimal concentration of K⁺ ions was added to assure charge equilibration. Hydrogen Mass Repartitioning (HMR)^[56] was consistently used, allowing, in combination with the Rattle and Shake algorithms,^[57] a timestep of 4 fs to integrate Newton's equations of motion. Finally, the trajectories were analyzed and visualized with VMD,^[58] as well as a dedicated script to retrieve G4 structural parameters,^[59] while CPPTRAJ^[60] was used for clustering.

Acknowledgements

All the calculations were performed on the local cluster of the LPCT and on the regional cluster ExpLor of the University of Lorraine. CNRS and Université de Lorraine and Université Paris Cité, France, and Università di Palermo and from the Ministero dell'Università e Ricerca Scientifica e Tecnologica, are thanked for their support. T.M. thanks University of Palermo for funding a joint Ph.D. program. A.M. thanks ANR (Agence Nationale de la Recherche) and CGI (Commissariat à l'Investissement d'Avenir) for their financial support of this work through Labex SEAM (Science and Engineering for Advanced Materials and devices) ANR 11 LABX 086 and ANR 11 IDEX 05 02. The support of the IdEx "Université Paris 2019" ANR-18-IDEX-0001 and of the Platform P3MB is gratefully acknowledged.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

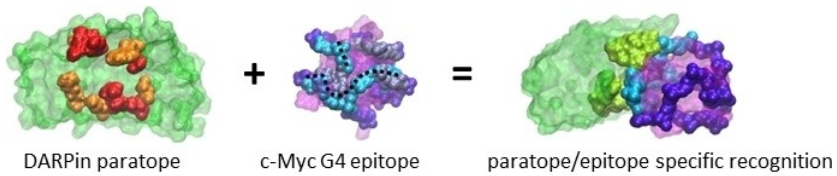
Keywords: *c-Myc* promoter · DARPin · epitope/paratope recognition · guanine quadruplex · molecular dynamics

- [1] H. J. Lipps, D. Rhodes, *Trends Cell Biol.* **2009**, *19*, 414–422.
- [2] M. L. Bochman, K. Paeschke, V. A. Zakian, *Nat. Rev. Genet.* **2012**, *13*, 770–780.
- [3] C. K. Kwok, C. J. Merrick, *Trends Biotechnol.* **2017**, *35*, 997–1013.
- [4] J. Spiegel, S. Adhikari, S. Balasubramanian, *Trends Chem.* **2020**, *2*, 123–136.
- [5] M. García-Díaz, D. Sánchez-García, J. Soriano, M. L. Sagristà, M. Mora, Á. Villanueva, J. C. Stockert, M. Cañete, S. Nonell, *MedChemComm* **2011**, *2*, 616–619.
- [6] R. Hänsel-Hertsch, D. Beraldi, S. V. Lensing, G. Marsico, K. Zyner, A. Parry, M. Di Antonio, J. Pike, H. Kimura, M. Narita, et al., *Nat. Genet.* **2016**, *48*, 1267–1272.
- [7] J. F. Moruno-Manchon, P. Lejault, Y. Wang, B. McCauley, P. Honarpisheh, D. A. M. Scheihing, S. Singh, W. Dang, N. Kim, A. Urayama, et al., *eLife* **2020**, *9*, 10.7554/eLife.52283.
- [8] R. Simone, P. Fratta, S. Neidle, G. N. Parkinson, A. M. Isaacs, *FEBS Lett.* **2015**, *589*, 1653–1668.
- [9] G. Biffi, D. Tannahill, J. Miller, W. J. Howat, S. Balasubramanian, *PLoS One* **2014**, *9*, 10.1371/journal.pone.0102711.
- [10] J. Carvalho, J. L. Mergny, G. F. Salgado, J. A. Queiroz, C. Cruz, *Trends Mol. Med.* **2020**, *26*, 848–861.
- [11] Y.-Z. Xu, P. Jenjaroenpun, T. Wongsurawat, S. D. Byrum, V. Shponka, D. Tannahill, E. A. Chavez, S. S. Hung, C. Steidl, S. Balasubramanian, et al., *NAR Cancer* **2020**, *2*, 10.1093/narcan/zcaa029.
- [12] N. Kosiol, S. Juranek, P. Brossart, A. Heine, K. Paeschke, *Mol. Cancer* **2021**, *20*, 10.1186/s12943-021-01328-4.
- [13] N. Kumari, S. V. Vartak, S. Dahal, S. Kumari, S. S. Desai, V. Gopalakrishnan, B. Choudhary, S. C. Raghavan, *iScience* **2019**, *21*, 288–307.
- [14] S. Bielskutel, J. Plavec, P. Podbevšek, *J. Am. Chem. Soc.* **2019**, *141*, 2594–2603.
- [15] A. M. Fleming, J. Zhu, Y. Ding, C. J. Burrows, *ACS Chem. Biol.* **2017**, *12*, 2417–2426.
- [16] C. C. Hognon, T. Miclot, C. G. Iriepa, A. Francés-Monerris, S. Grandemange, A. Terenzi, M. Marazzi, G. Barone, A. Monari, C. Garcia-Iriepa, et al., *J. Phys. Chem. Lett.* **2020**, *11*, 5661–5667.
- [17] T. Miclot, C. Hognon, E. Bignon, A. Terenzi, M. Marazzi, G. Barone, A. Monari, *J. Phys. Chem. Lett.* **2021**, *12*, 10277–10283.

- [18] A. Francés-Monerris, C. Hognon, T. Miclot, C. García-Iriepa, I. Iriepa, A. Terenzi, S. Grandemange, G. Barone, M. Marazzi, A. Monari, *J. Proteome Res.* **2020**, *19*, 4291–4315.
- [19] E. Ruggiero, S. N. Richter, *Annu. Rep. Med. Chem.* **2020**, *54*, 101–131.
- [20] C. Zhao, G. Qin, J. Niu, Z. Wang, C. Wang, J. Ren, X. Qu, *Angew. Chem. Int. Ed.* **2021**, *60*, 432–438; *Angew. Chem.* **2021**, *133*, 436–442.
- [21] S. R. Wang, Q. Y. Zhang, J. Q. Wang, X. Y. Ge, Y. Y. Song, Y. F. Wang, X. D. Li, B. S. Fu, G. H. Xu, B. Shu, et al., *Cell Chem. Biol.* **2016**, *23*, 1113–1122.
- [22] N. Saranathan, P. Vivekanandan, *Trends Microbiol.* **2019**, *27*, 148–163.
- [23] M. C. Chen, R. Tippana, N. A. Demeshkina, P. Murat, S. Balasubramanian, S. Myong, A. R. Ferré-D'amaré, *Nature* **2018**, *558*, 465–483.
- [24] Z. Pietras, M. A. Wojcik, L. S. Borowski, M. Szewczyk, T. M. Kulinski, D. Cysewski, P. P. Stepien, A. Dziembowski, R. J. Szczesny, *Nat. Commun.* **2018**, *9*, 10.1038/s41467-018-05007-9.
- [25] J. C. Darnell, K. B. Jensen, P. Jin, V. Brown, S. T. Warren, R. B. Darnell, *Cell* **2001**, *107*, 489–499.
- [26] G. Farine, C. Migliore, A. Terenzi, F. Lo Celso, A. Santoro, G. Bruno, R. Bonsignore, G. Barone, *Eur. J. Inorg. Chem.* **2021**, 1332–1336.
- [27] V. Dharmodharan, P. I. Pradeepkumar, *ACS Chem. Biol.* **2019**, *14*, 2102–2114.
- [28] D. Drygin, A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. B. Ho, C. Proffitt, K. Trent, J. P. Whitten, et al., *Cancer Res.* **2009**, *69*, 7653–7661.
- [29] D. Dutta, M. Debnath, D. Müller, R. Paul, T. Das, I. Bessi, H. Schwalbe, J. Dash, *Nucleic Acids Res.* **2018**, *46*, 5355–5365.
- [30] S. M. Javadekar, N. M. Nilavar, A. Paranjape, K. Das, S. C. Raghavan, *DNA Res.* **2020**, *27*, 10.1093/dnares/dsaa024.
- [31] A. Henderson, Y. Wu, Y. C. Huang, E. A. Chavez, J. Platt, F. Brad Johnson, R. M. Brosh, D. Sen, P. M. Lansdorp, *Nucleic Acids Res.* **2017**, *45*, 6252.
- [32] H. Y. Liu, Q. Zhao, T. P. Zhang, Y. Wu, Y. X. Xiong, S. K. Wang, Y. L. Ge, J. H. He, P. Lv, T. M. Ou, et al., *Cell Chem. Biol.* **2016**, *23*, 1261–1270.
- [33] H. Fernando, R. Rodriguez, S. Balasubramanian, *Biochemistry* **2008**, *47*, 9365–9371.
- [34] O. Scholz, S. Hansen, A. Plückthun, *Nucleic Acids Res.* **2014**, *42*, 9182–9194.
- [35] P. R. Mittl, P. Ernst, A. Plückthun, *Curr. Opin. Struct. Biol.* **2020**, *60*, 93–100.
- [36] A. Plückthun, *Annu. Rev. Pharmacol. Toxicol.* **2015**, *55*, 489–511.
- [37] M. T. Stumpp, H. K. Binz, P. Amstutz, *Drug Discovery Today* **2008**, *13*, 695–701.
- [38] B. Heddi, V. V. Cheong, H. Martadinata, A. T. Phan, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9608–9613.
- [39] M. Y. Yaneva, V. V. Cheong, J. K. Cheng, K. W. Lim, A. T. Phan, *Biochem. Biophys. Res. Commun.* **2020**, *531*, 62–66.
- [40] A. Minard, D. Morgan, F. Raguseo, A. Di Porzio, D. Liano, A. G. Jamieson, M. Di Antonio, *Chem. Commun.* **2020**, *56*, 8940–8943.
- [41] I. M. Wallace, O. O'Sullivan, D. G. Higgins, C. Notredame, *Nucleic Acids Res.* **2006**, *34*, 1692–1699.
- [42] G. Parkinson, M. Lee, S. Neidle, *Exp. Mol. Pathol.* **2002**, *417*, 876–880.
- [43] A. Ambrus, D. Chen, J. Dai, R. A. Jones, D. Yang, *Biochemistry* **2005**, *44*, 2048–2058.
- [44] S. Bielskute, J. Plavec, P. Podbevšek, *Nucleic Acids Res.* **2021**, *49*, 2346–2356.
- [45] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, et al., *Nucleic Acids Res.* **2018**, *46*, W296–W303.
- [46] O. Grubisha, M. Kaminska, S. Duquerroy, E. Fontan, F. Cordier, A. Haouz, B. Raynal, J. Chiaravalli, M. Delepierre, A. Israël, et al., *J. Mol. Biol.* **2010**, *395*, 89–104.
- [47] H. K. Binz, P. Amstutz, A. Kohl, M. T. Stumpp, C. Briand, P. Forrer, M. G. Grütter, A. Plückthun, *Nat. Biotechnol.* **2004**, *22*, 575–582.
- [48] F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn, et al., *Nucleic Acids Res.* **2019**, *47*, W636–W641.
- [49] S. Moretti, F. Armougom, I. M. Wallace, D. G. Higgins, C. V. Jongeneel, C. Notredame, *Nucleic Acids Res.* **2007**, *35*, 10.1093/nar/gkm333.
- [50] G. C. P. Van Zundert, J. P. G. L. M. Rodrigues, M. Trellet, C. Schmitz, P. L. Kastiris, E. Karaca, A. S. J. Melquiond, M. Van Dijk, S. J. De Vries, A. M. J. J. Bonvin, *J. Mol. Biol.* **2016**, *428*, 720–725.
- [51] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- [52] J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hémin, W. Jiang, et al., *J. Chem. Phys.* **2020**, *153*, 044130.
- [53] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- [54] I. Ivani, P. D. Dans, A. Noy, A. Pérez, I. Faustino, A. Hospital, J. Walther, P. Andrio, R. Goñi, A. Balaceanu, et al., *Nat. Methods* **2016**, *13*, 55–58.
- [55] P. Mark, L. Nilsson, *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- [56] C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.
- [57] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952–962.
- [58] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33–38.
- [59] V. Tsvetkov, G. Pozmogova, A. Varizhuk, *J. Biomol. Struct. Dyn.* **2016**, *34*, 705–715.
- [60] V. V. Shitov, N. A. Semenov, N. Y. Gozman, *Telecommun. Radio Eng. (English Transl. Elektrosyaz Radiotekhnika)* **1984**, *38–39*, 14–16.

Manuscript received: June 13, 2022
Accepted manuscript online: July 6, 2022
Version of record online: ■■■, ■■■■

RESEARCH ARTICLE



DARPins are proteins able to specifically recognize G-quadruplexes. Their small size combined with their easy design makes them a good competitor to antibodies for the identification and localization of G-quadruplex.

Using molecular dynamics calculations, we show that the selectivity of DARPins towards G-quadruplexes is achieved by the recognition of a structural motif adopted by the DNA filament folding.

T. Miclot, Dr. E. Bignon, Prof. A. Terenzi, Prof. S. Grandemange, Prof. G. Barone*, Prof. A. Monari**

1 – 11

G-Quadruplex Recognition by DARPins through Epitope/Paratope Analogy

