Supporting Information for "Resolving the Ligand Binding Specificity in c-MYC G-quadruplex DNA: Absolute Binding Free Energy Calculations and SPR Experiment"

Nanjie Deng,[1] Lauren Wickstrom,[2] Piotr Cieplak,[3] Clement Lin,[4] and Danzhou Yang[4]**

(1) Department of Chemistry and Physical Sciences, Pace University, 1 Pace Plaza, New York, NY 10038

(2) Borough of Manhattan Community College, the City University of New York, Department of Science, New York, NY 10007

(3) Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA 92037

(4) Department of Medicinal Chemistry and Molecular Pharmacology, College of Pharmacy, Purdue University, West Lafayette, IN 47907

Email: nanjie.deng@gmail.com, telephone: (212) 346-1592; yangdz@purdue.edu, telephone: (765) 494-814

Derivation of a new PMF expression of absolute binding free energy

We consider the calculation of absolute binding free energy between ligand *L* and receptor *R* as illustrated in Fig. S1. To derive the expression of absolute binding free energy, we start from the statistical mechanical expression for $\Delta G^{\circ}_{\mathrm{bind}}$ [1–3]

$$\Delta G^{\circ}_{\mathrm{bind}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N} Z_{0,N}}{Z_{R,N} Z_{L,N}} \quad (1)$$

Here $V_0 = \frac{1}{C^o} = 1660\ \text{Å}^2$ is the inverse of the standard concentration of 1M solution. $Z_{X,N}$ is the configuration integral ($Z = \int e^{-U(r)/k_B T} dr$) of a single solute species X solvated in a box of volume *V* containing *N* water molecules. Thus, $Z_{RL,N}$ represents a system containing one receptor-ligand complex in the bound state solvated by *N* water molecules. $Z_{0,N}$ represents a system of pure solvent. If the solvent box *V* is sufficiently large, we can rewrite the denominator in the logarithm of the right hand side of Eq. (1) as

$$\Delta G^{\circ}_{\mathrm{bind}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N} Z_{0,N}}{Z_{R,N} Z_{L,N}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N} Z_{0,N}}{Z_{R+L,N} Z_{0,N}} \quad (2)$$

Here $Z_{P+L,N}$ represents the configuration integral of volume *V* containing the receptor *R*, the unbound ligand *L* and *N* waters. The transformation in Eq. (2) corresponds to transferring *L* from its own solvent box to the solvent box that contains *R*. Note that *R* and *L* remains unbound in the final solvent box. Thus, Eq. (2) becomes

$$\Delta G^{\circ}_{\mathrm{bind}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N} Z_{0,N}}{Z_{R,N} Z_{L,N}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N}}{Z_{R+L,N}} \quad (3)$$

The evaluation of the right hand side of Eq. (3) is made possible by inserting the following intermediate states

$$\Delta G^{\circ}_{\mathrm{bind}} = -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N}}{Z_{R+L,N}}$$

$$= -k_B T \ln \frac{V}{V_0} \frac{Z_{RL,N}}{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}} \frac{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}} \frac{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L,N}} \quad (4)$$

Here $Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}$ represents a bound complex *RL* in which the ligand *L*'s external degrees of freedom, which are defined by the polar angles (θ, ϕ) and three Euler angles (Θ, Φ, Ψ) (Fig. S1), are restrained to their equilibrium values in the bound state by a set of harmonic restraints:

$U_\theta = \frac{1}{2}k_\theta(\theta-\theta_0)^2, U_\phi = \frac{1}{2}k_\phi(\phi-\phi_0)^2, U_\Theta = \frac{1}{2}k_\Theta(\Theta-\Theta_0)^2, U_\Phi = \frac{1}{2}k_\Phi(\Phi-\Phi_0)^2, and\ U_\Psi = \frac{1}{2}k_\Psi(\Psi-\Psi_0)^2$. Fig. S1 gives the definition of the polar angles (θ, ϕ) and three Euler angles (Θ, Φ, Ψ).

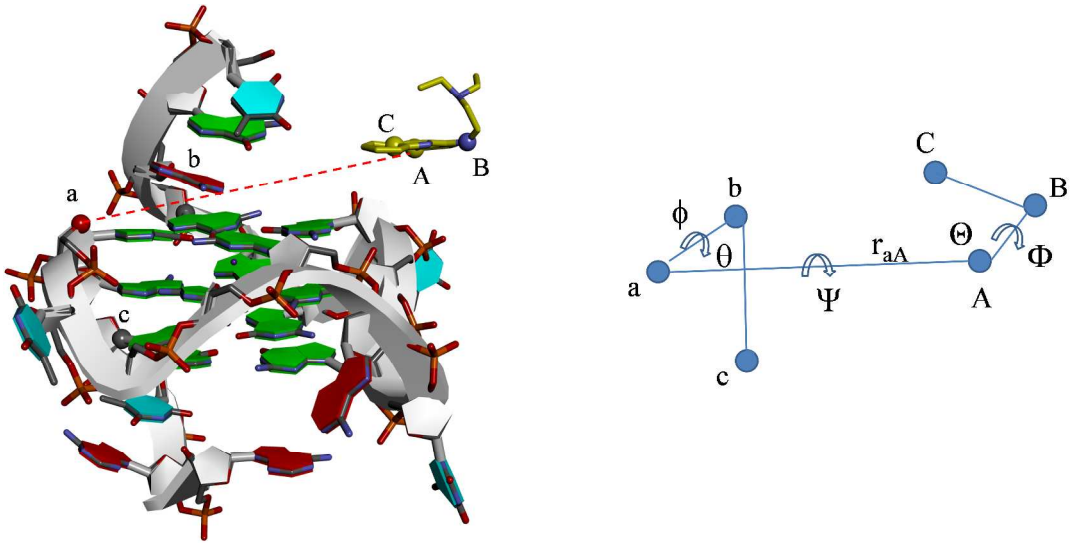

Figure S1. The coordinate frame in which the orientation and the position of the ligand relative to the DNA receptor are defined, where the atoms a,b,c and A,B,C belong to the receptor and ligand, respectively. The ligand position is defined in the polar coordinates by the distance r_{aA} and two angles θ: b-a-A;ϕ: c-b-a-A. The ligand orientation is defined by the three Euler angles: Θ: a-A-B; Φ: a-A-B-C; Ψ: b-a-A-B.

$Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}$ represents a system in which the ligand L is not only subject to the polar and orientational restraints $(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$, but also subject to the harmonic restraint $U_{r^*} = \frac{1}{2}k_r(r-r^*)^2$, which forces the ligand atom A to be close to a bulk location r^*.

The different terms inside the logarithm of the right hand side of Eq. (4) can be recognized as corresponding to the following thermodynamic transformations:

The first term inside the logarithm in the right hand side of Eq. (4) corresponds to the free energy of switching on the angular restraints $(U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$ on the ligand in the bound state, i.e.

$$\frac{Z_{RL,N}}{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}} = e^{\Delta G_{restr}^{bound}/k_BT} \qquad (5.1)$$

which can be computed using free energy perturbation (FEP) on the bound complex. In this work, this is done by using 24 λ windows to switch on the polar angular and orientational

restraints ($U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi$): 12 λ windows (λ = 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.2, 0.35, 0.5, 0.65, 0.8, 1.0) are used to switch on the polar angle restraints (U_θ, U_ϕ), followed by another 12 λ for the ligand orientational restraints (U_Θ, U_Φ, U_Ψ). The Bennett acceptance ratio (BAR)[4] is used to compute the free energy change in Eq. (5.1).

The second term inside the logarithm of the right hand side of Eq. (4), $\frac{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}}$, equals to the ratio of the probability of an angularly restrained bound complex and the probability of a unbound, but "cross linked" receptor-ligand "complex". The cross-linking is maintained by the distance and angular restraints $\left(U_{r^*}, U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi\right)$. To evaluate the ratio $\frac{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}}$ in the Eq. (4), we make use of the potential of mean force (PMF). Let $w(r)$ be the 1D PMF along the r_{aA} axis. $w(r)$ can be computed using umbrella sampling simulation on the ligand in the presence of all the angular restraints ($U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi$): see below. Using the 1D PMF $w(r)$, we can write

$$\frac{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}} = \frac{\int_{bound} e^{-w(r)/k_BT} dr}{\int_{bulk} e^{-w(r)/k_BT} e^{-U_{r^*}/k_BT} dr} = \frac{\int_{bound} e^{-w(r)/k_BT} dr}{\int_{bulk} e^{-w(r)/k_BT} e^{-\frac{1}{2}k_r(r-r^*)^2/k_BT} dr} \qquad (5.2a)$$

When the force constant k_r is sufficiently large (stiff spring), the integrand in the denominator of the Eq. (5.2) is everywhere zero except within a very small region in the neighborhood of the bulk location r^*, and the right hand side of Eq. (5.2a) becomes

$$\frac{\int_{bound} e^{-w(r)/k_BT} dr}{\int_{bulk} e^{-w(r)/k_BT} e^{-\frac{1}{2}k_r(r-r^*)^2/k_BT} dr} = \frac{\int_{bound} e^{-w(r)/k_BT} dr}{e^{-w(r^*)/k_BT} \int_{-\infty}^{\infty} e^{-\frac{1}{2}k_r(r-r^*)^2/k_BT} d(r-r^*)} =$$

$$\frac{\int_{bound} e^{-[w(r)-w(r^*)]/k_BT} dr}{\int_{-\infty}^{\infty} e^{-\frac{1}{2}k_r(\Delta r)^2/k_BT} d\Delta r} = \frac{\int_{bound} e^{-[w(r)-w(r^*)]/k_BT} dr}{(\frac{2\pi k_BT}{k_r})^{\frac{1}{2}}} \qquad (5.2b)$$

Substituting Eq. (5.2b) into Eq. (5.2a) yields

$$\frac{Z_{RL(\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}} = \frac{\int_{bound} e^{-[w(r)-w(r^*)]/k_BT} dr}{(\frac{2\pi k_BT}{k_r})^{\frac{1}{2}}} \qquad (5.2)$$

Finally, the last term inside the logarithm of the right hand side of Eq. (4), $\frac{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L,N}}$, represents the free energy of removing all the distance and angular restraints $(U_{r^*}, U_\theta, U_\phi, U_\Theta, U_\Phi, U_\Psi)$ on the bulk ligand such that the ligand is allowed to occupy the standard volume $1/C°$ and rotate freely. When the force constants used in the harmonic restraints are sufficiently strong, which is the case in this work, the rigid-rotor approximation holds, which allows this term to be evaluated analytically [2], i.e.

$$\frac{Z_{R+L(r^*,\theta,\phi,\Theta,\Phi,\Psi),N}}{Z_{R+L,N}} \approx \frac{r^{*2} sin\theta_0 sin\Theta_0 (2\pi k_B T)^3}{8\pi^2 V (k_r k_\theta k_\phi k_\Theta k_\Phi k_\Psi)^{\frac{1}{2}}} \qquad (5.3)$$

where θ_0 and Θ_0 are the equilibrium values of θ and Θ, respectively. $k_r k_\theta k_\phi k_\Theta k_\Phi k_\Psi$ in the denominator is the product of the force constants for the corresponding harmonic restraints on (r, θ, ϕ, Θ, Φ, Ψ) .

Combining the above Eq. (5) and Eq. (4), we obtain the final expression for the absolute binding free energy

$$\Delta G^\circ_{bind} = -\Delta G^{bound}_{restr} - w(r^*) - k_B T \ln \frac{\int_{bound} e^{-w(r)/k_B T} dr}{(2\pi k_B T/k_r)^{\frac{1}{2}}} - k_B T \ln \frac{C^o r^{*2} sin\theta_0 sin\Theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_\theta k_\phi k_\Theta k_\Phi k_\Psi)^{\frac{1}{2}}} \qquad (6)$$

The last term in the Eq. (6) is the free energy of turning on all the restraints on the ligand external degrees of freedom in the bulk solution. Thus, we can rewrite Eq. (6) as

$$\Delta G^\circ_{bind} = -\Delta G^{bound}_{restr} - w(r^*) - k_B T \ln \frac{\int_{bound} e^{-w(r)/k_B T} dr}{(2\pi k_B T/k_r)^{\frac{1}{2}}} + \Delta G^{bulk}_{restr} \qquad (7)$$

where $\Delta G^{bulk}_{restr} = -k_B T \ln \frac{C^o r^{*2} sin\theta_0 sin\Theta_0 (2\pi k_B T)^3}{8\pi^2 (k_r k_\theta k_\phi k_\Theta k_\Phi k_\Psi)^{\frac{1}{2}}}$

It should be noted that, the 1D-PMF w(r) implicitly contains the Jacobian factor that is proportional to $2\pi r^2$. To see this, we first note that in the presence of the ligand polar angle restraints (U_θ, U_ϕ), the ligand moves along an 1D axis inside a cone defined by the restraints (U_θ, U_ϕ), and whose cross-section surface area at position *r* is given by $2\pi r^2 \Delta\theta^2$, where $\Delta\theta^2$ is

the mean-square fluctuation in θ angle. $\Delta\theta^2$ is related to the angular force constant k_θ and the temperature factor by $\Delta\theta^2 = \frac{k_B T}{k_\theta}$. Thus, the 1D probability density P(r) actually can be written as a product of the surface area of the cone $2\pi r^2 \Delta\theta^2$ and a term $P'(r)$ that only depends on r, that is

$$P(r) = 2\pi r^2 \frac{k_B T}{k_\theta} P'(r) \qquad (7.1)$$

Since the PMF w(r) is related to the 1D probability density P(r) by $w(r) = -k_B T ln[P(r)] + C$, where C is a constant defined by set the zero of w(r) at the bound state to zero. Combine the above with Eq. (7.1)

$$w(r) = -k_B T ln\left[2\pi r^2 \frac{k_B T}{k_\theta} P'(r)\right] + C \qquad (7.2)$$

It is important to note that the implicit r^2-dependence in the w(r) is necessary to make the absolute binding free energy expression of the Eq. (6) independent of the choice of the bulk location r*. To see this, note that the last term in the right hand side of the Eq. (6) has an r^{*2} dependence. This r^{*2} dependence is cancelled out by the r^{*2}-dependence in the term -w(r*), which makes the overall expression of ΔG°_{bind} independent of the choice of r*.

As shown in the Table 3 of the main text, the integral term $-k_B T \ln \frac{\int_{bound} e^{-w(r)/k_B T} dr}{(2\pi k_B T/k_r)^{\frac{1}{2}}}$ in the Eq. (6) is very small for both 5' and 3' complexes. This means that the ratio $\frac{\int_{bound} e^{-w(r)/k_B T} dr}{(2\pi k_B T/k_r)^{\frac{1}{2}}}$ has a value close to one for both the complexes at the 5' and 3' ends. The physical meaning of this ratio is as follows: the numerator $\int_{bound} e^{-w(r)/k_B T} dr$ provides an estimate for the effective 1D line width that corresponds to the bound state, while the denominator $(2\pi k_B T/k_r)^{\frac{1}{2}}$ provides an estimate for the 1D line width that corresponding to a ligand harmonically restrained to the r* location in the bulk region. The fact that the two line widths are close to each other (for example,

for the 5' end site, the two line width are 1.26 Å and 0.99 Å) is mainly the result of the choice of the harmonic force constant k_r rather than the consequence of some fundamental physical factors. In fact, one could choose a stronger k_r, which would lead to a more negative $-k_BT \ln \frac{\int_{bound} e^{-w(r)/k_BT} dr}{(2\pi k_BT/k_r)^{\frac{1}{2}}}$.

Although in this work, the zero point of the PMF w(r) is set to be the bound state, the presence of the w(r*) term in the Eq. (6) doesn't depends on the choice of the zero point of w(r). Instead, it follows from the fact that PMF w(r) is related to the 1D probability density P(r) by $w(r) = -k_BTln[P(r)] + C$, where C is an arbitrarily chosen constant. It can be shown that any choice of the C or equivalently the zero point of the w(r) will lead to the same expression of the ΔG°_{bind} in the Eq. 6.

We compute the ligand 1D PMF $w(r)$ using umbrella sampling simulations in explicit ionic solvent [5], in which a series of MD simulations is performed with the harmonic distance restraint on the DNA atom *a* and ligand atom *A*: see Fig. S1. The biasing potential in the *i*-th simulation window is $U_{r,i} = \frac{1}{2}k_r(r - r_i^0)^2$, where r_i^0 is the reference distance for the *i*-th sampling window. The full range of the distance space is covered using 24 umbrella windows for the 5' site complex and 26 umbrella windows for the 3' complex. A single force constant k_r = 1000 kJ mol^{-1} nm^{-2} is used for the distance restraint in all the sampling windows. The force constants used in the angular restraints are: $k_\theta = k_\phi = k_\Theta = k_\Phi = k_\Psi = 1000\ kJ\ rad^{-1}mol^{-1}$. In each umbrella window, a 25-ns MD simulation is performed, starting from the last simulation snapshot of the previous window. The first 5 ns are treated as equilibration, which allows the system to adjust to the current umbrella potential. The last 20 ns of sampling data are used for the calculation of PMF. The biased probability distributions along the distance *r* accumulated in these sampling

windows are unbiased and combined using the Weighted Histogram Analysis Method (WHAM) method to yield the unbiased distribution and the potential of mean force $w(r)$ [6,7]. The WHAM program implemented by Grossfield is used to calculate the PMF [8]. The statistical uncertainties are estimated by dividing the trajectory in each sampling window into two blocks; the difference in the results obtained using the first half of the trajectory and the second half of the trajectory gives an estimate of the error bar at each intermediate distance: see Fig. 3 and Fig. 4 of the main text.

The double decoupling method

In DDM, the expression Eq. (1) for the absolute binding free energy $\Delta G^{\circ}_{\mathrm{bind}}$ is also evaluated by inserting intermediate states [1,2], i.e.

$$\Delta G^{\circ}_{\mathrm{bind}} = -k_B T \ln \frac{V Z_{RL,N} Z_{0,N}}{V_0 Z_{R,N} Z_{L,N}} = -k_B T \ln\left(\frac{V}{V_0} \frac{Z_{RL,N}}{Z_{R\cdots L,N}} \frac{Z_{R\cdots L,N}}{Z_{R_{wat}\cdots L_{gas}}} \frac{Z_{R_{wat}\cdots L_{gas}}}{Z_{R,N} Z_{L_{gas}}} \frac{Z_{L_{gas}} Z_{0,N}}{Z_{L,N}}\right) \quad (8)$$

Here $Z_{R\cdots L,N}$ is the configuration integral of a solvated ligand-receptor complex in the bound state, in which the ligand is harmonically restrained to remain within the receptor binding site by a ligand-receptor distance restraint; $Z_{R_{wat}\cdots L_{gas}}$ is the configuration integral of a bound complex in which the ligand is harmonically restrained to remain within the receptor binding site but otherwise does not interact with its environment; $Z_{L_{gas}}$ describes an ideal-gas phase ligand.

Thus, Eq. (8) allows $\Delta G^{\circ}_{\mathrm{bind}}$ to be computed as

$$\Delta G^{\circ}_{\mathrm{bind}} = -\Delta G^{coupled}_{restr} - \Delta G^{bound}_{decoupl} + \Delta G^{gas}_{restr} + \Delta G^{bulk}_{decoupl} \quad (9)$$

Each term in Eq. (9) can be computed from simulation or analytically:

$$-\Delta G^{coupled}_{restr} = -k_B T \ln \frac{Z_{RL,N}}{Z_{R\cdots L,N}} \quad (9.1)$$

$$-\Delta G^{bound}_{decoupl} = -k_B T \ln \frac{Z_{R\cdots L,N}}{Z_{R_{wat}\cdots L_{gas}}} \quad (9.2)$$

$$\Delta G^{gas}_{restr} = -k_B T \ln \frac{V}{V_0} \frac{Z_{R_{wat}\cdots L_{gas}}}{Z_{R,N} Z_{L_{gas}}} \quad (9.3)$$

$$\Delta G_{decoupl}^{bulk} = -k_B T \ln \frac{Z_{L_{gas}} Z_{wat}}{Z_{L_{wat}}} \quad (9.4)$$

In Eq. (9.1), $\Delta G_{restr}^{coupled}$ is the free energy of switching on the ligand-receptor harmonic distance restraint to force the ligand to remain inside the binding site, when the ligand is fully coupled to its environment (i.e. receptor and solvent). This term is typically small and can be computed with a single simulation of the ligand-receptor complex in solution using the Zwanzig free energy perturbation formula [9]. $\Delta G_{decoupl}^{bound}$ is the free energy of turning off (i.e. decoupling) the nonbonded interactions between the ligand and its environment, when the ligand remains in the binding site in the presence of the ligand-receptor harmonic distance restraint; this term is computed using thermodynamic integration (TI) by simulating the decoupling process. $\Delta G_{restr}^{gas}$is the free energy of switching on the harmonic distance restraint that forces the gas-phase ligand to remain inside the receptor binding site [2]; this term is evaluated analytically by

$$\Delta G_{restr}^{gas} = -k_B T \ln \frac{V}{V_0} \frac{Z_{R_{wat}\cdots L_{gas}}}{Z_{R,N} Z_{L_{gas}}} = -k_B T \ln \frac{V}{V_0} \frac{1}{V} \left(\frac{2\pi k_B T}{k_r}\right)^{\frac{3}{2}} = -k_B T \ln \frac{1}{V_0} \left(\frac{2\pi k_B T}{k_r}\right)^{\frac{3}{2}} \quad (10)$$

where k_r is the force constant of the ligand-receptor distance restraint. In Eq. (9.4) $-\Delta G_{decoupl}^{bulk}$ is the free energy of inserting the gas-phase ligand into the solvent. This term is the ligand solvation free energy, which is also calculated using TI.

Therefore, a DDM calculation involves two legs of decoupling simulations in order to compute the $\Delta G_{decoupl}^{bound}$ and $-\Delta G_{decoupl}^{bulk}$. In the each leg of the decoupling simulations the Coulomb intermolecular interaction is turned off first using 11 λ-windows, and the Lennard-Jones intermolecular interaction is then switched off in 17 λ-windows. (Coulomb decoupling: λ = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0; Lennard-Jones decoupling: λ = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.94, 0.985, 1.0). To achieve convergence at each λ, a decoupling simulation is performed for 15 ns. The first 5 ns of the trajectory were treated as

equilibration and the last 10 ns trajectory were used to compute $\Delta G^{\circ}_{\text{bind}}$. For each $\Delta G^{\circ}_{\text{bind}}$ calculation, three independent DDM runs with different starting velocities were carried out. The statistical error bars are estimated from the standard deviations of the three independent sets of DDM simulations. A total of ≈ 2.5 μs simulation data were collected to compute a single $\Delta G^{\circ}_{\text{bind}}$ value. The total simulation time which includes the equilibration for all the DDM simulations is ≈ 10 μs.

Binding free energy decomposition based on DDM

As seen from Eq. (9), a DDM calculation [1,2,10,11] involves two legs of decoupling simulations, in which a ligand is decoupled from the receptor binding pocket or from the bulk solution by gradually turning off the nonbond interactions (Coulomb and Lennard-Jones) between the ligand and its environment. These two decoupling simulations yield $\Delta G^{bound}_{decoupl}$ and $\Delta G^{bulk}_{decoupl}$, the free energy of decoupling the ligand from the receptor binding site and that from the bulk solvent, respectively. When the decoupling is done by first turning off the electrostatic Coulomb interactions between the ligand and its environment, followed by turning off the intermolecular Lennart-Jones interactions [10,12], the decoupling simulations provide estimates for the following free energy terms

$$\Delta G^{bound}_{decoupl} = \Delta G^{bound}_{decoupl-Coulomb} + \Delta G^{bound}_{decoupl-LJ}$$

$$\Delta G^{bulk}_{decoupl} = \Delta G^{bulk}_{decoupl-Coulomb} + \Delta G^{bulk}_{decoupl-LJ} \qquad (11)$$

Combining Eq. (11) with the Eq. (9), the absolute binding free energy $\Delta G^{\circ}_{\text{bind}}$ can be rewritten as

$$\Delta G^{\circ}_{\text{bind}} = -\Delta G^{coupled}_{restr} + \Delta G_{Coulomb} + \Delta G_{LJ} + \Delta G^{gas}_{restr} \qquad (12)$$

where $\Delta G_{Coulomb}$ is the difference in the free energies of turning of the Coulomb interaction between the ligand and the binding site and that between the ligand and the bulk solvent, i.e.

$$\Delta G_{Coulomb} = -(\Delta G^{bound}_{decoupl-Coulomb} - \Delta G^{bulk}_{decoupl-Coulomb}) \quad (12.1)$$

The ΔG_{LJ} has a similar meaning, i.e.

$$\Delta G_{LJ} = -(\Delta G^{bound}_{decoupl-LJ} - \Delta G^{bulk}_{decoupl-LJ}) \quad (12.2)$$

To a certain extent, these two terms $\Delta G_{Coulomb}$ and ΔG_{LJ} reflect the contributions of the effective electrostatic interactions and the nonpolar intermolecular interactions to the binding free energy, respectively [11,13], and they therefore contain physical insights into the thermodynamic driving forces of binding. However, it should be noted that the decomposition of total binding free energy depends on the choice of alchemical path. For example, such a decomposition is only physically meaningful when the Coulomb interactions are switched off before the Lennard-Jones interactions are turned off, as is done here. While such decomposition can be useful for understanding the nature of ligand binding, the interpretation it provides is qualitative and should be viewed with caution.

Lastly, it should be noted that the simulations of the DDM calculation are performed in finite-size, periodic solvent box; for charged ligands this can have a non-negligible effect on the calculated electrostatic decoupling free energies $\Delta G^{bound}_{decoupl-Coulomb}$ and $\Delta G^{bulk}_{decoupl-Coulomb}$. Here we used a scheme developed by Rocklin et al.[14] to compute the correction for these effects. The electrostatic finite-size corrections include the following contributions: periodicity-induced net-charge interactions, periodicity-induced net-charge undersolvation, discrete solvent effects, and residual integrated potential effects.[14] Calculating the residual integrated potential effects requires three Poisson-Boltzmann calculations (PB) of electrostatic potentials generated by different combinations of receptor and ligand charge distributions. Here the PB calculations are performed using the APBS program[15]. With these corrections, the final expression for the absolute binding free energy from DDM is written as

$$\Delta G^{\circ}_{\text{bind}} = -\Delta G^{coupled}_{restr} + \Delta G_{Coulomb} + \Delta G_{LJ} + \Delta G^{finite\_size}_{elec\_corr} + \Delta G^{gas}_{restr} \quad (13)$$

Hydration Site Analysis Calculation

To perform HSA calculations [16,17], a molecular dynamics simulation of the unliganded c-MYC DNA-quadruplex was performed using the AMBER parmbsc0 force field [18] in a TIP3P [19] cubic water box, using the AMBER14 package. The water box was built with a minimum 10.0 Å buffer between solute atoms and box boundary. After preparation, structures were minimized for up to 10000 cycles, with all solute and ions harmonically restrained using a force constant of 100 kcal mol^{-1} $Å^{-2}$. The minimization step was followed by a 200 ps NVT equilibration where the system was heated at a constant volume from 100 K to 300 K and a 350 ps NPT equilibration where the restraints were reduced 100 kcal mol^{-1} $Å^{-2}$ to 10 kcal mol^{-1} $Å^{-2}$. The system was again minimized for up to 10000 cycles maintaining the previous restraints and equilibrated to a temperature of 300 K and a pressure of 1 atm under NPT conditions for 6 ns. An NVT production run was started from a configuration in which the box volume was close to the average of the last 5 ns of the previous NPT simulation. The production MD simulations were run for 100 ns and snapshots were collected every 1.0 ps. The first 1 ns of the production run was discarded.

During the production run, position restraints with a force constant of 10 kcal mol^{-1} $Å^{-2}$ were used on all of the heavy atoms and SHAKE was used to constrain any bonds involving hydrogen atoms. The temperature was maintained with a Langevin thermostat. A stochastic Langevin dynamics integrator with a friction constant of 1 ps^{-1} and a time step of 1 fs was used to integrate the equations of motion and to provide constant temperature control at T = 300 K. Electrostatic interactions were computed using the particle-mesh Ewald (PME) method [20].

Hydration site analysis was used to determine solvation thermodynamics and structural properties of water molecules inside the 5'end and 3'end binding cavities. In this work, we performed HSA analysis on a subset of 10,000 frames from the 100 ns production trajectory. High density spherical regions (hydration sites) of 1□ radius were identified using a clustering procedure [16] on the water molecules that were found within 3 Å of the heavy atoms of the quinolone rings in an aligned holo conformation. The resulting hydration sites were each populated by retrieving all water molecules, which had oxygen atoms within 1.0 □ from the corresponding hydration site center. The hydration sites were then enumerated according to their occupancies, with the highest populated site given the index 0.

The energy of each hydration site (E_{total}) is calculated as a function of the pairwise solute-water energies (E_{sw}) and water-water energies (E_{ww}) [16],

$$E_{tot} = \frac{1}{2}\,(E_{sw} + E_{ww})$$

where a factor of one-half takes into account the assignment of pairwise interactions to each molecule in the pair.

Hydrogen bonds were identified based on the following geometric criteria: the donor−acceptor heavy atom distance is less than or equal to 3.5 Å and the hydrogen−donor−acceptor angle is less than or equal to 30°.[17] The total number of hydrogen (HB_{tot}) is equal to sum of the solute-water (HBsw) and water-water hydrogen bonds (HBww).

Fractional enclosure (f_{enc}) was calculated based on the following function:[17]

$$f_{enc} = 1 - \frac{N_{nbr}}{N_{nbr-bulk}}$$

where N_{nbr} is mean number of water molecules found in the first hydration shell of a hydration site or first shell neighbors and $N_{nbr\text{-}bulk}$ is mean number of first-shell neighbors of a TIP3P water

molecule in neat water simulation as defined in this previous work. This quantity indicates the degree to which the water in a hydration site is blocked from contact with other water molecules.

Table S1. Thermodynamic properties of cavity waters at the 5' site computed using HSA.

index	Occupancy	Esw	Eww	Etot	HBww	HBsw	HBtot	Enclosure	Ligand-overlap
0	0.77	-4.14	-5.81	-9.95	1.91	1.09	3.00	0.49	YES
1	0.61	-2.39	-7.95	-10.34	2.95	0.03	2.98	0.34	YES
2	0.68	-4.19	-6.21	-10.40	2.09	1.60	3.69	0.31	NO
3	0.42	-1.27	-8.31	-9.58	2.85	0.04	2.90	0.23	YES
4	0.37	-1.33	-8.45	-9.78	2.89	0.16	3.05	0.21	YES
5	0.46	-3.07	-5.96	-9.03	2.21	0.70	2.91	0.43	YES
6	0.36	-2.38	-7.50	-9.89	3.16	0.16	3.32	0.14	YES
7	0.35	-1.17	-8.71	-9.88	3.06	0.01	3.08	0.20	YES
8	0.34	-2.98	-6.95	-9.93	2.76	0.53	3.30	0.19	YES
9	0.30	-3.21	-6.58	-9.79	2.78	0.50	3.27	0.18	YES
10	0.35	-0.96	-8.53	-9.50	2.92	0.05	2.97	0.20	YES
11	0.30	-2.56	-7.13	-9.70	2.90	0.37	3.27	0.17	YES
BULK	0.14	0	-9.53	-9.53	3.33	0	3.33	0	

Esw, solute-water energy; Eww, water-water energy; Etot, total energy; HBww, water-water hydrogen bonds; HBsw, solute-water hydrogen bonds; Ligand-overlap, hydration sites are located in the same region or proximal to the ligand location in the binding cavity, BULK, bulk water, values were obtained from neat TIP3P water simulation (ref. Haider et al) [17].

Table S2. Thermodynamic properties of cavity waters at the 3' site computed using HSA.[a]

index	Occupancy	Esw	Eww	Etot	HBww	HBsw	HBtot	Enclosure	Ligand-overlap
0	0.87	-4.32	-4.38	-8.70	1.02	1.57	2.59	0.67	YES
1	0.76	-4.55	-5.52	-10.06	2.15	0.91	3.06	0.44	YES
2	0.55	-3.05	-6.26	-9.31	2.23	0.91	3.14	0.32	YES
3	0.51	-3.11	-6.75	-9.86	2.48	0.80	3.28	0.28	YES
4	0.47	-2.40	-7.35	-9.75	2.67	0.94	3.61	0.14	NO
5	0.44	-1.55	-7.98	-9.53	2.85	0.25	3.10	0.25	YES
6	0.42	-3.35	-6.36	-9.71	2.51	0.73	3.24	0.25	YES
7	0.35	-0.89	-8.90	-9.79	3.07	0.01	3.08	0.15	YES
8	0.41	-1.86	-8.37	-10.22	3.23	0.08	3.31	0.15	YES
9	0.36	-1.08	-8.29	-9.37	2.78	0.21	2.99	0.21	YES
10	0.35	-0.49	-9.17	-9.66	2.93	0.01	2.94	0.22	YES
11	0.37	-1.49	-8.23	-9.73	2.95	0.04	2.99	0.22	YES
BULK	0.14	0	-9.53	-9.53	3.33	0	3.33	0	

a. See footnotes for Table S1.

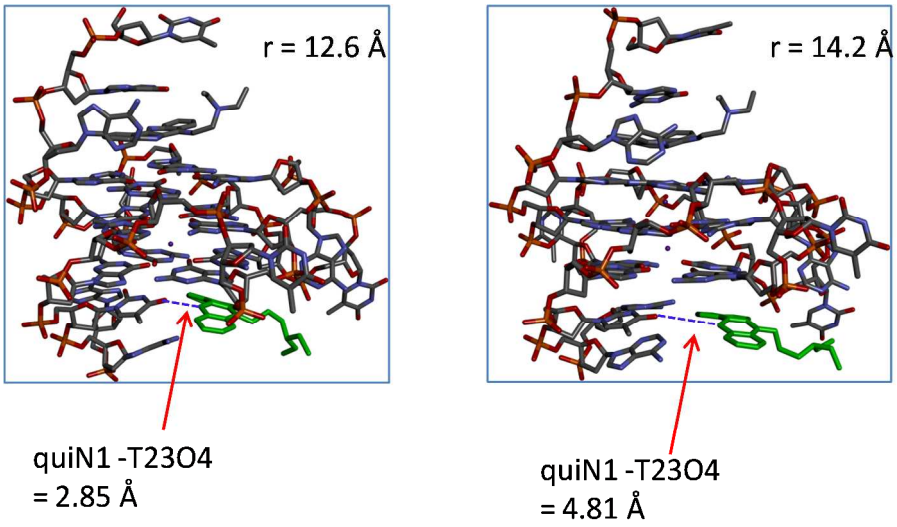

Figure S2. Variations of the intermolecular distance between quindoline N1 and T23O4 as the ligand-DNA distance is increased from r = 12.6 Å to 14.2 Å.

REFERENCES

(1) Gilson, M.; Given, J.; Bush, B.; Mccammon, J. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophysical Journal* **1997**, *72* (3), 1047–1069.

(2) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *Journal of Physical Chemistry B* **2003**, *107* (35), 9535–9551.

(3) Gallicchio, E.; Levy, R. M. Recent Theoretical and Computational Advances for Modeling Protein–ligand Binding Affinities. In *Advances in Protein Chemistry and Structural Biology*; Elsevier, 2011; Vol. 85, pp 27–80.

(4) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *Journal of Computational Physics* **1976**, *22* (2), 245–268.

(5) Deng, N.-J.; Cieplak, P. Free Energy Profile of RNA Hairpins: A Molecular Dynamics Simulation Study. **2010**, *98* (4), 627–636.

(6) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *Journal of Computational Chemistry* **1992**, *13* (8), 1011–1021.

(7) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. Temperature Weighted Histogram Analysis Method, Replica Exchange, and Transition Paths †$. *Journal of Physical Chemistry B* **2005**, *109* (14), 6722–6731.

(8) Grossfield, A. *WHAM: The Weighted Histogram Analysis Method*; http://membrane.urmc.rochester.edu/content/wham.

(9) Zwanzig, R. W. High□Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22* (8), 1420–1426.

(10) Deng, Y.; Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *Journal of Chemical Theory and Computation* **2006**, *2* (5), 1255–1273.

(11) Deng, N.; Zhang, P.; Cieplak, P.; Lai, L. Elucidating the Energetics of Entropically Driven Protein–Ligand Association: Calculations of Absolute Binding Free Energy and Entropy. *The Journal of Physical Chemistry B* **2011**, *115* (41), 11902–11910.
(12) Deng, Y.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *Journal of Physical Chemistry B* **2009**, *113* (8), 2234–2246.
(13) Lin, Y.-L.; Meng, Y.; Huang, L.; Roux, B. Computational Study of Gleevec and G6G Reveals Molecular Determinants of Kinase Inhibitor Selectivity. *Journal of the American Chemical Society* **2014**, *136* (42), 14753–14762.
(14) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. Calculating the Binding Free Energies of Charged Species Based on Explicit-Solvent Simulations Employing Lattice-Sum Methods: An Accurate Correction Scheme for Electrostatic Finite-Size Effects. *The Journal of Chemical Physics* **2013**, *139* (18), 184103.
(15) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proceedings of the National Academy of Sciences* **2001**, *98* (18), 10037–10041.
(16) Nguyen, C. N.; Cruz, A.; Gilson, M. K.; Kurtzman, T. Thermodynamics of Water in an Enzyme Active Site: Grid-Based Hydration Analysis of Coagulation Factor Xa. *Journal of Chemical Theory and Computation* **2014**, *10* (7), 2769–2780.
(17) Haider, K.; Wickstrom, L.; Ramsey, S.; Gilson, M. K.; Kurtzman, T. Enthalpic Breakdown of Water Structure on Protein Active-Site Surfaces. *The Journal of Physical Chemistry B* **2016**, *120* (34), 8743–8756.
(18) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophysical Journal* **2007**, *92* (11), 3817–3829.
(19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926.
(20) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577.