

# Molecular modeling and biophysical analysis of the c-MYC NHE-III<sub>1</sub> silencer element

Derek J. Cashman · Robert Buscaglia ·  
Matthew W. Freyer · Jamie Dettler ·  
Laurence H. Hurley · Edwin A. Lewis

Received: 19 July 2007 / Accepted: 7 November 2007 / Published online: 18 December 2007  
© Springer-Verlag 2007

**Abstract** G-Quadruplex and i-Motif-forming sequences in the promoter regions of several oncogenes show promise as targets for the regulation of oncogenes. In this study, molecular models were created for the c-MYC NHE-III<sub>1</sub> (nuclease hypersensitivity element III<sub>1</sub>) from two 39-base complementary sequences. The NHE modeled here consists of single folded conformers of the polypurine intramolecular G-Quadruplex and the polypyrimidine intramolecular i-Motif structures, flanked by short duplex DNA sequences. The G-Quadruplex was based on published NMR structural data for the c-MYC 1:2:1 loop isomer. The i-Motif structure

is theoretical (with five cytosine–cytosine pairs), where the central intercalated cytosine core interactions are based on NMR structural data obtained for a tetramolecular [d(A<sub>2</sub>C<sub>4</sub>)<sub>4</sub>] model i-Motif. The loop structures are in silico predictions of the c-MYC i-motif loops. The porphyrin meso-tetra(N-methyl-4-pyridyl)porphine (TMPyP4), as well as the ortho and meta analogs TMPyP2 and TMPyP3, were docked to six different locations in the complete c-MYC NHE. Comparisons are made for drug binding to the NHE and the isolated G-Quadruplex and i-Motif structures. NHE models both with and without bound cationic porphyrin were simulated for 100 ps using molecular dynamics techniques, and the non-bonded interaction energies between the DNA and porphyrins calculated for all of the docking interactions.

D. J. Cashman · R. Buscaglia · M. W. Freyer · J. Dettler ·  
E. A. Lewis (✉)  
Department of Chemistry and Biochemistry,  
Northern Arizona University,  
P.O. Box 5698, South Beaver Street,  
Flagstaff, AZ 86011-5698, USA  
e-mail: edwin.lewis@nau.edu

L. H. Hurley  
College of Pharmacy, University of Arizona,  
Tucson, AZ 85721, USA

L. H. Hurley  
Arizona Cancer Center,  
1515 North Campbell Ave.,  
Tucson, AZ 85724, USA

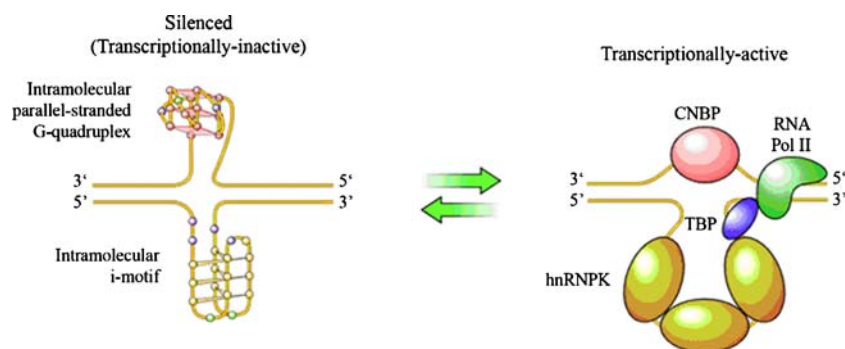
L. H. Hurley  
BIO5 Institute for Collaborative Bioresearch,  
University of Arizona,  
1657 East Helen Street,  
Tucson, AZ 85721, USA

L. H. Hurley  
Department of Chemistry, University of Arizona,  
Tucson, AZ 85721, USA

**Keywords** G-quadruplex · i-Motif · Oncogenes · c-MYC ·  
Silencer element · NHE III<sub>1</sub> · TMPyP4

## Introduction

The development of anticancer agents that target nucleic acids has been a popular and effective means of anticancer drug design [1]. Because these agents are not highly selective towards a particular sequence or gene, issues of toxicity and other undesirable side effects still remain [1, 2]. Recent efforts in anticancer drug design have focused on small molecules that bind to higher order DNA structures, such as the G-Quadruplex or i-Motif, as a way to enhance the selectivity of such compounds [3–8]. These higher order DNA structures have been proposed to form in the promoter regions of 17 out of 30 known oncogenes, such as c-MYC, VEG-F, k-Ras, bcl-2, as well as the human telomere [5, 9, 10]. The human c-MYC protein is a 65



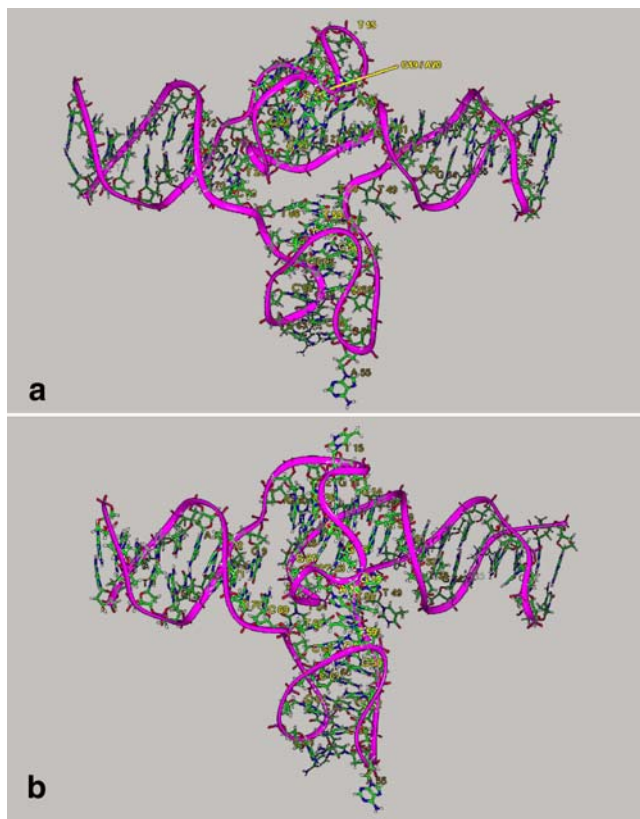
**Fig. 1** Diagram of the c-MYC NHE-III<sub>1</sub> (nuclease hypersensitivity element III<sub>1</sub>) “silencer” element, showing the equilibrium between the transcriptionally inactive and active forms. On the left is the

transcriptionally inactive form, containing both the G-Quadruplex and i-Motif structures. On the right, is the transcriptionally active form, with multiple transcription factors bound

kDa nuclear phospho-protein that has been implicated in various physiological processes— cell growth, proliferation, loss of differentiation, and cell death (apoptosis) [11, 12]. Furthermore, the aberrant overexpression of the c-MYC oncogene has been shown to be associated with a broad range of human cancers [13–15]. A segment of the c-

MYC P<sub>1</sub> promoter is termed the nuclease hypersensitivity element III<sub>1</sub> (NHE-III<sub>1</sub>), and is responsible for 80–90% of c-MYC transcription [16–21].

The NHE-III<sub>1</sub> element can form both transcriptionally active and silenced (transcriptionally inactive) forms (Fig. 1) [22]. The polypurine strand of the NHE-III<sub>1</sub> is a guanine-rich sequence consisting of five consecutive runs of either three or four guanines (Fig. 2) [23]. This strand can form G-Quadruplex structures and has been demonstrated to act as a silencer element and suppress c-MYC gene transcription [24]. The G-rich strand can fold to form a number of different intramolecular G-Quadruplex structures. The most stable of these structures, based on NMR and thermodynamic data, is the 1:2:1 loop isomer [23]. The complementary polypyrimidine strand consists of four runs of either three or four cytosines (Fig. 2). The polypyrimidine strand has been shown to form an intramolecular i-Motif structure, in which three hemi-protonated cytosine base pairs are intercalated in between three or four hemi-protonated cytosine base pairs [8, 25]. A recent report of the solution NMR structure of the c-MYC i-Motif has shown the presence of at least two equilibrating structures, both having a total of four cytosine–cytosine pairs, and either two proximate but unpaired cytosines or a looped out cytosine between two intercalated cytosines (Yang and Hurley 2007, unpublished data). Although studies have shown that the formation of G-Quadruplex and i-Motif structures may not be highly favored under physiological conditions, it has been demonstrated that drug binding can drive the equilibrium towards the formation of these structures [26–28]. This has been proposed as a therapeutic mechanism for downregulating oncogene expression [29–32].



**Fig. 2** Molecular models of the average structure of the final 20 ps of the molecular dynamics simulation of the c-MYC NHE-III<sub>1</sub> “silencer” element. The G-Quadruplex structure is at the top-center, and the i-Motif is at the bottom-center of each picture. **a** “Rotation #1” of the G-Quadruplex, with the T15 loop at the top and rear and the G19/A20 loop at the top and front of the picture. **b** “Rotation #2” of the G-Quadruplex, with the T15 loop at the top and front of the image, and the G19/A20 loop at the front and adjacent to the G-Quadruplex/i-Motif interface

In this paper, we have used molecular modeling techniques to construct a model of the complete c-MYC NHE-III<sub>1</sub> silencer element. This model is based on the existing NMR structure of the isolated G-Quadruplex 1:2:1 loop isomer [23] and a theoretical structure of the isolated i-Motif, with the central intercalated cytosine core interactions based on NMR structural data [33] and the i-Motif

### G-Quadruplex Forming Strand

5' – TGGGG AGGGTGGGGAGGGTGGGGAA GG – 3'  
3' – ACCCC TCCCACCCCTCCCACCCCTT CC – 5'

### i-Motif Forming Strand

**Scheme 1** 27mer oligonucleotide sequences for the c-MYC G-Quadruplex and i-Motif NHE-III<sub>1</sub> components. The *underlined bases* represent the guanines involved in G-tetrad interactions or the cytosines involved in hemi-protonated cytosine-cytosine interactions

loop structures predicted in silico. The objective of this study is to provide structural information for the rational design of potential drug compounds with high affinity for the c-MYC NHE III<sub>1</sub>. It is envisioned that the NHE III<sub>1</sub> structure might have new drug binding sites created in the cavity between the G-Quadruplex and i-Motif or in areas where the G-Quadruplex or i-Motif loops come into close proximity. Although not investigated here, these new drug binding sites have potential for binding compounds with greater selectivity for targeting oncogenes.

### Materials and methods

#### Construction of complete silencer element

The complete silencer element was assembled in four stages (G-Quadruplex, i-Motif, and two flanking duplex structures) using the molecular modeling software InsightII (Accelrys, San Diego, CA). The G-Quadruplex portion was obtained by starting with a 22-mer G-Quadruplex NMR solution structure, obtained from the Protein Data Bank (pdb accession code: *Ixav*) [23]. The relevant bases were mutated back to wild type c-MYC using the Biopolymer module of InsightII. The i-Motif structure was built using the NMR solution structure of a tetrameric i-Motif, [d (A<sub>2</sub>C<sub>4</sub>)<sub>4</sub>], (pdb accession code: *Iybl*) as a scaffold [33], to which the relevant bases of the wild type c-MYC polypyrimidine i-Motif forming sequence were added using the Biopolymer module of InsightII. Bases **A9** and **A13** were added to one end of the i-Motif scaffold, forming loops that connect two of the four strands of the tetrameric i-Motif scaffold. At the opposite end of the scaffold, base **T13** was added between **C12** and **C14**, forming the final loop in the model structure for the c-MYC i-Motif. Bases **T4** and **T22** were added at the remaining two ends, forming

a thymidine triplet structure by forming hydrogen bonds: **T4 O4 - T13 N3**, **T13 O4 - T22 N3**, and **T22 O4 - T4 N3**. Two double-stranded DNA duplex segments were then created using the Biopolymer module, based on the wild type sequence of the c-MYC promoter (Scheme 1), using the bases that were not already incorporated into either the G-Quadruplex or i-Motif structures. GC base pairs were added to the exterior ends of each of these duplexes, bringing the total length of each to 9 base pairs long, such that the final sequence of the complete silencer models is as depicted in Scheme 2. Each strand of the two flanking duplex structures was then attached to the appropriate available strand of the G-Quadruplex and i-Motif components of the NHE. Because there is some potential variability in the orientation of the G-Quadruplex, two models of the complete silencer element were made; one with the 1-2-1 loops of the G-Quadruplex on the exterior of the NHE, and the other with the 1-2-1 loops of the G-Quadruplex on the interior of the NHE. An initial 600 cycles of steepest descent energy minimization was performed on the completed structures in vacuo, with the atoms of the G-Quadruplex and i-Motif portions fixed, to relieve any major areas of stress that may have been caused near the initial points of attachment.

#### Energy minimization and molecular dynamics

The models were solvated using a layer of 4,617 water molecules and 82 potassium ions. This amounts to a single water layer with a thickness of between 9 and 12 Å and roughly corresponds to the maximum number of water molecules that can be accommodated by the Insight II program. It would be better to use periodic boundary conditions, essentially enclosing the whole system in a box, but the solvated silencer element model is again too large (i.e., contains too many water molecules) to be analyzed by the Insight II program. For these reasons, we used a single thick water layer, and did not see a significant number of water molecules escaping into the vacuum during the simulation. The entire solvated silencer element was subjected to 1,000 cycles of steepest descent energy minimization using the Amber force field. Molecular dynamics (MD) calculations were carried out using the Discover module of InsightII and the Amber force field parameterization. The systems were heated slowly with

### QUADRUPLEX

5' – GCTTATGGG GAGGGTGGGGAGGGTGGGGAA GGTGGGGGG – 3'  
3' – CGAATACCC CTCCCACCCCTCCCACCCCTT CCACCCCCC – 5'

### DUPLEX i-MOTIF DUPLEX

**Scheme 2** 39mer oligonucleotide sequences for the polypurine and polypyrimidine strands in the construction of the model c-MYC NHE-III<sub>1</sub>. These sequences correspond to the wild type c-MYC P1 promoter

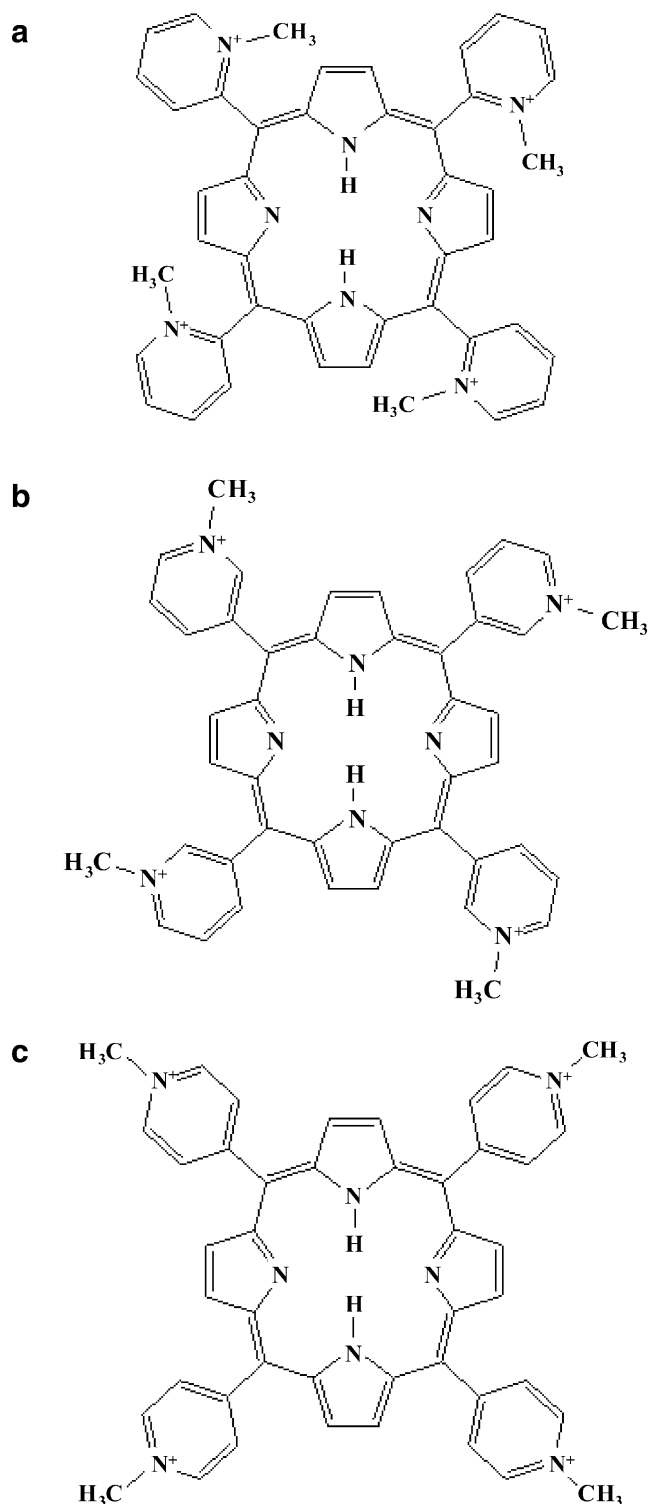
gradual removal of positional restraints on the DNA following this protocol: (1) 3,000 fs MD (T 100 K) holding DNA fixed; (2) 3,000 fs MD (T 100 K), DNA free; (3) 3,000 fs MD (T 200 K) holding DNA fixed; (4) 3,000 fs MD (T 200 K), DNA free; (5) 3,000 fs MD (T 298.15 K), holding DNA fixed. The production run consisted of unrestrained dynamics at 298.15 K for 100 ps. A non-bonded interaction cut-off value of 12.00 Å was used for all simulations, with a distance-dependent dielectric constant, and the 1–4 non-bonded interactions scaled by 0.5. An average structure was obtained of the final 20 ps of the simulation, and this structure was then subjected to 2,000 cycles of conjugate gradient energy minimization. Analysis of the structures and trajectories utilized the Analysis module of InsightII.

#### Docking and scoring with TMPyP4 and analogues

The cationic porphyrin meso-tetra(N-methyl-4-pyridyl)porphine, TMPyP4 (Scheme 3), was docked into six different sites on the pre-dynamics (starting) structure of the complete silencer element. The six different potential binding sites of the silencer element that were analyzed are: (1) immediately above the thymidine triplet region of the i-Motif, (2) the opposite end of the i-Motif (bottom), (3) the 5' "end" of the G-Quadruplex, adjacent to the **G12-G16-G21-G25** tetrad on the 5' end, (4) intercalated between the **G12-G16-G21-G25** and the **G13-G17-G22-G26** tetrads, (5) intercalated between the **G13-G17-G22-G26** and the **G14-G18-G23-G27** tetrads, and (6) adjacent to the **G14-G18-G23-G27** tetrad on the 3' end. Each complexed structure was subjected to the energy minimization and molecular dynamics steps described above for the free DNA structures. Due to a minor anomaly in the amber force field with regards to the atoms of the small molecule porphyrins, it was necessary to apply constraints in Discover to the atoms of two of the pyrrole rings throughout the minimization and dynamics steps, to maintain the planarity of the porphyrin ring system. To compare the exterior binding models of TMPyP4 with that of analog molecules TMPyP2 and TMPyP3, the average structure of the final 20 ps of the simulation was taken, and TMPyP4 was modified using the Builder module of InsightII to TMPyP2 and TMPyP3 (Scheme 3). These structures were subjected to an additional 2,000 steps of conjugate gradient energy minimization. The non-bonded interaction energies (van der Waal's and coulombic interactions) were calculated for the final complexes using the Docking module of InsightII.

#### Results and discussion

An analysis of the non-bound average structures of the final 20 ps of the NHE-III<sub>1</sub> silencer simulation reveals some insightful observations potentially useful in the design of



**Scheme 3** Structures for **a** TMPyP2, **b** TMPyP3, and **c** TMPyP4

drugs targeting this structure. There is a reasonably large, crescent-shaped space located at the interface between the G-Quadruplex structure and the i-Motif (Fig. 2a,b). In the first structure (designated as "rotation #1" in this paper), the 1-base **T15** loop and the 2-base **G19/A20** loop are located on the opposite side of the G-Quadruplex from the



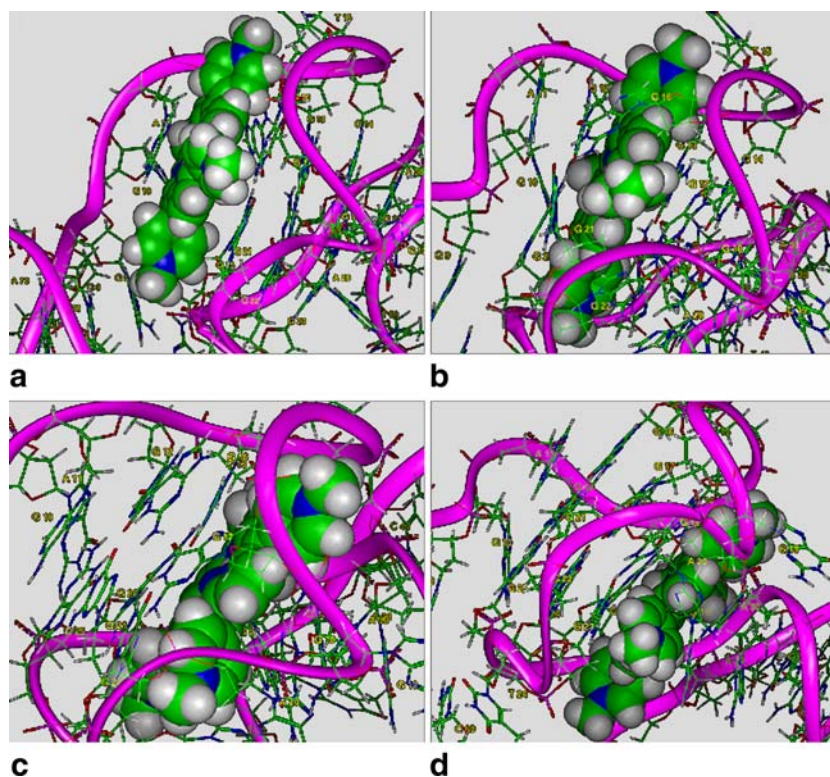
i-Motif. Although there are no bases in the G-Quadruplex that interact directly with the i-Motif in this structure, the 1-base **T24** loop appears to make favorable Van der Waal's contacts with the deoxyribose ring of the **C69** base, which is immediately to the 3' end of base **T68**, one of the bases forming the thymidine triplet structure of the i-Motif (Fig. 2a). In the second structure (designated as "rotation #2" in this paper), the 1-base **T15** loop is still located on the opposite side of the G-Quadruplex from the i-Motif, though it is rotated to a more forward position (towards the front of the picture) than in rotation #1. The 2-base **G19/A20** loop is then located adjacent to the i-Motif, with **A20** forming a stacking interaction with base **T59** of the thymidine triplet structure in the i-Motif. There is still a crescent-shaped structure forming the interface between the G-Quadruplex and i-Motif; however, in rotation #2 it is more protected, with the **G19/A20** loop on one side (in front of the thymidine triplet in Fig. 2b) and the 1-base **T24** loop on the other side (behind the thymidine triplet in Fig. 2b).

A close-up look at the thymidine triplet structure of the i-Motif in rotation #1 reveals that the triplet is largely intact and stable, with bases **T50**, **T59**, and **T68** forming a dome-like structure, as opposed to being planar. There are three hydrogen bonding interactions stabilizing this structure, between the **T50** C=O and **T59** N-H (1.8 Å), the **T59**

C=O and **T68** N-H (1.8 Å), and the **T68** C=O and **T50** N-H (1.8 Å). Conversely, the thymidine triplet structure of the i-Motif in rotation #2 appears to be less stable and organized, likely due to the stacking interactions between base **A20** (quadruplex) with base **T59** (i-Motif), which appear to be moving the **T59** ring towards the quadruplex and out of plane with the thymidine triplet. There is still a hydrogen bonding interaction between the **T50** C=O and the **T59** N-H (1.8 Å), though now we see somewhat of a bifurcated interaction between the **T68** N-H with both the **T59** C=O (2.4 Å) and the **C60** C=O (1.8 Å). The **C60** base is stacked immediately below **T59**. Due to its awkward positioning, the **T68** base now forms a second bifurcated electrostatic interaction between the **T68** C=O and the **C51** NH<sub>2</sub> (1.8 Å) and the **T50** N-H (3.5 Å).

#### Binding of porphyrin compounds to G-quadruplex structure

All three of the porphyrin compounds were also docked to four different locations within the G-Quadruplex structure; at the 5' end, at the 3' end, and intercalated between the G-tetrads. Figure 3 shows molecular models of the docked complexes of TMPyP4 to each of the four docking sites of the G-Quadruplex. Overall, the non-bonded interaction energy scores were significantly more favorable in the



**Fig. 3** Molecular models of TMPyP4 bound to the G-Quadruplex portion of the NHE-III<sub>1</sub> silencer element. **a** TMPyP4 docked to the 5' end, between bases **G10/A11** and the **G12-G16-G21-G25** tetrad. **b** TMPyP4 intercalated near the 5' end, between the **G12-G16-G21-G25**

and the **G13-G17-G22-G26** tetrads. **c** TMPyP4 intercalated near the 3' end, between the **G13-G17-G22-G26** and the **G14-G18-G23-G27** tetrads. **d** TMPyP4 docked to the 3' end, between bases **G28/A30** and the **G14-G18-G23-G27** tetrad

docking sites for the G-Quadruplex than for the i-Motif sites (Table 1). This is most likely attributable to the fact that the DNA bases comprising the G-tetrads of the quadruplex appear to be more tightly packed than for the i-Motif. Additionally, it is important to note that for docking of the porphyrin compounds at the ends of the G-Quadruplex, they still appear to be partially intercalated, with a compound at the 5' end sandwiched between the **G12-G16-G21-G25** tetrad on one side and the **G10/A11** bases on the other side (Fig. 3a). A compound at the 3' end is sandwiched between the **G14-G18-G23-G27** tetrad on one side and the **G28/A30** bases on the other (Fig. 3d).

Compared to docking to the ends of the i-Motif, there are no bases on both sides of the porphyrin ring structure, so the term “end-binding” more accurately describes docking to the i-Motif sites than the “ends” of the G-Quadruplex. Although docking to the thymidine triplet end of the i-Motif shows interactions on both sides of the porphyrin compound, the interactions here with the G-Quadruplex are not stacking interactions with the porphyrin ring system, but rather interactions between bases in the G-Quadruplex loops with the pyridinium rings. This is also evident in the increased magnitude of the non-bonded interaction energy of docking to the thymidine triplet region versus the

opposite end of the i-Motif (Table 1), yet it is still less than the non-bonded interaction energy of docking to the ends or intercalating with the G-Quadruplex.

The most favorable non-bonded interaction energy was calculated in intercalating TMPyP4 between the **G12-G16-G21-G25** and the **G13-G17-G22-G26** tetrads (5' intercalated) of the “rotation #1” model of the silencer element, with a calculated energy of  $-118.7 \text{ kcal mol}^{-1}$ . The least favorable energy was seen in docking TMPyP2 to the 5' end of the “rotation #2” model of the silencer element, with a calculated energy of  $-69.6 \text{ kcal mol}^{-1}$ . Overall, intercalating TMPyP4 between either of the G-tetrads produced more favorable non-bonded interaction energy than binding to either of the ends, due to the increased Van der Waal's energy. Isothermal titration calorimetry (ITC) data from our laboratory also predicts that there are two binding mechanisms to the G-Quadruplex DNA, both intercalative as well as end binding. [3] However, the experimental data indicates that end binding is approximately 20 times more favorable than intercalation, due to the fact that end binding is entropically driven, while intercalation is enthalpically driven [3]. Although the non-bonded interaction energy scores calculated by InsightIII are definitely useful for rapid screening of different compounds and different orientations

**Table 1** Non-bonded interaction energies ( $\text{kcal mol}^{-1}$ ) between the NHE-III<sub>1</sub> silencer element and the compounds TMPyP2, TMPyP3, and TMPyP4. Interactions are partitioned into Van der Waal's energy and electrostatic energy. The total energy for each interaction is given

	TMPyP2		TMPyP3		TMPyP4	
	Rotation #1	Rotation #2	Rotation #1	Rotation #2	Rotation #1	Rotation #2
<b>i-Motif (T-triplet)</b>						
Van der Waal's energy	-43.8	-47.0	-46.6	-50.5	-49.4	-53.9
Electrostatic energy	-9.2	-7.8	-2.7	-7.9	-1.2	-1.9
Total energy	-53.0	-54.8	-49.2	-58.4	-50.6	-55.8
<b>i-Motif (bottom)</b>						
Van der Waal's energy	-39.6	-35.2	-43.3	-35.9	-42.5	-38.5
Electrostatic energy	-3.4	-3.2	-0.9	-2.3	-0.5	-0.6
Total energy	-43.0	-38.3	-44.3	-38.2	-42.9	-39.1
<b>G-Quadruplex (3' end)</b>						
Van der Waal's energy	-80.1	-85.5	-90.1	-95.5	-95.8	-93.8
Electrostatic energy	-7.7	-4.9	-6.8	-6.2	-4.1	-1.7
Total energy	-87.7	-90.4	-96.9	-101.6	-99.9	-95.5
<b>G-Quadruplex (3' intercalated)</b>						
Van der Waal's energy	N/A	N/A	N/A	N/A	-107.1	-105.2
Electrostatic energy	N/A	N/A	N/A	N/A	-3.1	-0.6
Total energy	N/A	N/A	N/A	N/A	-110.1	-105.8
<b>G-Quadruplex (5' intercalated)</b>						
Van der Waal's energy	N/A	N/A	N/A	N/A	-118.6	-104.9
Electrostatic energy	N/A	N/A	N/A	N/A	-0.1	-2.6
Total energy	N/A	N/A	N/A	N/A	-118.7	-107.6
<b>G-Quadruplex (5' end)</b>						
Van der Waal's energy	-67.4	-61.4	-75.9	-75.0	-77.0	-81.8
Electrostatic energy	-3.2	-8.1	-3.3	-8.6	-2.5	-4.5
Total energy	-70.6	-69.6	-79.3	-83.6	-79.5	-86.3

of compounds versus a drug target, it is clearly not 100% accurate as there is no entropy component involved in the calculation.

There were no major differences in the calculated energies of the methyl para-substituted TMPyP4 versus the methyl meta-substituted TMPyP3 compounds with respect to end binding. This is due primarily to the fact that the pyridinium rings are rotated at an angle to the porphyrin ring system and in the same direction of the nucleotide strand of the DNA, which allow enough room for either a para- or meta-substitution on the ring between the strands of the DNA. ITC experiments demonstrated that TMPyP3 and TMPyP2 do not intercalate in the G-Quadruplex, thus structures containing intercalated TMPyP3 and TMPyP2 molecules were not modeled. The ortho- and meta-substituted TMPyP2 and TMPyP3 place the methyl group closer to the center of the porphyrin ring, generating much more steric hindrance than TMPyP4. These “turned out” pyridyl rings would require a much wider intercalation cavity between G-tetrads in order to intercalate. The “turned out” pyridyl rings not only prevent intercalation, but also decrease the binding affinity for TMPyP3 and TMPyP2 to the exterior of the G-Quadruplex structure. This decrease in binding affinity was observed in ITC and is demonstrated *in silico* with lower overall energy scores in the end binding structures as compared to TMPyP4.

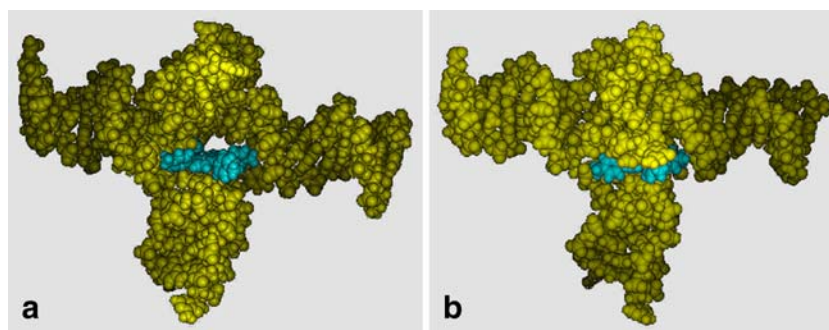
#### Binding of porphyrin compounds to the i-motif

The cationic porphyrin compounds TMPyP2, TMPyP3, and TMPyP4 were all bound to both ends of the i-Motif structure, one immediately adjacent to the thymidine triplet, near the interface between the G-Quadruplex and i-Motif, and the other at the opposite end of the i-Motif. Because the interactions between each porphyrin compound and the DNA is non-bonded in nature, the Docking module of InsightII was used to calculate the non-bonded interaction energy between both species, providing an empirical score

describing the interaction (Table 1). The total non-bonded interaction energy is the summation of two components: Van der Waal’s energy, and electrostatic energy.

The interaction energies for the binding of all three compounds to the thymidine triplet region of the i-Motif were all reasonably similar, ranging from  $-49.2 \text{ kcal mol}^{-1}$  for the binding of TMPyP3 to the triplet in “rotation #1”, to  $-58.4 \text{ kcal mol}^{-1}$  for the binding of TMPyP3 to the triplet in “rotation #2.” TMPyP3 (as well as TMPyP2 and TMPyP4) do not appear to be making any significant interactions with any part of the G-Quadruplex in their interactions with “rotation #1. However, if we look at “rotation #2,” we can see several intriguing interactions, as two of the loops of the quadruplex are now closer to the i-Motif. There are some favorable Van der Waal’s interactions between **A20** of the G-Quadruplex and one of the pyridinium rings. Also, the meta-substituted methyl group present on this pyridinium ring is also making favorable hydrophobic contacts with the deoxyribose ring of **T59** on the i-Motif. On the opposite side of the porphyrin ring, the pyridinium ring and methyl group are making favorable hydrophobic contacts with the **T24** base of the G-Quadruplex. All three porphyrin compounds also stack against the thymidine triplet of the i-Motif and appear to contribute to its stability. The binding of all three compounds to the opposite end (bottom) of the i-Motif were also quite similar, ranging from  $-38.2 \text{ kcal mol}^{-1}$  for the binding of TMPyP3 to the “rotation #2” structure, and  $-44.2 \text{ kcal mol}^{-1}$  for the binding of TMPyP3 to the “rotation #1” structure.

We also observed that the interaction energy in binding to both ends of the i-Motif was dominated by Van der Waal’s energy, with very little electrostatic component. This is due to the aromatic and hydrophobic nature of the porphyrin compounds. The planar, aromatic porphyrin ring structure tends to form stacking interactions with multiple pairs of bases of the DNA, and the pyridinium rings can interact favorably with other bases or to the deoxyribose rings of the DNA. Although all four compounds carry a +1



**Fig. 4** CPK models of the NHE-III<sub>1</sub> with a TMPyP4 molecule bound to the top of the i-motif structure. **a** “Rotation #1” of the G-Quadruplex, with the **T15** loop at the top and rear and the **G19/A20** loop at the top and front of the picture. **b** “Rotation #2” of the G-

Quadruplex, with the **T15** loop at the top and front of the image, and the **G19/A20** loop at the front and adjacent to the G-Quadruplex/i-Motif interface



charge on each of the four nitrogen atoms of the pyridinium rings, the solvent accessibility of the charge is reduced by the attached methyl group present on the pyridinium ring.

#### Binding of porphyrin compounds to the c-MYC NHE

CPK structures showing a TMPyP4 molecule bound to the top of the i-Motif in the NHE cavity are shown in Fig. 4. Figure 4a shows “rotation 1” of the G-Quadruplex; the cavity produced between the G-Quadruplex and the I-motif can be seen. Figure 4b shows “rotation 2” of the G-Quadruplex, which creates a smaller drug binding cavity at the interface between the G-Quadruplex and i-Motif components of the NHE. The empty interface cavity sizes for the two NHE structures are equivalent to the volume of 46 and 37 water molecules for “rotation 1” and “rotation 2”, respectively. The docked TMPyP4 molecule displaces 8 water molecules from “rotation 1” and 10 water molecules from “rotation 2”. A drug with a larger volume or with the ability to bridge the G-Quadruplex and i-Motif interface would have higher affinity for the NHE cavity.

#### Conclusions

In conclusion, it is important to note that the NHE-III<sub>1</sub> silencer element of the c-MYC gene promoter is extraordinarily complex. This particular sequence is very flexible, and contains DNA in three different forms (G-Quadruplex, i-Motif, and two flanking duplex structures). Furthermore, it is possible that the G-Quadruplex may fold up into multiple folded structures, such as the 1-2-1 loop isomer or the 1-6-1 loop isomer. We have attempted to compensate for multiple rotations of the 1-2-1 loop isomer of the G-Quadruplex by incorporating two different rotations of it into our models. Additional rotations of the G-Quadruplex were initially analyzed but discarded due to acute turns in the connection points between the G-Quadruplex and the flanking duplex DNA regions. A recent report by Yang et al. has found that the c-MYC i-Motif can adopt at least two structures similar to the one modeled here, further adding to the complexity of the NHE structure(s) and solution equilibria (Yang and Hurley 2007, unpublished data).

The primary importance of the modeled NHE structures is that they provide a look into a potential biological target for anti-cancer therapeutics. Adding additional complexity to the situation is the fact that obtaining a crystal or NMR structure of the complete NHE-III<sub>1</sub> silencer will be extremely difficult due to the size and structural diversity of the NHE. Therefore, molecular modeling studies provide a reasonable alternative for the analysis of such a structure.

These modeling studies have also provided a rapid and efficient analysis of six different binding sites for potential

drugs targeting this structure. Although several studies have analyzed these potential binding sites [3, 30, 33], one site of particular importance is the thymidine triplet region [34] at the interface between the i-Motif and G-Quadruplex structures. Porphyrin compounds may not be the ideal drug candidates, primarily due to their potent toxicity [35], as well as violating multiple parts of Lipinski's Rules of Five [36]. However, TMPyP4 and its analogs have proven very useful in the study of compounds that selectively bind to G-Quadruplex structures, and it is hoped that information on its binding, as well as better three-dimensional structural data on the NHE-III<sub>1</sub> silencer element, can lead to better drug candidates.

**Acknowledgments** The authors would like to thank Drs. Jonathan (Brad) Chaires of the University of Louisville, James G. Brown Cancer Center, and David Wilson of Georgia State University, Department of Chemistry and Biochemistry for their many helpful discussions. Funding for this research was received from Northern Arizona University Proposition 301 TRIF Grant to E.A.L. and Arizona Biomedical Research Commission (0014 to L.H.H. and E.A.L. and 0015 to E.A.L. and L.H.H.).

#### References

- Hurley LH (2002) *Nature* 2:188–200
- Slapak CA, Kufe DW (1998) Principles of cancer therapy. In: Isselbacher KJ et al (eds) *Harrison's principles of internal medicine*, 14th edn. McGraw-Hill, New York, pp 523–537
- Freyer MW, Buscaglia R, Kaplan K, Cashman D, Hurley LH, Lewis EA (2006) *Biophys J* 92(6):2007–2015
- Han H, Hurley LH (2000) *Trends in Pharmaceutical Sciences* 21(4):136–142
- Hurley LH (2001) *Biochem Soc Trans* 29:692–696
- Hurley LH, Wheelhouse RT, Sun D, Kerwin SM, Salazar M, Fedoroff OY, Han FX, Izbicka E, von Hoff DD (2000) *Pharmacological Therapeutics* 85:141–158
- Mergny JL, Helene C (1998) *Nat Med* 4:1366–1367
- Mills M, Lacroix L, Arimondo PB, Leroy JL, Francois JC, Klump H, Mergny JL (2002) *Curr Med Chem Anticancer Agents* 2:627–644
- Mergny JL, Riou JF, Mailliet P, Teulade-Fichou MP, Gilson E (2002) *Nucleic Acids Res* 30:839–865
- Todd AK, Johnston M, Neidle S (2005) *Nucleic Acids Res* 33(9):2901–2907
- Gartel AL, Shchors K (2003) *Exp Cell Res* 283:17–21
- Pelengaris S, Khan M (2003) *Arch Biochem Biophys* 416:129–136
- Marcu KB, Bossone SA, Patel AJ (1992) *Ann Rev Biochem* 61:809–858
- Pelengaris S, Rudolph B, Littlewood T (2000) *Curr Opin in Genetic Devel* 10:100–105
- Slamon DJ, deKernion JB, Verma IM, Cline MJ (1984) *Science* 224:256–262
- Cooney M, Czernuszewicz G, Postel EH, Flint SJ, Hogan ME (1988) *Science* 241:456–459
- Michelotti EF, Tomonaga T, Krutzsch H, Levens D (1995) *J Biol Chem* 270:9494–9499
- Postel EH, Berberich SJ, Rooney JW, Kaetzel DM (2000) *J Bioenerg Biomembranes* 32:277–284
- Simonson T, Pecinka P, Kubista M (1998) *Nucleic Acids Res* 26:1167–1172



20. Simonsson T, Pribylova M, Vorlickova M (2000) *Biochem Biophys Res Comm* 278:158–166
21. Tomonaga T, Levens D (1996) *Proc Natl Acad Sci USA* 93 (12):5830–5835
22. Collins I, Weber A, Levens D (2001) *Mol Cell Biol* 21(24):8437–8451
23. Ambrus A, Chen D, Dai J, Jones RA, Yang D (2005) *Biochemistry* 44:2048–2058
24. Siddiqui-Jain A, Grand CL, Bearss DJ, Hurley LH (2002) *Proc Natl Acad Sci USA* 99:11593–11598
25. Mathur V, Verma A, Maiti S, Chowdhury S (2004) *Biochem Biophys Res Comm* 320:1220–1227
26. Halder K, Mathur V, Chugh D, Verma A, Chowdhury S (2005) *Biochem Biophys Res Comm* 327:49–56
27. Li W, Miyoshi D, Nakano S, Sugimoto N (2003) *Biochemistry* 42:11736–11744
28. Risitano A, Fox KR (2003) *Biochemistry* 42:6507–6513
29. Grand CL, Han H, Munox RM, Weitman S, von Hoff DD, Hurley LH, Bearss DJ (2002) *Mol Cancer Ther* 1:565–573
30. Han FX, Wheelhouse RT, Hurley LH (1999) *JACS* 121:3561–3570
31. Read MA, Neidle S (2000) *Biochemistry* 39:13422–13432
32. Wheelhouse RT, Sun D, Han H, Han FX, Hurley LH (1992) *J Am Chem Soc* 114:3261–3262
33. Esmaili N, Leroy J-L (2005) *Nucleic Acids Res* 33:213–224
34. Hurley LH, Von Hoff DD, Siddiqui-Jain A, Yang D (2006) *Semin Oncol* 33:498–512
35. Koo MS, Ozawa T, Santos RA, Lamborn KR, Bollen AW, Deen DF, Kahl SB (2007) *J Med Chem* 50:820–827
36. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) *Adv Drug Deliv Rev* 46:3–26