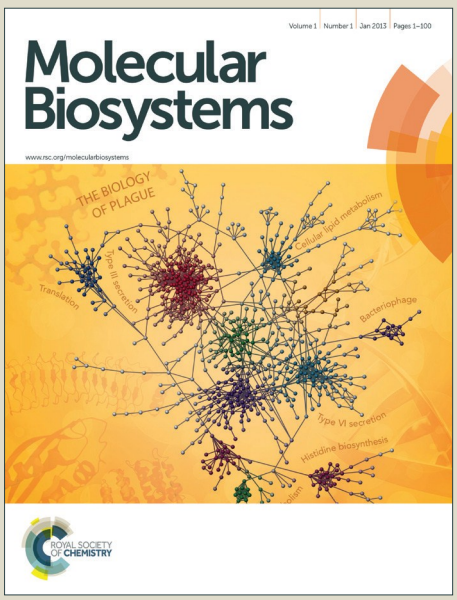
# Molecular BioSystems

Accepted Manuscript





ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

COMMUNICATION

# Reverse Watson-Crick G-G base pair in G-Quadruplex formation

Soma Mondal[a], Jyotsna Bhat[a], Jagannath Jana[a], Meghomukta Mukherjee[a] and Subhrangsu Chatterjee[a]*

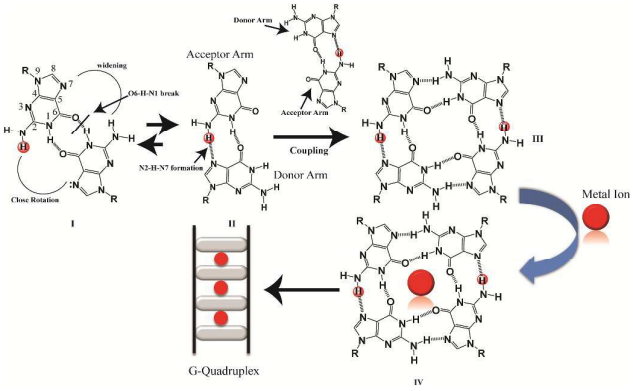
Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

**A stable intermediate dimeric G-rich form as a precursor of tetrameric G-quadruplex structures has been detected by MALDI-TOF Spectrometry. Molecular Dynamics Simulation offered detail insights at atomic level assigning reverse Watson-Crick G-G base pairing (not Hoogsteen) in the G-rich dimer. In support, cisplatin formed stable adduct binding to dimeric G-rich structure eliminating the possibility of G-G Hoogsteen hydrogen bond formation.**

Besides the canonical B-DNA that has right handed double helical structure stabilized by Watson-Crick base pairing[1], repetitive G-rich sequences can fold into non-canonical structure known as G-quadruplex. A G-quadruplex structure consists of π-π stacking of planar G-tetrads stabilized by Hoogsteen type hydrogen-bonds[2, 3]. G-rich sequences are abundantly distributed in telomeric region of human DNA and in promoter of many oncogenes[4-7], thus contributing a significant role in many biological processes[8, 9]. G-quadruplexes can be formed from single strands (intramolecular), two strands (bimolecular) and four strands (tetramolecular)[10]. Furthermore, intermolecular G-quadruplexes are gaining importance nowadays for engineering of nano-molecular devices owing to their stable and rigid three dimensional scaffolds and inherent electronic properties[11-14]. In addition, intermolecular G-quadruplex structures bring in frontier concepts in developing anti-cancer[15] and anti-HIV therapeutics[16] for near future. Among all the intermolecular G-quadruplex structures, tetramolecular G-quadruplex structure emerged as a new vision due to their symmetry and thermodynamic stability. Many studies related to the mechanism of tetramolecular G-quadruplex formation has been already investigated[17, 18] through kinetics[19, 20] and thermodynamic studies[21, 22]. Structures adopted by short G-rich sequences




Scheme S1.Proposed mechanism of G-Quartet formation from G-G Reverse Watson Crick G-G base-pairs

containing one flanking base at the 3' or 5' or both the ends have been shown to form parallel quadruplex structures[23]. As from the recent literature on intermolecular G-quadruplex formation it was observed that there has been no scientific clue, evidence on atomic level how an intermolecular G-quadruplex form. The possible G-G base pairing conformation, pathway of forming G-G Hoogsteen hydrogen bonded base pairing state in the three dimensional scaffolds of Quadruplexes have not been extensively mined out. Now here in this report we cultivated results which showed that monomeric to dimeric G-rich DNA formation can happen through G-G reverse Watson-Crick type hydrogen bonding. The closed, stabilized G-G reverse Watson-Crick base pairs in dimeric form open up to form active functional G-G Hoogsteen base pairs which act as a building block of G-quartet structure [Scheme 1]. Sharma et al has previously reported the participation of G-G reverse Watson-Crick hydrogen bond in multi-modality for edge interactions[24] in RNA. We studied four DNA sequences listed in [Table S1] using NMR, CD spectroscopy, gel electrophoresis and MALDI-TOF spectrometry. To support experimental findings we further performed molecular dynamics simulations. CD is mainly used to reveal information regarding the conformation of G-quadruplex structures. The CD spectra of all the GG4 sequences were recorded after overnight incubation at 4˚C [Figure S1(A),(C)]. CD spectra of GG4(A) shows a positive peak around 260 nm characteristics of parallel quadruplex, whereas GG4(T) shows a positive peak around 285 nm with a

[a] Dr. Subhrangsu Chatterjee
*Bose Institute, Centenary Campus, P-1/12 CIT Scheme VIIM, Kankurgachi, Department of Biophysics, Kolkata-54*
*http://bic.boseinst.ernet.in/subhro_c/*
*E-mail: subhro_c@jcbose.ac.in,*
*Phone: 0091-33-2569-3340*

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

minima around 261 nm which indicates antiparallel type structure[25]. The CD spectra of GG4(C) displays two positive peaks one at 285 nm and the other at 256 nm which is the characteristic peak of C rich duplex DNA[26]. Interestingly, for G rich sequences Laure et al.[27] has shown that for the sequences where 7 ≤ n ≤ 13( n= no .of guaninyl residues) the G-tract remains unfolded, therefore though GG4(G) shows a positive peak around 260 nm it remains single stranded. GG4(AT) exhibits a positive peak around 262 nm, quite different from the summation of GG4(A) and GG4(T) CD spectra which indicates the existence of other conformations including that of the GG4(A) and GG4(T). On the contrary, GG4(GC) displays a positive peak around 260 nm which is similar as the

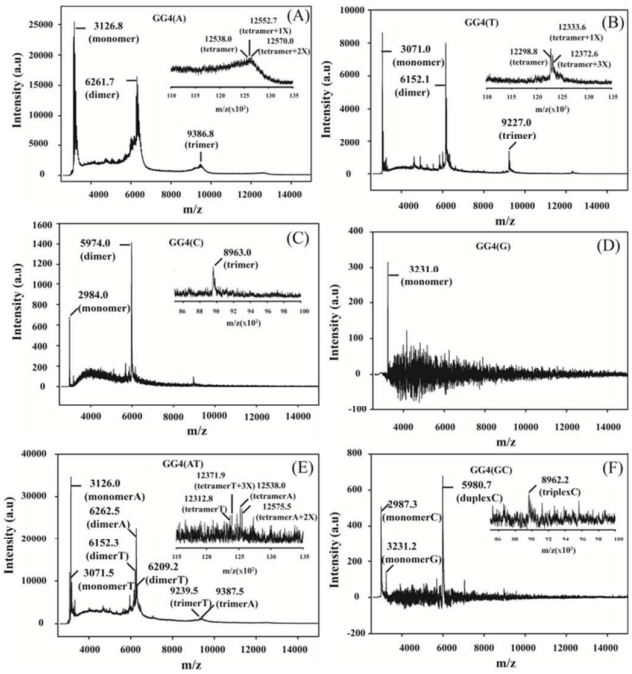



**Figure 1.**MALDI-TOF spectra of GG4 sequences after two months incubation at 4˚C using 3-HPA as matrix (A)GG4(A), inset [zoomed region of GG4(A) tetramer] (B)GG4(T), inset [zoomed region of GG4(T) tetramer] (C)GG4(G)(D)GG4(C), inset [zoomed region of GG4(C) triplex] (E)GG4(AT), inset [zoomed region of GG4(AT) tetramer] (F)GG4(GC), inset [zoomed region of GG4(C) triplex].X= ammonium ion.

summation of ellipticities of both GG4(C) and GG4(G) indicating the presence of duplex conformations of GG4(C) and single stranded conformation of GG4(G). Next, the CD spectra of GG4 sequences were taken after two months incubation at 4˚C [Figure S1(B),(D)]. GG4(A) and GG4(T) both show positive peak near 260 nm accompanied by an enhancement in CD signal, thus GG4(T) undergoes structural transition from antiparallel type structure to parallel stranded G-quadruplex structure. In contrast, the spectral patterns for GG4(C), GG4(G) and GG4 (GC) remain unaltered after two months indicating presence of stable duplex and single strand. In case of GG4(AT), after two months of incubation the CD spectra showed a positive peak near 262 nm, strikingly this spectra resembles the spectra generated from the summation of individual ellipticities of GG4(A) and GG4(T), this observation directly confirms the existence of primary conformations of GG4(A) and GG4(T) in their mixtures. MALDI-TOF mass spectrometry[28, 29] was employed for these sequences to unravel the intermediates that lead to build stable quartet conformation of GG4 sequences with time. MALDI-TOF spectra were recorded after the sequences were annealed in ammonium citrate dibasic and incubated at 4˚C for overnight with successive time intervals of 1 month and 2 month respectively. G-quadruplex conformation remains the same in both $K^+$ and $NH4^+$ solution [Figure S2]. When MALDI-TOF spectra were recorded after overnight incubation we find an equilibrium between monomer and dimer for all the sequences except GG4(G) which exists as a monomer [Figure S3 and Table S3]. Interestingly, for GG4(C) a stable trimer (triplex) is also detected. Mixed strand stoichiometry is detected for GG4(AT) dimer [Figure S3(E) inset]. In GG4(GC) we observed an equilibrium between GG4(C) monomeric and duplex forms but no mixed duplex is detected involving monomeric GG4(G)and GG4(C). Spectra recorded after 1 month incubation reveals formation of trimer (less amount) for all sequences along with monomer and dimer except for GG4(G) which still exists as monomer [Figure S4 and Table S4]. Mixed strand stoichiometry is again detected for GG4(AT) dimer [Figure S4(E) inset].GG4(GC) reveals no mixed conformation. The spectra recorded after two months incubation reveals the formation of tetramers with very low intensity for GG4(A), GG4(T) and GG4(AT) [Figure 1(A) inset, 1(B) inset, 1(E) inset] in MALDI-TOF spectrometry. When GG4(A) tetramer formation is just initiated GG4(T) comprised a proper tetramolecular G-quadruplex bound to three ammonium ions. In addition we detect some slipped G-quadruplex structures[17, 18], consisting of none, one and two ammonium ions for the GG4(A), GG4(T) and GG4(AT) sequences. In course of time there is an enhancement in signal in MALDI footprints for the dimers and trimers of all the sequences [Figure 1 and Table S5]. Now in MALDI-TOF spectrometry GG4(A) and GG4(T) both showed presence of dimer much higher than that of the trimer as well as tetramer. Though intermolecular structures are highly concentration dependent one dimensional proton spectra were performed at 10˚C to find out the nature of secondary structures adopted by the GG4 series. The imino protons appear around 11.0 ppm [Figure S5(A)] for GG4(A), GG4(T) and GG4(AT) indicating formation of G-quadruplex structures[17]. For GG4(C) and GG4(GC) imino protons are detected around 13.0 ppm [Figure S5(B)] suggesting formation of duplex structures[30]. Single stranded conformation of GG4(G) is evidenced with no appearance of imino signals in down field region [FigureS5(B)]. Gel electrophoresis data revealed slow migration rate of GG4(A), GG4(T) and GG4(AT) followed by GG4(C) and GG4(GC) indicating the presence of G-quadruplex and duplex forms respectively. For GG4(G) there is no band seen indicating single stranded nature of GG4(G) [Figure S6] (Ethidium bromide does not bind with single stranded DNA). We further continued our studies with GG4(A), GG4(T) and GG4(AT) as these sequences has the potential to form G-quarduplex structures. The two most probable mechanism of tetramolecular G-quarduplex formation reported till date are (i) monomer-dimer-tetramer in which coupling of dimer is the rate determining step[19] (ii) monomer-dimer-trimer-tetramer where tetramer formation is the rate determining step[18].We mainly focussed on the dimeric structure obtained in high yield in the mass spectrometry whose structural elucidation is not very clear. We simulated the dimeric G-rich structure (duplex) consisting of G-G Hoogsteen base pairs keeping the strand orientation in parallel direction to provide detail insights at atomic level. In simulation we observed the rearrangement of hydrogen bonding pattern of G-G base pairs. In first phase , G-G bases get unpaired and guanine bases are opened up. During further production run G-G pairing is

re-established by means of reverse Watson-Crick hydrogen bonding[31] and remained stable throughout the period [Figure 2(A)], (structural transition can be clearly seen in AVI movie provided in the supporting information). It has already been reported that to form parallel duplex structure base pair in reverse Watson-Crick is more favoured than normal Watson-Crick bonding [32]. As seen in G)N2H-N7(Acceptor G) distance is represented in three dimensional energy landscape of conversion of Hoogsteen pair to Watson-Crick pair [Figure 2(D) inset]. Similar results are observed for GG4(A) and GG4(T). From the Molecular dynamics simulation in the explicit solvent for all dimeric G rich structures (GG4(A), GG4(T), GG4(AT)) it is very well understood that the reverse Watson-Crick hydrogen

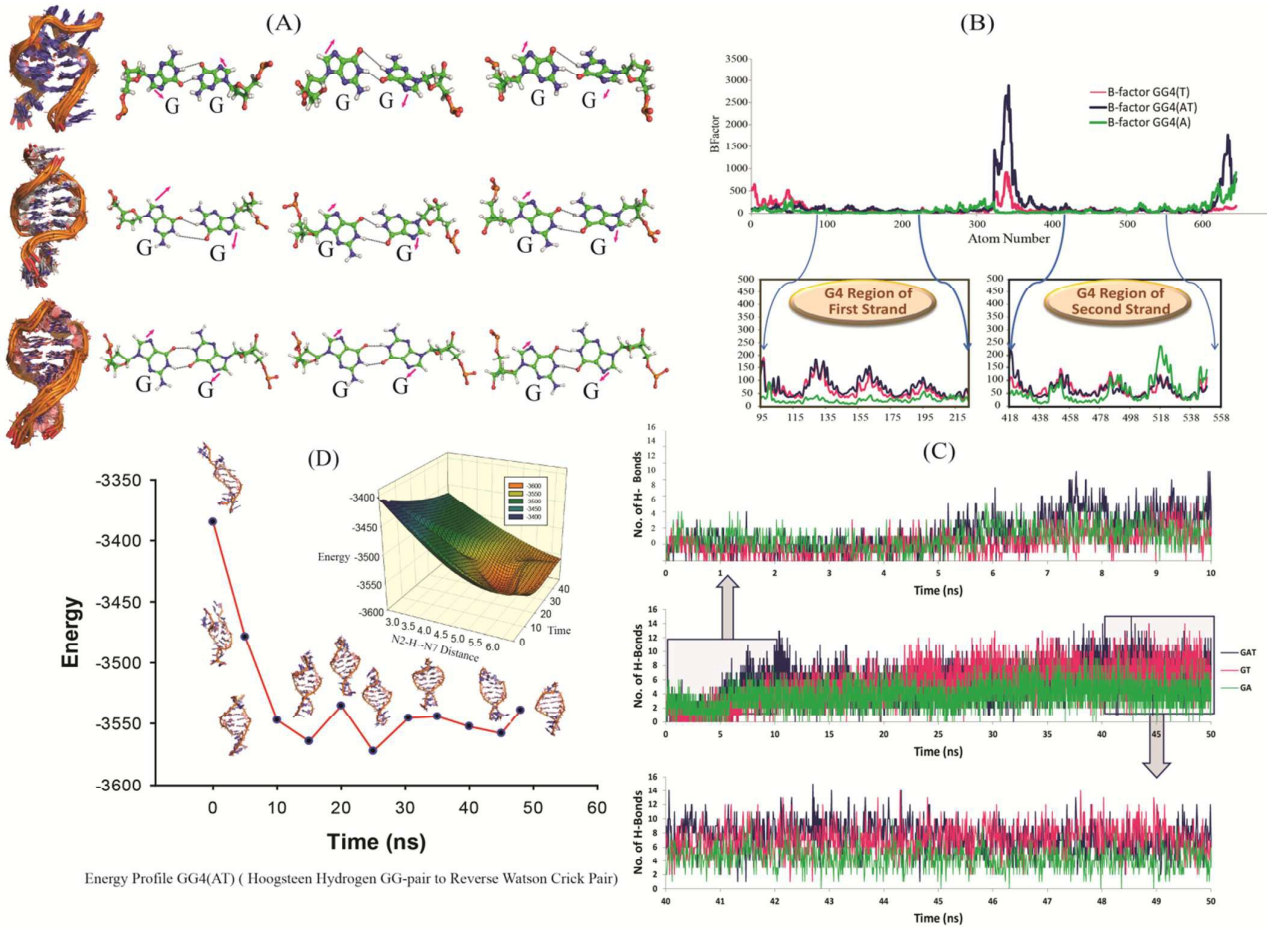



**Figure 2.(A)**Ensemble structure of GG4A(top), GG4T(middle) and GG4AT(bottom) along with reverse Watson Crick hydrogen bonding **(B)**B-factor estimation of each atom is illustrated in top region and G rich regions of first and second strand of GG4T, GG4AT and GG4A are enlarged for detailed understanding. **(C)**Total number of hydrogen bonds at each time step over entire simulation run(50ns), up arrow points at enlarged graph of number of hydrogen bonds during first 10 ns simulation run and down arrow points at enlarged graph of number of hydrogen bonds during last 10 ns simulation run **(D)**Change in total energy in GG4AT with respect to time and the corresponding change in N2H-N7 distance (inset)

[Figure 2(C)], for first 10 ns total number of hydrogen bonds are fewer in number [GG4(A):3, GG4(T):2 and GG4(AT):3(average values)] however, hydrogen bonding is increasing with the progression of simulation run. During 40-50 ns simulation period the total number of hydrogen bonds increased to GG4(A):4, GG4(T):7 and GG4(AT):8(average values). Also these hydrogen bonds are stable as illustrated by hydrogen bond occupancy values [Table S6]. B-factor analyses of all the atoms showed that the G-region of all the three systems are most stable compared to the terminal bases [Figure 2(B)].MMPBSA energy calculation suggested that the total energy content of the system is decreasing during the simulation thus resulting into more stable conformations [Table S7]. Change in total energy in GG4(AT) with respect to time is shown in [Figure 2(D), Figure S9] also the corresponding change in (Donor bonding involved in G-G base pairs stabilize the dimeric form of DNA. The simulation though was augmented with G-G Hoogsteen base pairing reverse Watson-crick G-G paired dimers become the end result. This observation was validated by MALDI-TOF of G-G dimer in presence of cisplatin (Molar mass 300 g/mol). As cisplatin binds to N7 atoms of two consecutive guanines[33] dimeric G-G pair has to be in reverse Watson-Crick hydrogen bonded conformation to facilitate the cisplatin DNA adduct formation. Here we could find a signal of G-G dimer bound to one and two cisplatin moiety shown in [Figure 4(A, B, C)] . This experimental observation is aligned with that of the theoretical molecular dynamics simulation. Interestingly in G-G pairing the imidazole ring of both the Gs are not involved in the base-pairing. If N1-H-O6 hydrogen bond in pairing is cleaved formation of N(2)-H-N7 bond simultaneously occur to save the

energy loss. This Watson-Crick pairing switched to Hoogsteen pairing opens one acceptor and a donor arm which ultimately couple with another G-G Hoogsteen base-pairing to form G

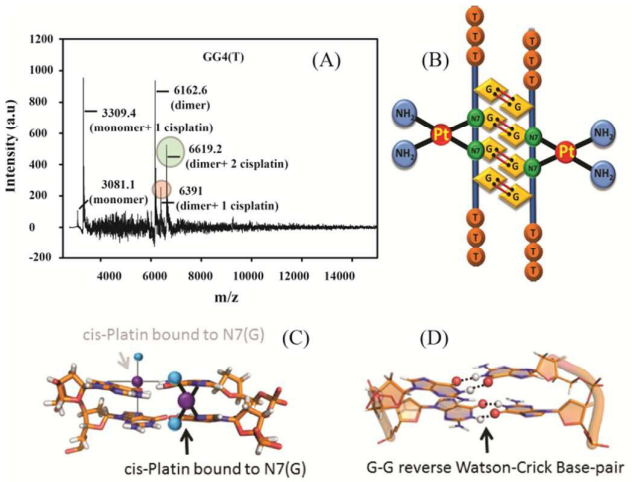



**Figure 3**.(A)MALDI of GG4(T) with cisplatin (B)Model structure of cisplatin bound to dimeric structure of GG4(T) (C)cisplatin bound to N7 of guanine (D)G-G reverse Watson-Crick base pair.

quadruplex structure [Scheme 1 ,S2]. The central N-H-O hydrogen bond being the backbone of the switch of O6-H-N1 to N2-H-N7 makes the coupling of the pairs of dimers (GG) to produce quadruplex. In a quadruplex structure there are two types of hydrogen bonds[34] (i) N1(Donor G)-H-O6 (Acceptor G) (ii) N7(Acceptor G) -H - N2(Donor G). The central N-H-O signal is sharp in NMR spectroscopy confirming the slow exchange of the central proton. On the contrary the peripheral N2-H-N7 does not appear in NMR confirming fast exchange.

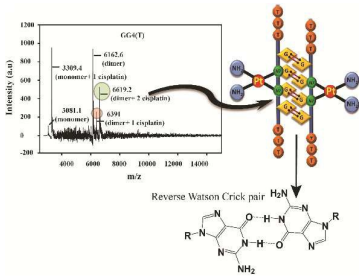
In short it is proven that the stable dimeric G rich form is dictated by G-G reverse Watson-Crick bonded base pairs. However, reverse Watson-Crick G-G pair cannot lead to form G quartet structures without switching to Hoogsteen pairing. This emphasizes $GG^{Reverse\ Watson\text{-}Crick}$ converts to $GG^{Hoogsteen}$ functional pairing which in turn couple with another $GG^{Hoogsteen}$ pair to form tetramolecular G-quarduplex structure. Thus $GG^{Reverse\ Watson\text{-}Crick}$ is conceptualized as the dormant, locked conformation of G-G base pair whereas $GG^{Hoogsteen}$ is the open, active, functional form.

## Notes and references

SC would like to thank DST Ramanujan Fellowship. SM, JB, and JJ would like to thank UGC, DBT and CSIR, Govt of India for their fellowship. SC thanks DBT research grant BT/PR6627/GBD/27/440/2012. SC thanks Mr. Amarendra Nath Biswas for carrying out MALDI-TOF experiments.

1. J. D. Watson and F. H. C. Crick, 1953.
2. J. Choi and T. Majima, *Chem Soc Rev*, **40**, 5893-5909.
3. A. T. Phan, V. Kuryavyi and D. J. Patel, *Curr Opin Struct Biol*, 2006, **16**, 288-298.
4. G. Biffi, D. Tannahill, J. McCafferty and S. Balasubramanian, *Nat Chem*, **5**, 182-186.
5. E. Y. Lam, D. Beraldi, D. Tannahill and S. Balasubramanian, *Nat Commun*, **4**, 1796.
6. A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc Natl Acad Sci U S A*, 2002, **99**, 11593-11598.
7. S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat Rev Drug Discov*, **10**, 261-275.
8. K. Hirashima and H. Seimiya, *Nucleic Acids Res*, **43**, 2022-2032.
9. M. C. Chen, P. Murat, K. Abecassis, A. R. Ferré-D'Amaré and S. Balasubramanian, *Nucleic Acids Res*, **43**, 2223-2231.
10. S. Burge, G. N. Parkinson, P. Hazel, A. K. Todd and S. Neidle, *Nucleic Acids Res*, 2006, **34**, 5402-5415.
11. F. A. Aldaye, A. L. Palmer and H. F. Sleiman, *Science*, 2008, **321**, 1795-1799.
12. P. Alberti, A. Bourdoncle, B. Saccà, L. Lacroix and J. L. Mergny, *Org Biomol Chem*, 2006, **4**, 3383-3391.
13. A. M. Chiorcea-Paquim, P. V. Santos, R. Eritja and A. M. Oliveira-Brett, *Phys Chem Chem Phys*, **15**, 9117-9124.
14. O. Doluca, J. M. Withers, T. S. Loo, P. J. Edwards, C. González and V. V. Filichev, *Org Biomol Chem*, **13**, 3742-3748.
15. P. J. Bates, D. A. Laber, D. M. Miller, S. D. Thomas and J. O. Trent, *Exp Mol Pathol*, 2009, **86**, 151-164.
16. A. Rajendran, M. Endo, K. Hidaka, P. L. Tran, J. L. Mergny, R. J. Gorelick and H. Sugiyama, *J Am Chem Soc*, **135**, 18575-18585.
17. C. Bardin and J. L. Leroy, *Nucleic Acids Res*, 2008, **36**, 477-488.
18. F. Rosu, V. Gabelica, H. Poncelet and E. De Pauw, *Nucleic Acids Res*, **38**, 5217-5225.
19. J. R. Wyatt, P. W. Davis and S. M. Freier, *Biochemistry*, 1996, **35**, 8002-8008.
20. J. L. Mergny, A. De Cian, A. Ghelab, B. Saccà and L. Lacroix, *Nucleic Acids Res*, 2005, **33**, 81-94.
21. J. Gros, A. Aviñó, J. Lopez de la Osa, C. González, L. Lacroix, A. Pérez, M. Orozco, R. Eritja and J. L. Mergny, *Chem Commun (Camb)*, 2008, 2926-2928.
22. P. L. Tran, A. De Cian, J. Gros, R. Moriyama and J. L. Mergny, *Top Curr Chem*, **330**, 243-273.
23. J. Zhou, G. Yuan, J. Liu and C. G. Zhan, *Chemistry*, 2007, **13**, 945-949.
24. P. Sharma, A. Mitra, S. Sharma, H. Singh and D. Bhattacharyya, *J Biomol Struct Dyn*, 2008, **25**, 709-732.
25. S. Paramasivan, I. Rujan and P. H. Bolton, *Methods*, 2007, **43**, 324-331.
26. M. Morikawa, K. Kino, T. Oyoshi, M. Suzuki, T. Kobayashi and H. Miyazawa, *Biomolecules*, **4**, 140-159.
27. L. Joly, F. Rosu and V. Gabelica, *Chem Commun (Camb)*, **48**, 8386-8388.
28. N. Nagesh, A. Krishnaiah, V. M. Dhople, C. S. Sundaram and M. V. Jagannadham, *Nucleosides Nucleotides Nucleic Acids*, 2007, **26**, 303-315.
29. L. Yuan, T. Tian, Y. Chen, S. Yan, X. Xing, Z. Zhang, Q. Zhai, L. Xu, S. Wang, X. Weng, B. Yuan, Y. Feng and X. Zhou, *Sci Rep*, **3**, 1811.
30. R. H. Sarma, M. H. Sarma, L. Dai and K. Umemoto, *FEBS Lett*, 1997, **418**, 76-82.
31. V. Kocman and J. Plavec, *Nat Commun*, 2014, **5**, 5831.
32. L. A. Yatsunyk, O. Piétrement, D. Albrecht, P. L. Tran, D. Renčiuk, H. Sugiyama, J. M. Arbona, J. P. Aimé and J. L. Mergny, *ACS Nano*, 2013, **7**, 5701-5710.
33. J. Reedijk, *Chem Rev*, 1999, **99**, 2499-2510.
34. A. J. Dingley, R. D. Peterson, S. Grzesiek and J. Feigon, *J Am Chem Soc*, 2005, **127**, 14466-14472.

**TOC Graphic**




**Cisplatin binds to N7 of Guanine in Reverse Watson Crick G-G Pair.**