

Article

# Towards Computational Modeling of Ligand Binding to the ILPR G-Quadruplex

Xiaotong Zhang <sup>1</sup>, John Barrow <sup>2</sup>, Tanja van Mourik <sup>1,\*</sup> and Michael Bühl <sup>1,\*</sup><sup>1</sup> EaStCHEM School of Chemistry, University of St Andrews, North Haugh, St Andrews KY16 9ST, UK<sup>2</sup> School of Medicine, Medical Sciences and Nutrition, Institute of Education in Healthcare and Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK

\* Correspondence: tanja.vanmourik@st-andrews.ac.uk (T.v.M.); buehl@st-andrews.ac.uk (M.B.)

**Abstract:** Using a combination of unconstrained and constrained molecular dynamics simulations, we have evaluated the binding affinities between two porphyrin derivatives (TMPyP4 and TEGPy) and the G-quadruplex (G4) of a DNA fragment modeling the insulin-linked polymorphic region (ILPR). Refining a well-established potential of mean force (PMF) approach to selections of constraints based on root-mean-square fluctuations results in an excellent agreement between the calculated and observed absolute free binding energy of TMPyP4. The binding affinity of IPLR-G4 toward TEGPy is predicted to be higher than that toward TMPyP4 by 2.5 kcal/mol, which can be traced back to stabilization provided by the polyether side chains of TMPyP4 that can nestle into the grooves of the quadruplex and form hydrogen bonds through the ether oxygen atoms. Because our refined methodology can be applied to large ligands with high flexibility, the present research opens an avenue for further ligand design in this important area.

**Keywords:** G-quadruplex; potential of mean force; ligand with high flexibility

## 1. Introduction

A G-quadruplex (G4) is a secondary structure of nucleic acids that are formed within specific repetitive guanine-rich DNA sequences. Rather than pairwise aggregation leading to the well-known double helix motif, four guanine moieties bind with each other, forming a G-quartet. Several G-quartets stack to form a G4 structure, which is stabilized by monovalent cations, located in the central cavity, either parallel with a quartet (coordinated to the carbonyl oxygens pointing inwards) or in between two successive quartets (see Figure 1a). Depending on the loop structure of the DNA strands involved, G4 DNA can adopt three conformations: parallel (see Figure 1b), anti-parallel, and hybrid [1–3].

Previous experiments have shown that G4 structures may be associated with gene regulation when they form in the promoter region, the DNA sequence to which proteins bind to initiate transcription [1,4–10]. Of particular interest is the formation of G4 structures in the insulin-linked polymorphic region (ILPR), as the G4 structure formed in ILPR (ILPR G4) may regulate the transcriptional activity of insulin [11–17]. Although the reason for the altered insulin gene expression by the ILPR G4 is unclear, previous studies suggest that the transcription factor Pur-1 can bind to G4 structures within the ILPR, thereby enhancing the recognition of RNA polymerase in the promoter region [11]. To further investigate the relationship between the ILPR G4 and gene regulation in situ, it is necessary to explore bespoke ligands, such as small molecules that display high affinity towards the ILPR G4 DNA. Because parallel conformations are more likely to prevail in the presence of intracellular molecular crowding [1,18,19], ligands with a high affinity for parallel G4s are particularly desirable.

**Citation:** Zhang, X.; Barrow, J.; van Mourik, T.; Bühl, M. Towards Computational Modeling of Ligand Binding to the ILPR G-Quadruplex. *Molecules* **2023**, *28*, 3447. <https://doi.org/10.3390/molecules28083447>

Academic Editor: Teobald Kupka

Received: 23 February 2023

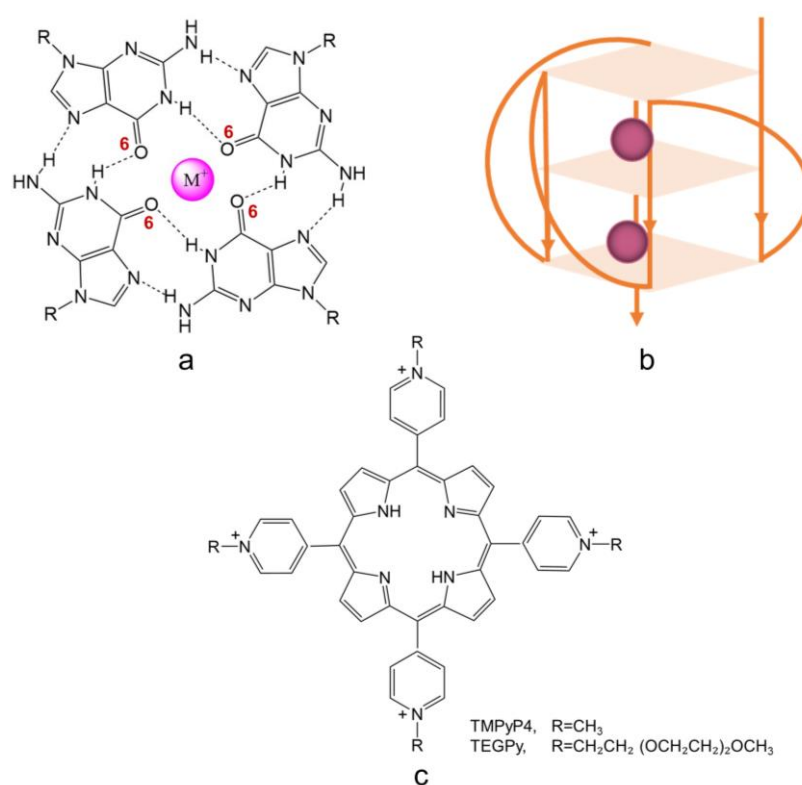
Revised: 31 March 2023

Accepted: 2 April 2023

Published: 13 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).



**Figure 1.** Schematic illustration of (a) a G-quartet stabilized by monovalent cations (such cations are usually coordinated to the inwardly pointing carbonyl oxygen atoms (O6) of the guanine bases forming the quartet); (b) a G4 structure with parallel conformation; and (c) G4 binding ligands with aromatic structural features; R=CH<sub>3</sub> or CH<sub>2</sub>CH<sub>2</sub> (OCH<sub>2</sub>CH<sub>2</sub>)<sub>2</sub>OCH<sub>3</sub>, representing TMPyP4 and TEGPy, respectively.

The past decade has seen widespread interest in the development of potent G4 ligands. Among these ligands, TMPyP4 (5,10,15,20-tetra(N-methyl-4-pyridyl) porphyrin, Figure 1c), a tetra-cationic porphyrin derivative, has been well studied owing to its high affinity towards the external G-quartet of G4s [1,20–23]. Although it is controversial whether TMPyP4 displays selectivity towards different G4 structures, its aromatic structural characteristic and high sensitivity to the G-quartet make it a prominent candidate in ligand design. To improve the ability of ligands to differentiate between different G4 structures, side chains can be introduced into the design of ligands to further interact with G4s through groove and loop binding [23,24]. One such ligand that has been studied experimentally is TEGPy (tetra-[(ethylene glycol)-pyridyl]-porphyrin) (Figure 1c) [25,26]. Apart from the same polyaromatic core as TMPyP4, TEGPy comprises four PEG (polyethylene glycol) arms that may have additional interactions with the grooves or loops of the ILPR G4 structure, making it a promising ligand model for studying higher specificity binding to the ILPR G4 DNA.

Experimentally, ligand affinity can be estimated from the equilibrium binding constants obtained by several methods, including surface plasmon resonance (SPR) [27] and isothermal titration calorimetry (ITC) [28]. Nevertheless, it is challenging for experimental methods to determine the binding affinity with specific binding modes, especially when the difference in binding affinity between different binding modes is small. In addition, it is difficult to assess the dynamics of specific interactions between ligands and target molecules. Thus, computational approaches such as free-energy simulations are invaluable for providing quantitative mechanistic information on binding specificity and affinity. They can also predict binding constants for as-yet-unknown ligands, which could build the basis for the *in silico* design of inhibitors or regulators.

Recently, Deng et al., reported the binding affinity of c-MYC G4-quindoline in two binding pockets using the potential of mean force (PMF) approach [29]. Their results are in general agreement with those obtained by SPR, which demonstrates the validity of the PMF approach to study the c-MYC G4-quindoline system. Previous computational studies have demonstrated the superiority of free-energy simulations in characterizing the thermodynamics of ligand-G4 binding [29–31]. However, most of the studies have focused on relatively small ligand systems, and there is little research on ligands such as TEGPy that are large, highly flexible, and display multiple binding modes simultaneously towards G4s. This is because the flexibility and large size of such ligands will increase the conformational sampling spaces, which require extensive computational resources and may result in convergence problems for the free-energy simulations.

In this study, we investigate the absolute binding free-energy and ligand interactions of ILPR G4-ligand systems using two porphyrin derivatives through the Root-Mean-Square Fluctuation (RMSF)-based PMF protocol. By evaluating the RMSF of the ligands, we make a rational selection of restraint atoms to assist the system in reaching equilibrium in free-energy perturbation. This makes the free-energy approach applicable to large, flexible G4-ligand systems, which otherwise pose big challenges to this type of free-energy simulation. With our refinement of the PMF protocol, we can accurately calculate the absolute binding free energy of the ILPR G4-TMPyP4 system, and rationally predict the absolute binding free energy of the large and flexible TEGPy ligand towards the ILPR G4 structure. Interactions between the four side arms of TEGPy and the groove of the ILPR G4 structure will also be highlighted. The comparison between TEGPy and TMPyP4 allows assessment of the ligand affinity with different binding modes towards the ILPR G4 DNA, which is essential for probe design and for assisting experiments to unravel the relationship between G4 formation and regulation of the human insulin gene.

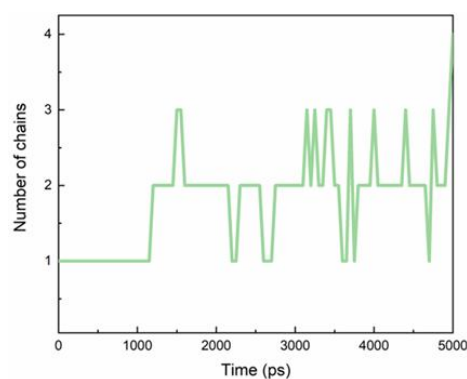
## 2. Results and Discussions

**Conformational sampling of the ILPR G4-ligand systems using unconstrained MD simulations.** To fully understand the structural features and dynamic behavior of the ILPR G4-ligand systems, unconstrained MD simulations were performed. The structural fluctuations of the ILPR G4-TMPyP4 and the ILPR G4-TEGPy systems during the simulations are analyzed using the RMSD and RMSF values, with the reference structure being the equilibrium structure obtained from the unconstrained MD simulations.

The geometry of the ILPR G4 and the ILPR G4-TMPyP4 complex displayed slight conformational changes during the 5 ns unconstrained MD simulation (Figure S1a). Both systems reached equilibrium at ~500 ps and maintained their structures throughout the simulation with a structural fluctuation of RMSD ~4.5 Å, suggesting that the  $\pi$ - $\pi$  interactions between the G4 and TMPyP4 moieties are well conserved. The computed RMSF of the G4 (Figure S1c) varies from loop regions (larger fluctuation) to quartets (smaller fluctuation), indicating a relatively rigid conformation in the quartet. This is due to the  $\pi$ - $\pi$  interactions between the outer quartet and TMPyP4, which in turn lead to similar small magnitudes of fluctuations in the atoms close to the aromatic ring of TMPyP4 (Figure S1d). Due to the additional side arms, the conformation of TEGPy is more flexible than that of TMPyP4, which in turn affects the G4 structure. As shown in Figure S2a, both the ILPR G4 and the ILPR G4-TEGPy systems present a structural fluctuation, with an RMSD ~5 Å, which is slightly larger than that of the ILPR G4-TMPyP4 system. However, the RMSF shows that TEGPy undergoes unequal structural changes compared to TMPyP4 (S2d vs S1d). This is apparent in the  $\pi$  plane, where the structural fluctuation of each atom is supposed to be very similar to each other. This finding highlights the importance of selecting the correct restraint atoms in the constrained MD simulations, or more specifically, the free-energy perturbation in restraining highly flexible ligands, such as TEGPy.

TEGPy was designed as a ligand for G4 structures because the four side arms are expected to interact with the G4 grooves [26]. We started our unconstrained MD simulation from a structure where these four arms had been placed manually into the four

grooves. The average number of side chains that nestled into the grooves during the MD simulation was evaluated by computing the distance between the outermost ether oxygen atom of the ligand and its nearest nitrogen atom from the guanine base of the quartet (Figure 2), assuming an arm is tightly bound when the arm-groove distance is below 3.5 Å. During the MD simulation, however, it turned out that only two or three of the arms are tightly bound in their grooves; the others are more flexible. This finding suggests the possibility that ligands with fewer side arms can be expected to bind just as well as, or even better than, ligands with four side arms. This conjecture will need to be confirmed by subsequent studies. For a detailed discussion of H-bonding between the side arms and the grooves, read the next paragraph.



**Figure 2.** The evolution of the average number of TEGPy arms that nestled in grooves of the ILPR G4.

**Absolute binding free energies for the ILPR G4-ligand systems from the RMSF-derived PMF approach.** The absolute binding free energies  $\Delta G_{bind}^{\circ}$  of the ILPR G4-TMPyP4 and ILPR G4-TEGPy systems and the individual components of these are listed in Table 1. The computed absolute binding free energy of the TMPyP4-ILPR G4 system is  $-9.18 (\pm 0.48)$  kcal·mol<sup>-1</sup>, which is in good agreement with the previous ITC results ( $-9.2$  kcal·mol<sup>-1</sup>) [32]. This validates the RMSF-based PMF approach, which accurately predicts the binding affinity between the ILPR G4 and its ligands. The computed absolute binding free energy of the TEGPy-ILPR G4 system is  $-11.50 (\pm 0.59)$  kcal·mol<sup>-1</sup>, approximately 2.3 kcal·mol<sup>-1</sup> larger (in absolute terms) than that of the complex with TMPyP4. This increase in the absolute binding free energy corresponds to a ca. 50-fold increase in binding affinity at room temperature (the computed free binding energies correspond to affinities of 0.005 nM<sup>-1</sup> and 0.240 nM<sup>-1</sup> for TMPyP4 and TEGPy, respectively). Since TEGPy might exhibit multiple binding modes towards the ILPR G4, the enhanced binding strength compared to TMPyP4 may be attributed to the additional binding, i.e., loop or groove binding. According to a recent study involving fluorescence correlation spectroscopy [33], TMPyP4 is indicated to bind slightly stronger to the 5' ends of human telomeric DNA sequences than to the 3' ends used in the present study (by a factor of ca. 7, albeit with quite a large uncertainty). Our predicted increase in binding affinity on going from TMPyP4 to TEGPy may thus be somewhat overestimated but should still be noticeable.

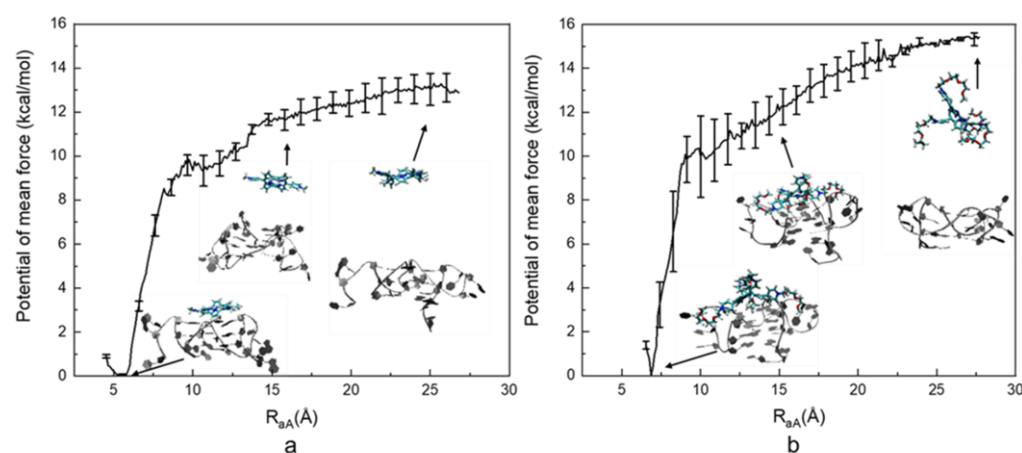
The contributions of each component of the absolute binding free energy in both systems were examined and are listed in Table 1. The computed free energy changes for restraining the ligand in its bound/unbound state are very close to the two different ligands ( $-5.62$  kcal·mol<sup>-1</sup> and  $-5.73$  kcal·mol<sup>-1</sup> for the TEGPy and TMPyP4 complex in its bound state; 9.54 and 9.43 kcal·mol<sup>-1</sup> for the TEGPy and TMPyP4 complex in its unbound state). Clearly, the strength in restraining the ligand to the bound state, or the unbound state (solvent) is very similar, and the main contribution to the difference in the ligand affinity of the two ligand systems comes from the potential of mean force.

**Table 1.** Absolute binding free energy in the ILPR G4- TMPyP4 and ILPR G4-TEGPy systems.

ILPR G4-Ligand Systems	$-\omega(r^*)$ (kcal·mol <sup>-1</sup> )	$-k_B T \ln \frac{\int_{bound} e^{-\omega(r)/k_B T} dr}{(2\pi k_B T)^{1/2} k_r}$ (kcal·mol <sup>-1</sup> )	$\Delta G_{restr}^{bulk}$ (kcal·mol <sup>-1</sup> )	$-\Delta G_{restr}^{bound}$ (kcal·mol <sup>-1</sup> )	$\Delta G_{bind}^\circ$ (kcal·mol <sup>-1</sup> )
ILPR G4-TEGPy <sup>a</sup>	-15.42 (±0.53)	-0.01 (±0.0006)	9.54 (±0.0)	-5.62 (±0.27)	-11.50 (±0.59)
ILPR G4-TMPyP4 <sup>a</sup>	-12.88 (±0.43)	-0.02 (±0.0007)	9.43 (±0.0)	-5.73 (±0.05)	-9.18 (±0.43)

<sup>a</sup> The error bars are obtained from standard deviations between results from two repeating 10 ns trajectories in the umbrella sampling windows (see computational details).

As indicated in Figure 3, the computed PMF of the ILPR G4-TMPyP4 system is  $-12.88 \pm 0.43$  kcal·mol<sup>-1</sup>. The lowest energy occurs at  $\sim 5.5$  Å, which is a rather deep minimum, with the free energy rising steeply on either side. As  $R_{aA}$  increases, the energy goes up rapidly until  $R_{aA}$  reaches  $\sim 8$  Å. After that, the energy rises at a moderate rate and then levels off after  $\sim 20$  Å, indicating the ligand is dissociated from the ILPR G4. The computed PMF of the TEGPy system is  $-15.42 \pm 0.53$  kcal·mol<sup>-1</sup>, which is  $2.54$  kcal·mol<sup>-1</sup> higher than that of the TMPyP4 system. The lowest energy occurs at  $\sim 6.5$  Å. As  $R_{aA}$  increases to  $\sim 10$  Å, the computed PMF of the TEGPy system rises more rapidly than that of the TMPyP4 system, as shown by the steeper slope in the profile, indicating a higher free-energy change. Then, as  $R_{aA}$  increases to  $12.5$  Å, the PMF gradually increases, suggesting weaker interactions between target and ligand molecules compared to the  $\pi$ - $\pi$  interactions observed at the previous stage. This phenomenon suggests that weaker interactions, such as hydrogen bonding, occur in the TEGPy system, leading to a higher affinity of the ligand. The energy profile finally levels off after  $25$  Å, where TEGPy is fully dissociated from the ILPR G4 DNA and reaches the solvent. It can be seen from the first intermediate conformation of the TEGPy system ( $R_{aA} \sim 6.5$  Å) that the ligand displays multiple interactions (groove and loop binding as well as  $\pi$ - $\pi$  interactions with the ILPR G4), which can explain the higher ligand affinity. In addition, compared to TMPyP4, TEGPy requires a longer distance to dissociate from its target, as some of the side arms can still interact with the external quartet or flanking bases of the outermost loop region. As exemplified in the third intermediate conformation in the TEGPy complex ( $R_{aA} \sim 20$  Å), a side chain of TEGPy is still close to the external quartet of ILPR G4, whereas TMPyP4 has already approached the solvation environment at a similar distance. It is clear that the loop and groove interactions formed in the ILPR G4-TEGPy system enable a higher affinity of TEGPy towards the ILPR G4 DNA.

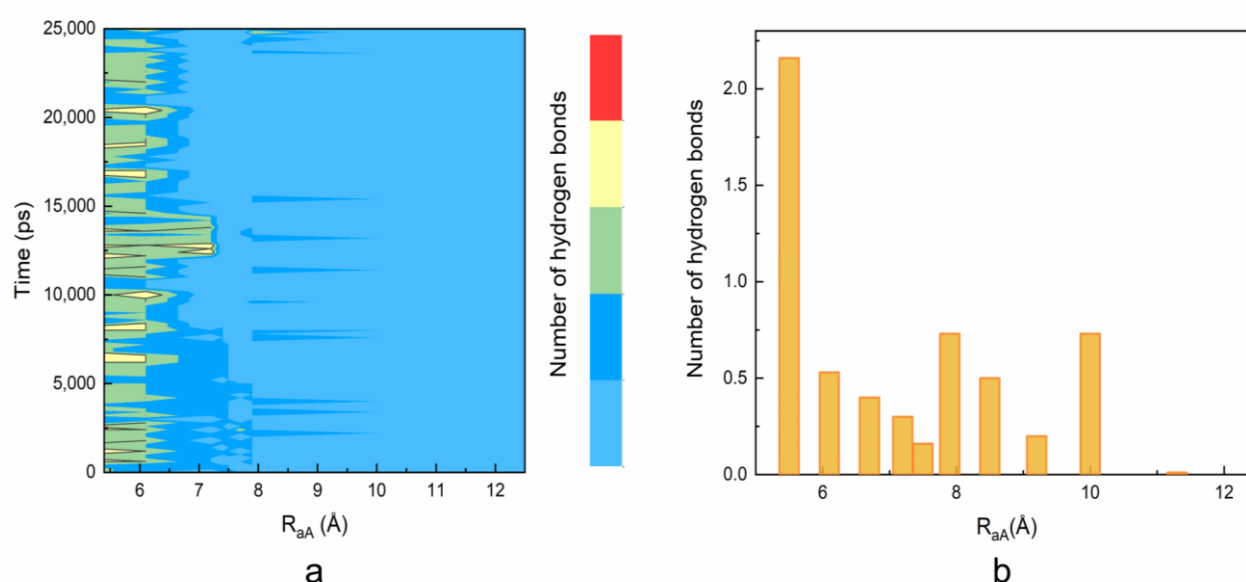


**Figure 3.** Computed potential of mean force of the ILPR G4-ligand systems. (a) PMF profile of the ILPR G4-TMPyP4 system. (b) PMF profile of the ILPR G4-TEGPy system. Intermediate conformations of the DNA-ligand complex at different reaction coordinates are also shown, where the DNA molecule is represented by new ribbons, and TMPyP4 is shown by licorice representations.

### The higher ligand affinity of TEGPy is attributed to the loop and groove binding.

To further understand the loop and groove interactions formed in the ILPR G4-TEGPy system, the number of intermolecular hydrogen bonds along the ligand dissociation was evaluated in Figure 2. An interaction was considered to be an intermolecular hydrogen bond if the distance between a potential donor and an acceptor atom is below 3.5 Å, and the donor-hydrogen-acceptor angle is larger than 150°. The number of intermolecular hydrogen bonds formed in the TEGPy system was measured using the H-bond plugin in VMD.

As shown in Figure 4a, two or three intermolecular hydrogen bonds are maintained between the ILPR G4 and TEGPy during the 25 ns constrained MD simulation with a  $R_{aA} \sim 5.5$  Å, which is in general agreement with that of the unconstrained MD simulation (Figure S3). This result validates the choice of restraint atoms, as the overall number of hydrogen bonds is not affected by the external restraint force. When  $R_{aA}$  elongates to 7 Å, the number of hydrogen bonds decreases to one and mostly ranges between zero and one when  $R_{aA}$  extends to 8 Å.

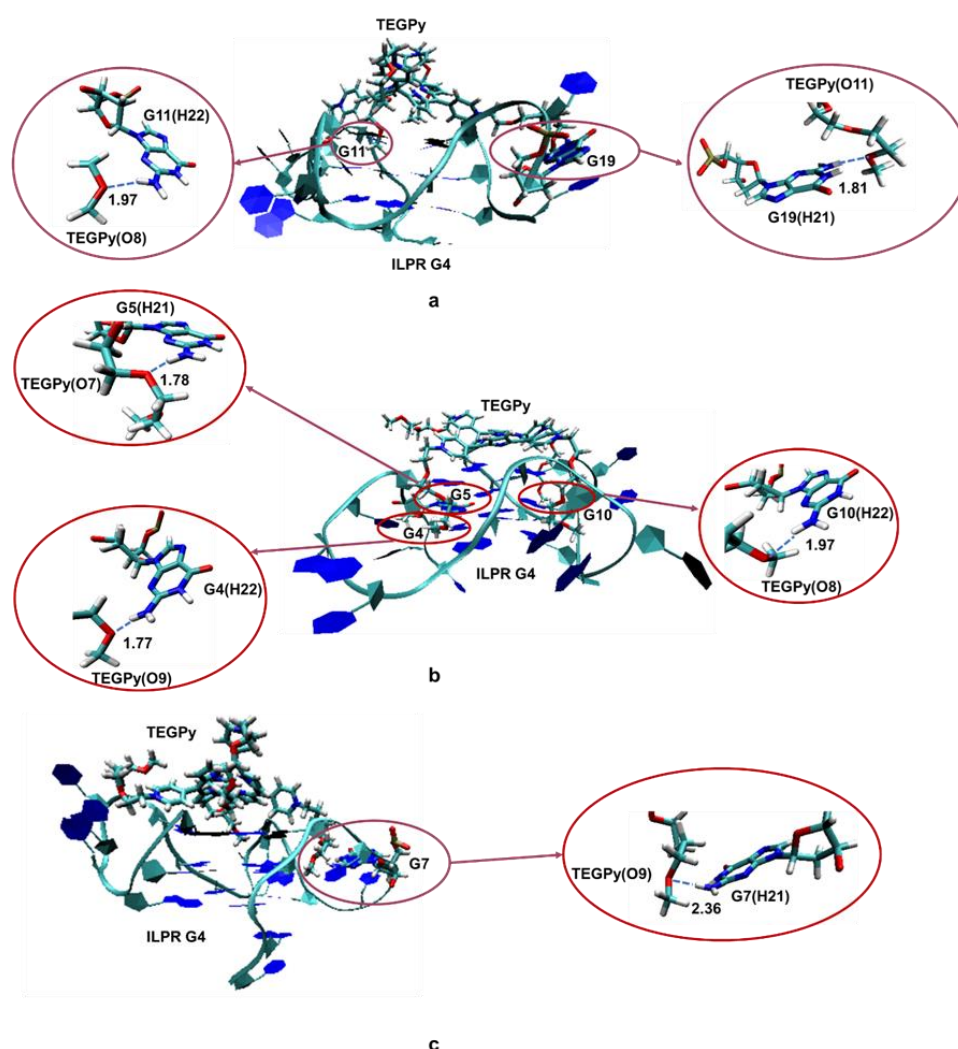


**Figure 4.** The evolution of the number of intermolecular hydrogen bonds in the TEGPy-ILPR G4 system along the ligand dissociation. (a) The number of intermolecular hydrogen bonds as a function of time and  $R_{aA}$ . The number of hydrogen bonds (ranging from 0 to 4) is indicated using a color scale. (b) The average number of intermolecular hydrogen bonds against  $R_{aA}$ . The number of hydrogen bonds is computed at each  $R_{aA}$  value as an average over the 25 ns constrained MD simulations.

Some hydrogen bonding between the ILPR and TEGPy is maintained at 10 Å, which confirms that TEGPy needs to reach a long distance from the ILPR G4 to fully dissociate from it. The average number of hydrogen bonds as a function of  $R_{aA}$  is shown in Figure 4b. In general, the number of hydrogen bonds decreases as the distance increases. There are two slight rebounds at around  $R_{aA} = 8$  Å and 10 Å and the number of hydrogen bonds levels off after 10 Å. We also investigated which regions of the ILPR G4-TEGPy system are involved in hydrogen bonding (Figure S4). Overall, the hydrogen bonding between TEGPy and the ILPR G4 is maintained within  $R_{aA} \sim 10$  Å, indicating the important role of hydrogen bonding in the ILPR G4-TEGPy binding. Some characteristic snapshots illustrating hydrogen bonding in the ILPR G4-TEGPy system are shown in Figure 5. There are two hydrogen bonds shown in Figure 5a. One occurs between the hydrogen atom (H22) of the amino group in guanine (G11, first quartet layer) and the oxygen atom (O8, Figure S5) of the ether group in TEGPy with a distance of 1.97 Å. The other one is formed between hydrogen atom H21 of G19 (flanking guanine base in the loop) and the oxygen atom (O11)

of the outermost ether group of the ligand (1.81 Å). In Figure 5b, hydrogen bonds exist between H21 of G5 (first layer guanine base), H22 of G4 (middle layer guanine base), H22 of G10 (middle layer guanine base), and O7, O9, and O8 of TEGPy, respectively. It can be seen from the given snapshots that O8 (oxygen atom of the outermost ether group in one arm of TEGPy) is involved in hydrogen bonding with two quartet layers in the groove region of the quadruplex. Similarly, O9, located in another side chain of TEGPy, interacts with both the loop region (G7) and groove region (G4) of the quadruplex through hydrogen bonding. In Figure 5c, there is a single hydrogen bond formed between the hydrogen atom (H21) of G7 (flanking guanine base in the loop) and the oxygen atom (O9) of the ligand (2.35 Å). The oxygen atoms participating in the hydrogen bonding are shown in Figure S5. These snapshots from the constrained MD simulation, together with the previous results, illustrate that groove binding and loop binding exist simultaneously in the ILPR G4-TEGPy system through hydrogen bonds, which contribute to the higher ligand affinity of TEGPy towards the G-quadruplex compared to TMPyP4.

How could the design of ILPR ligands be useful in therapeutic applications? One promising treatment for diabetes is the transplantation of tissue containing the insulin-



**Figure 5.** Schematic illustration of hydrogen bonds formed in the ILPR G4-TEGPy complex (see text for description of individual interactions (a–c)). The guanine bases participating in the hydrogen bonding with TEGPy and TEGPy are represented as licorice, whereas the remaining part of ILPR G4 DNA is shown in new ribbon. All distances presented in this figure are labeled with dashed lines and given in Å. See Figure S5 for the numbering scheme of the O atoms.

producing beta-cells from donor pancreases into the liver of the diabetic patient [34]. Because there is a drastic shortage of donor pancreas material, it is imperative to find an alternative source of beta-cells that can be transplanted. To date, nobody has managed to produce a replenishable source of beta-cells that are suitable as a therapy for diabetes, although there have been successful protocols established that can create beta-cells in vitro [35,36]. Stem cells could offer an ideal solution to this problem. One approach that could enhance stem cell differentiation is the use of novel molecules, such as ILPR ligands, which can alter the expression of important genes such as insulin. With the ILPR sequence being unique in the human genome, it is a promising target for creating a ligand that could affect insulin gene expression and potentially offer a pathway to novel therapeutics and diabetes therapies. The uniqueness of the ILPR sequence also offers the opportunity for small molecule delivery of therapeutics that will only target the beta-cells and no other cells in the body. Any small molecule that is able to bind to the ILPR DNA sequence would be an enabling technology for the following applications: (1) the delivery of anti-apoptosis drugs to the beta-cells: the breakdown (apoptosis) of the beta-cells is a major causative factor in diabetes, so stopping this process early could offer a preventative treatment; (2) the delivery of beta-cell stimulating factors: therapeutic agents could be delivered to the beta-cells to stimulate them to multiply—this could alleviate diabetes symptoms in patients who still have a small number of beta-cells remaining in their pancreas; and (3) imaging and identification of beta-cells: whilst not strictly therapeutic, the non-invasive ability to identify and image the beta-cells in patients would be a significant aid in the clinical management of diabetes.

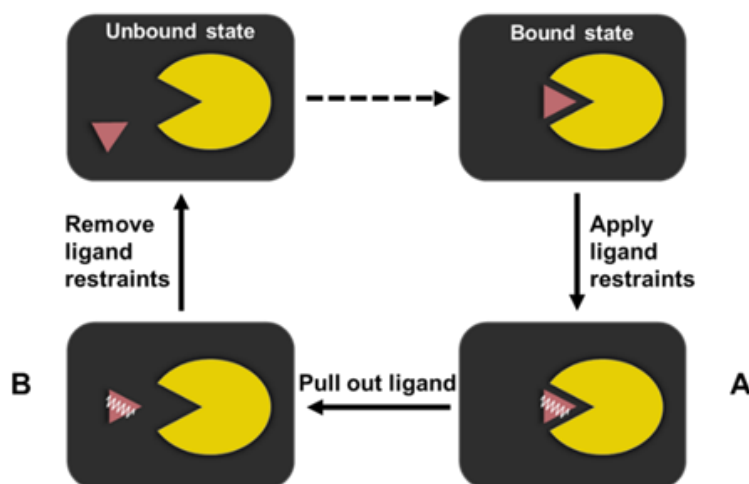
### 3. Materials and Methods

Due to the lack of a crystal structure of the ILPR G4, a related DNA fragment with a parallel conformation (Protein Data Bank entry 3CDM) was modified to provide the initial DNA model. Using the monomeric conformation of 3CDM, we mutated its sequence from [d(TAGGGTTA(GGG)<sub>3</sub>)<sub>2</sub>] to [d(TAGGGTGTGGG)<sub>2</sub>] (the monomeric sequence of the ILPR that was found to be most likely to form G4s and is associated with the high transcriptional activity of insulin) [37,38], using Chimera [39]. The experimentally refined 3CDM structure is not of the best quality but should be sufficient for our molecular modeling study (which involves structure minimizations and MD simulations). Both ligands were bound to the 3' ends of the ILPR G4 (the natural choice for TEGPy because the loop ends at the 5' ends are expected to interfere with the binding of the TEGPy side chains to the DNA).

We used nucleic acids parameters from ff94 [40] as implemented in Gromacs [41,42] for the ILPR-G4 ligand systems in solution. There have been some debates about the best force field to use for DNA [43–46]; the force field we are using has been validated for G4 systems in current research [47–49], as well as in the present work (see ESI and Figure S6 for further validation). GAFF2 [50] was used to model the ligands. The partial charges of the ligands were obtained by the AM1-BCC charge model [51]. The starting structure of the ILPR G4-ligand systems in the unconstrained MD simulation was fully optimized in the gas phase at the GFN2-xTB level of semiempirical density functional theory [52]. Each ILPR G4-ligand complex was initially placed in the center of a cubic periodic box and saturated with TIP4P water molecules. The distance between the outermost solute atoms and the nearest face of each box was set to 10 Å. The box size for the ILPR G4-TMPyP4 and the ILPR G4-TEGPy systems is 63 Å and 70 Å, respectively. Considering the negatively charged phosphate backbone of DNA, 16 sodium ions were added to the solvent to neutralize the system. The particle mesh Ewald (PME) method was used, with a cut-off of 10 Å, to describe the long-range electrostatic interactions of each system. Initially, 5000 energy minimization steps using the steepest-descent algorithm were conducted, followed by 100 ps pre-equilibration in the NVT ensemble followed by 100 ps in the NPT ensemble, respectively. Subsequently, a 5 ns production run was performed on each system at 300 K.



An RMSF-based PMF protocol employing the thermodynamic cycle of the ligand dissociation was used to compute the absolute free binding energy of the ILPR G4-ligand systems (Figure 6). This is based on the protocol originally proposed by Karplus and co-workers [53] and adapted by Deng et al. [29]. The bound state of the ligand systems is one equilibrated structure from the previously unconstrained MD simulation.

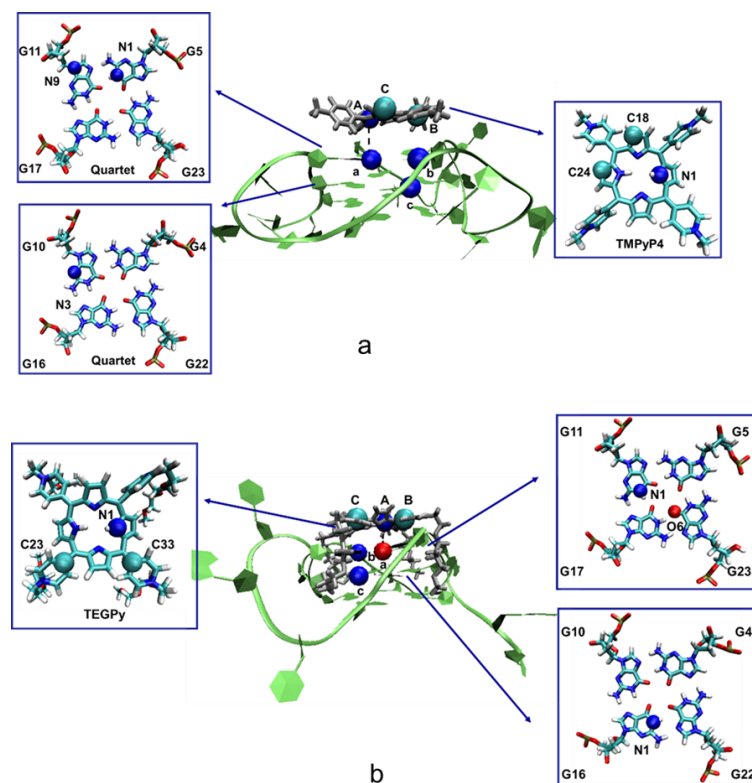


**Figure 6.** The RMSF-based PMF protocol based on the thermodynamic cycle adopted by the ILPR G4-ligand system. The ILPR G4 and its binding ligand are represented by the yellow Pac-man and purple triangle, respectively. The harmonic restraints applied to the ligand are indicated by a white spring.

Starting from the bound state, the orientation and position restraints of the binding ligand were gradually switched on (state A) via a series of  $\lambda$  values ranging from 0 to 1 (0 represents no restraints, whereas 1 denotes full restraints). The restraint atoms are selected by the evaluation of the all-heavy-atoms RMSF with respect to the bound state structure. Atoms that are expected to be in a more rigid region, but have larger computed RMSF, are considered good candidates for restraining the ligand orientation (Figures S1 and S2), because using these atoms larger fluctuations of the ligand (which would need long simulation times for correct sampling) can be suppressed. The appropriate choice of restraint atoms is key for applying the chosen protocol to this type of ligand (Figure 7). For the TMPyP4-ILPR G4 system, 16  $\lambda$  values were used in addition to 0 (0.01, 0.015, 0.05, 0.075, 0.10, 0.15, 0.20, 0.25, 0.35, 0.45, 0.50, 0.60, 0.70, 0.80, 0.90, and 1.00). These values are selected with reference to Deng's work, where 14  $\lambda$  values were used for quindoline (a ligand with a similar size to TMPyP4) [29]. As TEGPy is larger and more flexible than TMPyP4, more  $\lambda$  values were needed. Thus, for the ILPR G4-TEGPy system, 25  $\lambda$  values are used in addition to 0 (0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.12, 0.14, 0.16, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.60, 0.70, 0.80, 0.90, and 1.00). For each  $\lambda$  window, 5 ns equilibration at 300 K was performed, followed by a 5 ns production run. The difference in free energy upon switching on the restraints from the bound state to state A can then be estimated using the Bennett Acceptance Ratio (BAR), a procedure implemented in Gromacs [54], which also furnishes an estimated error based on the fluctuation of  $\partial H/\partial \lambda$  (for  $\Delta G_{restr}^{bound}$  in Equation (1)).

From state A to B, umbrella sampling along the reaction coordinate  $R_{aA}$  (Figure 7) was applied in both systems, with the ligand restraints kept at their full strength. For both ligand systems, there are 20 individual sampling windows, in which different values along  $R_{aA}$  were set. The initial conformation of each  $R_{aA}$  was generated by gradually pulling the ligand away from its binding target using the PULL code [55]. Conformations were saved every 1 Å. At each window, different values of  $R_{aA}$  were restrained with a force constant of 1000 kJ mol<sup>-1</sup>·nm<sup>-2</sup> for an initial 5 ns equilibration phase, to allow the simulations to adjust to the restraints. Subsequently, a 20 ns production run was conducted at 300 K.

An illustrative example showing the convergence of the mean force is shown in Figure S8 in the ESI. The PMF of the ligand-ILPR G4 systems was analyzed using the Weighted Histogram Analysis Method (WHAM) [56] as included in Gromacs. The statistical fluctuation of the free-energy change ( $\omega(r)$  in Equation (1)) was evaluated twice for each ligand, namely by the standard deviation of repeating two 10 ns trajectories and by that of repeating four 5 ns trajectories. The largest of the resulting standard deviations was used as the final error estimate (see Figure S8 for details).



**Figure 7.** Schematic illustration of the restraint atoms in the ILPR G4-ligand systems. **(a)** Restraint atoms in the ILPR G4-TMPyP4 complex. Atoms A (N1), B (C24) and C (C18) are located close to the aromatic core of TMPyP4, whereas a (N1, G5), b (N9, G11), and c (N3, G10) are in the first and second layer of the ILPR G4 structure. **(b)** Restraint atoms in the ILPR G4-TEGPy complex. Atoms a, b, and c and A, B, and C are located at the aromatic core of TEGPy and the quartets of the ILPR G4 structure, respectively. G stands for the guanine base. Note that restraining atoms with unfavorable RMSFs can lead to insufficient restraint of the ligand orientation, resulting in convergence problems in the constrained MD simulations (see Figure S9 and Table S1).

When the binding ligand escaped from the ILPR G4 and reached the bulk region (state B), all the restraints were switched off. The ligands can freely rotate in the solvent, and therefore, the free-energy change of the restraints in the solution can be estimated by the final term in Equation (1) (from state B to the unbound state). The absolute binding free energy of the ILPR G4-TMPyP4 and ILPR G4-TEGPy systems is computed by [29]

$$\Delta G_{bind}^{\circ} = -\Delta G_{restr}^{bound} - w(r^*) - k_B T \ln \frac{\int_{bound} e^{-w(r)k_B T} dr}{\left(\frac{2\pi k_B T}{k_r}\right)^{\frac{1}{2}}} + \Delta G_{restr}^{bulk} \quad (1)$$

$$\Delta G_{restr}^{bound} = k_B T \ln \frac{\langle f(H_A(p, q) - H_B(p, q) + C) \rangle_B}{\langle f(H_B(p, q) - H_B(p, q) - C) \rangle_A} + C \quad (2)$$

$$\Delta G_{restr}^{bulk} = -k_B T \ln \frac{C^0 r^{*2} \sin \theta_0 \sin \theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_{\theta} k_{\phi} k_{\psi})^{\frac{1}{2}}} \quad (3)$$

where the free energy of restraining the ligand at the bound state is expressed as  $-\Delta G_{restr}^{bound}$ . This can be estimated using BAR as shown in Equation (2), where  $f$  is the Fermi function  $f(x) = 1/(1 + \exp(x/k_B T))$ , and  $C = k_B T \ln \frac{Q_A n_B}{Q_B n_A}$ ,  $n_A$  and  $n_B$  are the numbers of configurations generated in the states A (the bound state without restraints) and B (the bound state with restraints), respectively.  $p$  and  $q$  represent the coordinates and momenta of the Hamiltonian. The potential of the mean force of the dissociation of the ligand is determined by  $w(r)$ , where  $r$  is the current location of the ligand and  $r^*$  is the location where the ligand reaches the bulk. This is to some extent ambiguous as long as the ligand is fully dissociated from its G4 target. For the ILPR G4-TMPyP4 and ILPR G4-TEGPy systems,  $r^*$  is taken to be 20 and 22.5 Å, respectively (the position where the PMF has levelled off, see Figure 3).  $k_r$  is the force constant of the harmonic restraint with a value of 1000 kJ·mol<sup>-1</sup>·nm<sup>-2</sup>. The free-energy change of restraining the ligand in the solvent is represented by  $\Delta G_{restr}^{bulk}$ , where  $k_\theta, k_\phi, k_\psi$  and  $k_\psi$  are the force constants for restraining the ligand orientation and position in both ILPR G4-ligand systems.  $k_\theta = k_\phi = k_\psi = 1000$  kJ·rad<sup>-1</sup>·mol<sup>-1</sup>. These restraint values are the average values obtained from the previous unconstrained MD simulations. The uncertainty of the third term in Equation (1) (starting with  $-k_B T$ ) was estimated from the values of the integral at the upper and lower bounds of  $r$ , according to the procedure by Deng et al. [29].

#### 4. Conclusions

We studied the absolute binding free energy and ligand interactions of the ILPR G4-ligand systems using a PMF-based protocol. Previous computational studies have demonstrated the usefulness of free-energy simulations in characterizing the thermodynamics of ligand-G4 binding [29–31]. However, most of the studies have focused on relatively small ligand systems and there has been limited research into ligands such as TEGPy that are large, highly flexible, and display multiple binding modes simultaneously towards G4s. This is because the flexibility and large size of such ligands increase the conformational sampling spaces, require extensive computational resources, and may result in convergence problems in the free-energy simulations. In this study, we found that TEGPy displays a rigid core and flexible side chains. A poor choice of restraint atoms in restraining such ligands can lead to improper free-energy perturbations. To overcome this issue, we have improved an RMSF-based PMF protocol to assist in selecting restraint atoms that confine the unnecessary fluctuations of atoms in the rigid region caused by the flexible part of the ligand, while maintaining the flexibility of the ligand. This novel protocol was validated for the ILPR G4 system by the excellent agreement of the computed absolute free binding energy of TMPyP4 ( $-9.18 \pm 0.48$  kcal·mol<sup>-1</sup>) with that obtained by the ITC experiment ( $-9.2$  kcal·mol<sup>-1</sup>). Using this approach, we rationally predict the absolute free binding energy ( $-11.50 \pm 0.59$  kcal·mol<sup>-1</sup>) and binding modes of the flexible and large modular ligand, TEGPy, in the ILPR G4 system. The results show that TEGPy displays a higher ligand affinity towards the ILPR G4 (ca. 50-fold increase in the binding constant at room temperature), due to its side chains interacting with the loop and groove regions through hydrogen bonding. However, in contrast to the binding mode proposed previously, we found that only some of the four side arms fit into the groove regions simultaneously. This finding suggests that ligands with fewer side arms might be expected to bind just as well as, or even better than, ligands with four side arms. This possibility provides a new perspective for designing ILPR G4 ligands and may assist in interpreting the important role that G4 formation plays in the regulation of the insulin gene.

Classical simulations are routinely used to study ligand-receptor binding, usually involving fairly small ligands (typically ca. <100 atoms, as opposed to the 174 atoms of TEGPy) Application of these standard tools to large, highly charged and, in the case of TEGPy, highly flexible ligands, poses a substantial challenge. The successful demonstration that a more refined methodology from the literature can be applied to such ligands and G4 structures is a significant development, opening the avenue for further ligand

design in this important field. The present research has helped to expand the limits on the complexity of receptors and ligands, further enhancing the applicability and usefulness of biomolecular simulations.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/molecules28083447/s1> (additional graphical and tabular material: Additional simulation details are shown in Figures S1 and S2; Additional simulation results are shown in Figures S3, S4 and S5; Figure S6: Force fields comparison between ff94 and parmbsc1; Figure S7: Convergence of mean constrained force; Error estimation for  $\omega(r^*)$  over repeating four trajectories are shown in Figures S8 and S9; Table S1: Computed free energy difference in restraining the ligand at the bound state of the ILPR G4-TEGPy system); pubXiaotong-G4-SI-\*.pdb (files with coordinates of the full systems with ligands bound).

**Author Contributions:** X.Z. original draft preparation and all simulations; J.B. review and editing; T.v.M. and M.B. supervision. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The research data supporting this publication can be accessed at <https://doi.org/10.17630/9505b210-dfec-401c-b32d-9151ae3810e3> (accessed on 1 April 2023)

**Acknowledgments:** X.Z. wants to thank the Chinese Scholarship Council for funding. We thank EaStCHEM and the School of Chemistry for their support. Computations were carried out on the St Andrews HPC cluster and computer clusters in the School of Chemistry.

**Conflicts of Interest:** There are no conflicts to declare.

## References

1. Neidle, S. Quadruplex nucleic acids as targets for anticancer therapeutics. *Nat. Rev. Chem.* **2017**, *1*, 41. <https://doi.org/10.1038/s41570-017-0041>.
2. Wong, H.; Payet, L.; Huppert, J. Function and Targeting of G-Quadruplexes. *Curr. Opin. Mol. Ther.* **2009**, *11*, 146–155.
3. Bochman, M.L.; Paeschke, K.; Zakian, V.A. DNA secondary structures: Stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **2012**, *13*, 770–780. <https://doi.org/10.1038/nrg3296>.
4. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474. <https://doi.org/10.1038/s41580-020-0236-x>.
5. Sun, D.; Hurley, L.H. The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **2009**, *52*, 2863–2874. <https://doi.org/10.1021/jm900055s>.
6. Keith, W.N.; Evans, T.R.J.; Glasspool, R.M. Telomerase and cancer: Time to move from a promising target to a clinical reality. *J. Pathol.* **2001**, *195*, 404–414. <https://doi.org/10.1002/path.1001>.
7. Huppert, J.L.; Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **2006**, *35*, 406–413. <https://doi.org/10.1093/nar/gkl1057>.
8. Spiegel, J.; Cuesta, S.M.; Adhikari, S.; Hänsel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **2021**, *22*, 1–15. <https://doi.org/10.1186/s13059-021-02324-z>.
9. Tian, T.; Chen, Y.-Q.; Wang, S.-R.; Zhou, X. G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem* **2018**, *4*, 1314–1344. <https://doi.org/10.1016/j.chempr.2018.02.014>.
10. Fedeles, B.I. G-quadruplex-forming promoter sequences enable transcriptional activation in response to oxidative stress. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2788–2790. <https://doi.org/10.1073/pnas.1701244114>.
11. Lew, A.; Rutter, W.J.; Kennedy, G.C. Unusual DNA structure of the diabetes susceptibility locus *IDDM2* and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12508–12512. <https://doi.org/10.1073/pnas.97.23.12508>.
12. Hammond-Kosack, M.C.U.; Dobrinski, B.; Lurz, R.; Docherty, K.; Kilpatrick, M.W. The human insulin gene linked polymorphic region exhibits an altered DNA structure. *Nucleic Acids Res.* **1992**, *20*, 231–236. <https://doi.org/10.1093/nar/20.2.231>.
13. Kennedy, G.C.; Rutter, W.J. Pur-1, a zinc-finger protein that binds to purine-rich sequences, transactivates an insulin promoter in heterologous cells. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 11498–11502. <https://doi.org/10.1073/pnas.89.23.11498>.
14. Bell, G.I.; Karam, J.H.; Rutter, W.J. Polymorphic DNA region adjacent to the 5' end of the human insulin gene. *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 5759–5763. <https://doi.org/10.1073/pnas.78.9.5759>.
15. Connor, A.C.; Frederick, K.A.; Morgan, E.J.; McGown, L.B. Insulin Capture by an Insulin-Linked Polymorphic Region G-Quadruplex DNA Oligonucleotide. *J. Am. Chem. Soc.* **2006**, *128*, 4986–4991. <https://doi.org/10.1021/ja056097c>.

16. Schonhoft, J.D.; Das, A.; Achamyeh, F.; Samdani, S.; Sewell, A.; Mao, H.; Basu, S. ILPR repeats adopt diverse G-quadruplex conformations that determine insulin binding. *Biopolymers* **2009**, *93*, 21–31. <https://doi.org/10.1002/bip.21289>.
17. Yu, Z.; Schonhoft, J.D.; Dhakal, S.; Bajracharya, R.; Hegde, R.; Basu, S.; Mao, H. ILPR G-Quadruplexes Formed in Seconds Demonstrate High Mechanical Stabilities. *J. Am. Chem. Soc.* **2009**, *131*, 1876–1882. <https://doi.org/10.1021/ja806782s>.
18. Liu, H.-Y.; Zhao, Q.; Zhang, T.-P.; Wu, Y.; Xiong, Y.-X.; Wang, S.-K.; Ge, Y.-L.; He, J.-H.; Lv, P.; Ou, T.-M.; et al. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem. Biol.* **2016**, *23*, 1261–1270. <https://doi.org/10.1016/j.chembiol.2016.08.013>.
19. Moye, A.L.; Porter, K.C.; Cohen, S.B.; Phan, T.; Zyner, K.G.; Sasaki, N.; Lovrecz, G.O.; Beck, J.L.; Bryan, T.M. Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat. Commun.* **2015**, *6*, 7643. <https://doi.org/10.1038/ncomms8643>.
20. Seenisamy, J.; Rezler, E.M.; Powell, T.J.; Tye, D.; Gokhale, V.; Joshi, C.S.; Siddiqui-Jain, A.; Hurley, L.H. The Dynamic Character of the G-Quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *J. Am. Chem. Soc.* **2004**, *126*, 8702–8709. <https://doi.org/10.1021/ja040022b>.
21. Martino, L.; Pagano, B.; Fotticchia, I.; Neidle, S.; Giancola, C. Shedding Light on the Interaction between TMPyP4 and Human Telomeric Quadruplexes. *J. Phys. Chem. B* **2009**, *113*, 14779–14786. <https://doi.org/10.1021/jp9066394>.
22. Gao, Y.; Guang, T.; Ye, X. Sedimentation velocity analysis of TMPyP4-induced dimer formation of human telomeric G-quadruplex. *RSC Adv.* **2017**, *7*, 55098–55105. <https://doi.org/10.1039/c7ra07758k>.
23. Ali, A.; Bansal, M.; Bhattacharya, S. Ligand 5,10,15,20-Tetra(*N*-methyl-4-pyridyl)porphine (TMPyP4) Prefers the Parallel Propeller-Type Human Telomeric G-Quadruplex DNA over Its Other Polymorphs. *J. Phys. Chem. B* **2014**, *119*, 5–14. <https://doi.org/10.1021/jp505792z>.
24. Huang, X.-X.; Zhu, L.-N.; Wu, B.; Huo, Y.-F.; Duan, N.-N.; Kong, D.-M. Two cationic porphyrin isomers showing different multimeric G-quadruplex recognition specificity against monomeric G-quadruplexes. *Nucleic Acids Res.* **2014**, *42*, 8719–8731. <https://doi.org/10.1093/nar/gku526>.
25. Laguerre, A.; Chang, Y.; Pirrotta, M.; Desbois, N.; Gros, C.P.; Lesniewska, E.; Monchaud, D. Surface-promoted aggregation of amphiphilic quadruplex ligands drives their selectivity for alternative DNA structures. *Org. Biomol. Chem.* **2015**, *13*, 7034–7039. <https://doi.org/10.1039/c5ob00692a>.
26. Diabate, P.D.; Laguerre, A.; Pirrotta, M.; Desbois, N.; Boudon, J.; Gros, C.P.; Monchaud, D. DNA structure-specific sensitization of a metalloporphyrin leads to an efficient in vitro quadruplex detection molecular tool. *New J. Chem.* **2016**, *40*, 5683–5689. <https://doi.org/10.1039/C6NJ01012A>.
27. Myszka, D.G. Survey of the 1998 Optical Biosensor Literature. *J. Mol. Recognit.* **1999**, *12*, 390–408. [https://doi.org/10.1002/\(SICI\)1099-1352\(199911/12\)12:6<390::AID-JMR482>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1099-1352(199911/12)12:6<390::AID-JMR482>3.0.CO;2-8).
28. Dutta, A.K.; Rösgen, J.; Rajarathnam, K. Using Isothermal Titration Calorimetry to Determine Thermodynamic Parameters of Protein–Glycosaminoglycan Interactions. *Methods Mol. Biol.* **2015**, *1229*, 315–324. [https://doi.org/10.1007/978-1-4939-1714-3\\_25](https://doi.org/10.1007/978-1-4939-1714-3_25).
29. Deng, N.; Wickstrom, L.; Cieplak, P.; Lin, C.; Yang, D. Resolving the Ligand-Binding Specificity in c-MYC G-Quadruplex DNA: Absolute Binding Free Energy Calculations and SPR Experiment. *J. Phys. Chem. B* **2017**, *121*, 10484–10497. <https://doi.org/10.1021/acs.jpcc.7b09406>.
30. Moraca, F.; Amato, J.; Ortuso, F.; Artese, A.; Pagano, B.; Novellino, E.; Alcaro, S.; Parrinello, M.; Limongelli, V. Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E2136–E2145. <https://doi.org/10.1073/pnas.1612627114>.
31. O'Hagan, M.P.; Haldar, S.; Morales, J.C.; Mulholland, A.J.; Galan, M.C. Enhanced sampling molecular dynamics simulations correctly predict the diverse activities of a series of stiff-stilbene G-quadruplex DNA ligands. *Chem. Sci.* **2020**, *12*, 1415–1426. <https://doi.org/10.1039/d0sc05223j>.
32. Freyer, M.W.; Buscaglia, R.; Kaplan, K.; Cashman, D.; Hurley, L.H.; Lewis, E.A. Biophysical Studies of the c-MYC NHE III1 Promoter: Model Quadruplex Interactions with a Cationic Porphyrin. *Biophys. J.* **2007**, *92*, 2007–2015. <https://doi.org/10.1529/biophysj.106.097246>.
33. Clovis, N.S.; Sen, S. G-Tetrad-Selective Ligand Binding Kinetics in G-Quadruplex DNA Probed with Fluorescence Correlation Spectroscopy. *J. Phys. Chem. B* **2022**, *126*, 6007–6015. <https://doi.org/10.1021/acs.jpcc.2c04181>.
34. Shapiro, A.J.; Lakey, J.R.; Ryan, E.A.; Korbitt, G.S.; Toth, E.; Warnock, G.L.; Kneteman, N.M.; Rajotte, R.V. Islet Transplantation in Seven Patients with Type 1 Diabetes Mellitus Using a Glucocorticoid-Free Immunosuppressive Regimen. *New Engl. J. Med.* **2000**, *343*, 230–238. <https://doi.org/10.1056/nejm200007273430401>.
35. Pagliuca, F.W.; Millman, J.R.; Gürtler, M.; Segel, M.; Van Dervort, A.; Ryu, J.H.; Peterson, Q.P.; Greiner, D.; Melton, D.A. Generation of Functional Human Pancreatic  $\beta$  Cells In Vitro. *Cell* **2014**, *159*, 428–439. <https://doi.org/10.1016/j.cell.2014.09.040>.
36. Rezanian, A.; Bruin, J.E.; Arora, P.; Rubin, A.; Batushansky, I.; Asadi, A.; O'Dwyer, S.; Quiskamp, N.; Mojibian, M.; Albrecht, T.; et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **2014**, *32*, 1121–1133. <https://doi.org/10.1038/nbt.3033>.
37. Paritala, H.; Firestone, S.M. Characterization of Insulin Ilpr Sequences for Their Ability to Adopt a G-Quadruplex Structure. *Nucleosides Nucleotides Nucleic Acids* **2010**, *29*, 81–90. <https://doi.org/10.1080/15257771003597691>.
38. Hammond-Kosack, M.C.U.; Kilpatrick, M.W.; Docherty, K. The human insulin gene-linked polymorphic region adopts a G-quartet structure in chromatin assembled in vitro. *J. Mol. Endocrinol.* **1993**, *10*, 121–126. <https://doi.org/10.1677/jme.0.0100121>.

39. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera? A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. <https://doi.org/10.1002/jcc.20084>.
40. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197. <https://doi.org/10.1021/ja00124a002>.
41. Lindahl, E.; Hess, B.; Van Der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317. <https://doi.org/10.1007/s008940100045>.
42. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. <https://doi.org/10.1021/ct700301q>.
43. Perez, A.; Marchán, I.; Svoboda, D.; Sponer, J.; Cheatham, T.E., III; Laughton, C.A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of  $\alpha/\gamma$  Conformers. *Biophys. J.* **2007**, *92*, 3817–3829. <https://doi.org/10.1529/biophysj.106.097782>.
44. Haider, S.; Neidle, S. Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes. In *G-Quadruplex DNA*; Baumann, P., Ed.; Methods in Molecular Biology; Humana Press: Totowa, NJ, USA, 2010; Volume 608, pp. 17–37, ISBN 978-1-58829-950-5.
45. Galindo-Murillo, R.; Robertson, J.C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T.E. Assessing the Current State of Amber Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127. <https://doi.org/10.1021/acs.jctc.6b00186>.
46. Ivani, I.; Dans, P.D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: A refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55–58. <https://doi.org/10.1038/nmeth.3658>.
47. Hou, X.-M.; Fu, Y.-B.; Wu, W.-Q.; Wang, L.; Teng, F.-Y.; Xie, P.; Wang, P.-Y.; Xi, X.-G. Involvement of G-triplex and G-hairpin in the multi-pathway folding of human telomeric G-quadruplex. *Nucleic Acids Res.* **2017**, *45*, 11401–11412. <https://doi.org/10.1093/nar/gkx766>.
48. Zeng, L.; Wu, Q.; Wang, T.; Li, L.-P.; Zhao, X.; Chen, K.; Qian, J.; Yuan, L.; Xu, H.; Mei, W.-J. Selective stabilization of multiple promoter G-quadruplex DNA by using 2-phenyl-1H-imidazole-based tanshinone IIA derivatives and their potential suppressing function in the metastatic breast cancer. *Bioorganic Chem.* **2020**, *106*, 104433. <https://doi.org/10.1016/j.bioorg.2020.104433>.
49. Pandith, A.; Nagarajachari, U.; Siddappa, R.K.G.; Lee, S.; Park, C.; Sannathammegowda, K.; Seo, Y.J. Loop-mediated fluorescent probes for selective discrimination of parallel and antiparallel G-Quadruplexes. *Bioorganic Med. Chem.* **2021**, *35*, 116077. <https://doi.org/10.1016/j.bmc.2021.116077>.
50. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174. <https://doi.org/10.1002/jcc.20035>.
51. Jakalian, A.; Jack, D.B.; Bayly, C.I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641. <https://doi.org/10.1002/jcc.10128>.
52. Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671. <https://doi.org/10.1021/acs.jctc.8b01176>.
53. Pettitt, B.M.; Karplus, M. The potential of mean force surface for the alanine dipeptide in aqueous solution: A theoretical approach. *Chem. Phys. Lett.* **1985**, *121*, 194–201. [https://doi.org/10.1016/0009-2614\(85\)85509-3](https://doi.org/10.1016/0009-2614(85)85509-3).
54. Bennett, C.H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245–268. [https://doi.org/10.1016/0021-9991\(76\)90078-4](https://doi.org/10.1016/0021-9991(76)90078-4).
55. Engin, O.; Villa, A.; Sayar, M.; Hess, B. Driving Forces for Adsorption of Amphiphilic Peptides to the Air–Water Interface. *J. Phys. Chem. B* **2010**, *114*, 11093–11101. <https://doi.org/10.1021/jp1024922>.
56. Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.; Kollman, P.A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021. <https://doi.org/10.1002/jcc.540130812>.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.