

University of Groningen

A computational study on the nature of DNA G-quadruplex structure

Gholamjani Moghaddam, Kiana

DOI:
[10.33612/diss.159767021](https://doi.org/10.33612/diss.159767021)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Gholamjani Moghaddam, K. (2021). *A computational study on the nature of DNA G-quadruplex structure*. University of Groningen. <https://doi.org/10.33612/diss.159767021>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

2

Theoretical Approach

Computational chemistry is an area of chemistry that uses computer modelling and simulation to tackle different chemical problems. Quantum mechanics (QM) has been utilized as a computational technique for the static modeling of molecular systems. However, investigation of time-dependent behavior of molecules is important in chemistry and biology. Molecular dynamics (MD) simulations address dynamical issues and excel at exploring structure-function relationships, fluctuations and conformational changes of biological systems such as proteins and nucleic acids. MD simulations cannot describe chemical reactions because the chemical bond formation and breaking are ignored in this method. To study reaction mechanisms involving electronic rearrangements, QM methods are required. However, QM approaches are not suitable for systems as large as the size of DNA because of their computational expense. Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations are a popular technique for investigating biomolecular reactions by combining two approaches in computational chemistry and taking advantage of both approaches. In addition, investigation of important biological mechanisms such as protein folding, lipid-protein, and protein-DNA interactions by means of atomistic MD simulations becomes demanding when the size of the systems or the timescale of the considered process is increased. Coarse-grain (CG) simulations have been developed to explore larger systems on time scales inaccessible to atomistic MD simulations. This method speeds up simulations by reducing the number of degrees of freedom. This chapter introduces the main computational methods used in this thesis.

2.1. Quantum Mechanical Methods

Molecular quantum mechanics is an essential tool to study atomic and molecular properties on a microscopic scale by solving the electronic Schrödinger equation. In fact, the solution of this differential equation, expressed in terms of the positions of the nuclei and the number of electrons, provides useful information such as energy, electron density and other properties of the molecule. Two approaches to solve the electronic Schrödinger equation have been developed over the past 50 years including wave function-based approaches and density functional theory (DFT). The simplest method based on the wave function-based approach is the Hartree–Fock (HF) method. HF theory only takes into account the average electron–electron interactions and consequently ignores the correlation between electrons, i.e. the difference between the total exact non-relativistic energy and HF energy in the same basis. The electron correlation can be divided into two dynamical and non-dynamical (static) electron correlation. The dynamical correlation arises from the Coulomb repulsion between electrons which is described by the post-Hartree-Fock methods such as Configuration interaction (CI)⁵⁴, coupled-cluster approaches (CC)⁵⁴ and Møller-Plesset perturbation theory (MP)⁵⁴. In contrast, the non-dynamical correlation comes from the near-degeneracy of electronic configurations. Such cases like bond-breaking processes can be described by the multi-reference methods such as multiconfigurational perturbation theory (CASPT)⁵⁵ and multiconfigurational self-consistent field (MCSCF)⁵⁶. Another method for improving the HF method is DFT, which takes into account an approximate treatment of the electron correlation with the advantage of being less demanding computationally compared to the above-mentioned methods.

In this thesis, we applied the DFT method for the ground state calculations. For the excited states calculations we used time-dependent density functional theory (TDDFT) and spin-flip time-Dependent density functional theory (SF-TDDFT) methods. The following sections will briefly introduce these methods used in this thesis.

2.1.1. Density Functional Theory

DFT^{57,58} is one of the most known quantum chemistry methods, not only because of its accuracy, but also because of its relatively less demanding computational cost than HF theory. DFT method takes into account electron correlation that is neglected in the HF method.

In DFT, the ground-state energy of an n -electron system can be described in terms of

the electron density, $\rho(r)$, instead of a wave function. This theory was developed by Kohn and Sham providing a practical tool of DFT calculations.

In the Kohn-Sham formalism, the ground state energy, E , can be expressed as

$$E = E_T + E_V + E_J + E_{XC} \quad (2.1)$$

E_T , E_V , E_J and E_{XC} are the kinetic energy, electron–nuclear interaction energy, Coulomb self-interaction of the electron density, $\rho(r)$, exchange–correlation energy, respectively.

The kinetic energy of an n -electron system, E_T , is expressed in terms of one-electron spatial orbitals ($\psi_i (i = 1, 2, \dots, n)$) called the Kohn–Sham orbitals as

$$E_T = \sum_{i=1}^n \left\langle \psi_i \left| -\frac{1}{2} \hat{\nabla}^2 \right| \psi_i \right\rangle \quad (2.2)$$

The electron–nuclear attraction energy can be written as sum over all M nuclei with atomic number Z_A and coordinate R_A .

$$E_V = - \sum_{A=1}^M Z_A \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_A|} d\mathbf{r} \quad (2.3)$$

The Coulomb interaction between two electron densities $\rho(\mathbf{r})$ at r_1 and r_2 is defined as

$$E_J = \frac{1}{2} \left\langle \rho(\mathbf{r}_1) \left| \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right| \rho(\mathbf{r}_2) \right\rangle \quad (2.4)$$

The exchange–correlation energy of the system, is represented as a functional of the density,

$$E_{XC} = \int f[\rho(\mathbf{r}), \hat{\nabla}\rho(\mathbf{r}), \dots] \rho(\mathbf{r}) d\mathbf{r} \quad (2.5)$$

The exact ground-state electron density is the sum over all occupied one-electron spatial orbitals ($\psi_i (i = 1, 2, \dots, n)$) (KS orbitals).

$$\rho(\mathbf{r}) = \sum_{i=1}^n |\psi_i(\mathbf{r})|^2 \quad (2.6)$$

The KS equations for the one-electron orbitals $\psi_i(r_1)$ can be written as

$$\left\{ -\frac{1}{2} \hat{\nabla}_i^2 - \sum_{A=1}^M \frac{Z_A}{|r - R_A|} + \int \frac{\rho(r_2)}{|r_1 - r_2|} dr_2 + V_{XC}(r_1) \right\} \psi_i(r_1) = \varepsilon_i \psi_i(r_1) \quad (2.7)$$

where ε_i , V_{XC} are the KS orbital energies and the exchange–correlation potential, respectively. V_{XC} is defined as functional derivative of $E_{XC}[\rho]$ with respect to $\rho(r)$:

$$V_{XC}[\rho] = \frac{\delta E_{XC}[\rho]}{\delta \rho} \quad (2.8)$$

All terms in KS equation are known except the V_{XC} term. Numerous functional forms have been developed for the exchange–correlation energy. This functional is often divided into an exchange and a correlation functional.

The simplest exchange–correlation functional, Local Density Approximation (LDA), is written as

$$E_{XC} = \int \rho(\mathbf{r}) \varepsilon_{XC}[\rho(\mathbf{r})] d\mathbf{r} \quad (2.9)$$

where $\varepsilon_{XC}[\rho(\mathbf{r})]$ is the exchange–correlation energy per electron in the uniform electron gas model. This functional is highly inaccurate where the electron density varies in the system such as in a molecule on an extended system. To improve the LDA functional, it is necessary to include the density gradient. Generalized gradient approximation (GGA) functional depends on the local density and its gradient, while the meta-GGA includes local density, its gradient, and its second derivative. Furthermore, hybrid functionals include a fraction of HF exchange (the exchange energy given by HF theory) to improve the performance. The most well-known hybrid functionals in DFT are B3LYP, PBE0, HSE, etc. In addition, in recent years many methods have been developed for describing non-covalent interactions, specially dispersion which can be combined with all DFT functionals.

2.1.2. Time-Dependent Density Functional Theory

TDDFT^{59,60} is one of the most widely used approaches in quantum chemistry for the calculation of electronic excitation energies, excited-state geometries, absorption spectra, etc. The time-dependent Kohn–Sham equations in TDDFT are single-particle equations which resemble those of eq. 2.7

$$\left\{ -\frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{|r - R_A|} + \int \frac{\rho(r_2, t)}{|r_1 - r_2|} dr_2 + V_{ext} + V_{XC}(r_1, t) \right\} \psi_i(r_1, t) = i\hbar \frac{\partial}{\partial t} \psi_i(r_1, t) \quad (2.10)$$

in which the density is given as

$$\rho(r, t) = \sum_{i=1}^n |\psi_i(r, t)|^2 \quad (2.11)$$

where the external potential V_{ext} , the exchange–correlation potential V_{XC} , the KS orbitals and the density are all time-dependent. The first approximation to the exchange–

correlation functional is the so-called adiabatic local density approximation (ALDA) in which the originally non-local time-dependent xc kernel is replaced with a time-independent local one. The ALDA approximation allows us to use a standard local ground-state xc potential in the TDDFT framework. Over the past decade, linear-response TDDFT (LR-TDDFT) in combination with exchange-correlation functionals, has become a known approach for studying the excited state properties of molecular systems. In short, the excitation energies, Ω , can be obtained as solution of an eigenvalue equation:

$$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} = \Omega \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (2.12)$$

where the coupling matrices A and B can be expressed as

$$A_{ia,jb} = (\epsilon_a - \epsilon_i)\delta_{ab}\delta_{ij} + \langle ib|aj \rangle - C_x \langle ib|ja \rangle + \langle ib|f^{xc}|aj \rangle \quad (2.13)$$

and

$$B_{ia,jb} = \langle ij|ab \rangle - C_x \langle ij|ba \rangle + \langle ij|f^{xc}|ab \rangle, \quad (2.14)$$

where $i, j \dots$ and $a, b \dots$ denotes occupied and unoccupied spin-orbitals, respectively. C_x refer to the fraction of HF exchange in the exchange correlation functional f^{xc} and two-electron integrals can be expressed as

$$\langle pq|rs \rangle = \int \int dr_1 dr_2 \Phi_p(r_1)\Phi_q(r_2) \frac{1}{r_{12}} \Phi_r(r_1)\Phi_s(r_2) \quad (2.15)$$

$$\langle pq|f^{xc}|rs \rangle = \int \int dr_1 dr_2 \Phi_p(r_1)\Phi_q(r_2) f^{xc}(r_1, r_2) \Phi_r(r_1)\Phi_s(r_2) \quad (2.16)$$

Standard TDDFT has well-known failure in description of static correlation effects because it relies on a single reference configuration of KS orbitals. In 2003, a new variation of TDDFT called spin-flip (SF) DFT⁶¹ was developed to tackle this issue which will be described in the following section.

2.1.3. Spin-Flip Time-Dependent Density Functional Theory

In SF-TDDFT methods^{61,62}, a high-spin triplet state with two unpaired α -electrons ($M_s = 1$) is used as a reference state. In order to obtain $M_s = 0$ for the target states, only $\alpha\beta$ blocks in LR-TDDFT are considered. Considering the orthogonality between the occupied α and β spin-orbitals, i.e. $\langle ib|aj \rangle = 0$ and also using a collinear exchange-correlation functional kernel approximation, i.e. $\langle ij|f^{xc}|ab \rangle = 0$, eqs 2.13 and 2.14 can be expressed as

$$A_{ia,jb} = (\epsilon_a - \epsilon_i)\delta_{ab}\delta_{ij} - C_x \langle ib|ja \rangle \quad (2.17)$$

and

$$B_{ia,jb} = 0 \quad (2.18)$$

Therefore, using a collinear kernel of SF-TDDFT is similar to LR-TDDFT within the Tamm-Dancoff approximation (TDA)⁶³ and only the following eigenvalue equation is necessary to be solved

$$AX = \Omega X \quad (2.19)$$

SF-TDDFT is an effective electronic structure method for describing bond-breaking⁶⁴, polyradicals^{65,66}, conical intersections^{67,68}, excited state processes^{69,70}, singlet fission phenomena⁷¹⁻⁷³. However, there is a serious drawback in SF-TDDFT which is in the form of spin contamination. In the SF-TDDFT method, the excited states are obtained by one-electron $\alpha \rightarrow \beta$ spin flip transitions from a triplet reference state. Considering a model including four electrons in four orbitals (one close shell, two open-shell and one virtual ones), only flip-down transitions within the open-shell orbitals can be properly described in the SF-TDDFT framework. However, other types of transitions can lead to spin-contaminated states. This drawback causes severe problems especially when the states cross and in such cases state-tracking techniques are required to improve the performance of SF-TDDFT. Apart from a state-tracking algorithm, the spin adapted (SA)SF-TDDFT method, being free of spin contamination, is a safer solution⁷⁴.

2.2. Atomistic Molecular Dynamics Simulation

Atomistic molecular dynamics (MD) simulation is a powerful technique which captures the dynamical behavior of a system in full atomic details. In order to calculate the dynamics of the system of N interacting atoms, the classical Newtonian equations of motion are solved according to the forces derived from a potential energy:

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \quad (2.20)$$

where F_i refers to the force that acts on each atom with mass m_i and position r_i .

In general, it is impossible to specify the properties of complex molecular systems analytically. MD simulation solves this issue by using numerical methods. For a molecule

with an available geometry and a set of velocities, the forces on each atom are computed and then new positions for the atoms are generated, from which an updated set of forces is calculated. Repetition of these steps produces a trajectory describing the time-evolution of the system.

The interactions between particles are described by a potential energy function, known as a force field, which is separated into terms representing bonded and non-bonded interactions (see Figure 2.1). The bonded interactions are represented by harmonic potentials for the bond stretching between two atoms i and j (r_{ij}), bond-angle vibration between atoms i , j and k (θ_{ijk}) and for the improper dihedral angle (ξ_{ijkl}) between planes (i,j,k) and (j,k,l), and by a cosine-based potential for the dihedral angles (ϕ_{ijkl}).

$$V_{bond}(r_{ij}) = \frac{1}{2} k_b (r_{ij} - b_0)^2 \quad (2.21)$$

$$V_{angle}(\theta_{ijk}) = \frac{1}{2} k_\theta (\theta_{ijk} - \theta_0)^2 \quad (2.22)$$

$$V_{dihedral}(\phi_{ijkl}) = k_\phi (1 + \cos(n\phi_{ijkl} - \phi_0)) \quad (2.23)$$

$$V_{improper}(\xi_{ijkl}) = \frac{1}{2} k_\xi (\xi_{ijkl} - \xi_0)^2 \quad (2.24)$$

where k is the harmonic force constant. b_0 , θ_0 , ϕ_0 and ξ_0 refer to the equilibrium values of the bond and angles, and n is the multiplicity of the dihedral angle potential.

The non-bonded interactions are represented by Coulomb's law and Lennard-Jones (LJ) potential for the electrostatic and Van der Waals interactions, respectively.

$$V_{Coulomb}(r_{ij}) = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_i q_j}{r_{ij}} \quad (2.25)$$

where q_i and q_j denote charges of two atoms at a distance r_{ij} . ϵ_0 and ϵ_r are the vacuum permittivity and relative permittivity, respectively.

$$V_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.26)$$

where the ϵ_{ij} and σ_{ij} are the strength and range of the interaction between particle i and j , respectively.

The parameters in the force fields are fitted to the quantum mechanical (QM) calculations and/or experimental data. In the biomolecular area, the most commonly used force fields are AMBER⁷⁵, CHARMM⁷⁶, GROMOS⁷⁷ and OPLS⁷⁸. In general, these force fields

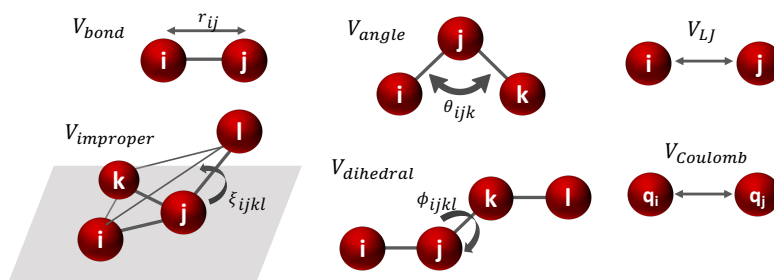


Figure 2.1 | Schematic representation of the bonded and non-bonded interactions.

2

differ in the data and procedure applied in their parametrization and in which particular systems they can describe more accurately.

2.3. Hybrid Quantum Mechanics/Molecular Mechanics Simulation

Hybrid quantum mechanics/molecular mechanics (QM/MM) approaches have become popular for modeling of chemical reactions in a big molecular system in which the system is divided into two parts, i.e. QM and MM regions⁷⁹. The QM region includes the active part of the system and they are treated explicitly at the level of QM theory, whereas, the remainder is considered as the MM region and described by a MM force field (see Figure 2.2).

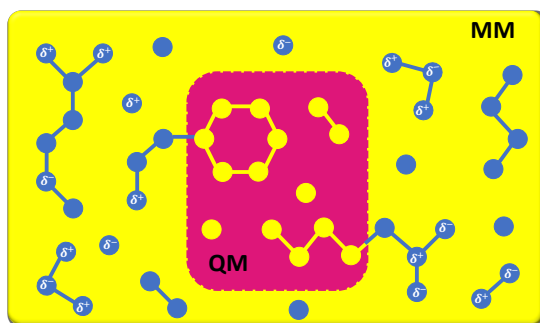


Figure 2.2 | Representation of the QM/MM concept. A small region, in which a chemical reaction occurs is treated at the level of QM theory. The remainder of the system is described by a MM force field.

The QM/MM potential energy includes interactions between atoms in the QM and

MM regions which are relatively straightforward to describe, and interactions between QM and MM atoms which are more difficult to describe. The more widely used approach for describing the interactions between the two regions is an additive coupling scheme in which the potential energy of a system can be expressed as

$$V_{QM/MM} = V_{QM} + V_{MM} + V_{QM-MM} \quad (2.27)$$

V_{QM} is the energy of the QM region and V_{MM} represents the energy of MM region calculated using a classical force field. The last term V_{QM-MM} is the interaction between the QM and MM regions including the bonded interactions at the boundary and also the electrostatic and van der Waals (vdW) interactions,

$$V_{QM-MM} = V_{QM-MM}^{bonded} + V_{QM-MM}^{vdW} + V_{QM-MM}^{el} \quad (2.28)$$

Electrostatic interactions between the QM and MM regions can be described by either mechanical embedding, electrostatic embedding or polarized embedding which will be explained in the following sections.

2.3.1. Mechanical Embedding

In the mechanical embedding, all interactions between the QM and MM regions are treated at the MM level. The MM charge model is applied to the QM atoms. In this scheme, there are no interactions between the charges in the MM region with the QM density which is problematic because the QM density is not polarized by the MM region. The vdW parameters of the QM atoms are included in the MM non-bonded energy described by a Lennard-Jones potential in Eq. 2.29.

$$V_{QM-MM} = \sum_m^{N_{MM}} \sum_i^{N_{QM}} \epsilon_{mi} \left[\left(\frac{\sigma_0}{R_{mi}} \right)^{12} - 2 \left(\frac{\sigma_0}{R_{mi}} \right)^6 \right] \quad (2.29)$$

where, R_{mi} represents the distance between QM atom i to MM atom m . ϵ_{am} and σ_0 are the standard Lennard-Jones parameters.

2.3.2. Electrostatic Embedding

The major deficiency of the mechanical embedding can be avoided by including the electrostatic interactions as one electron operators in QM Hamiltonian. In such an electrostatic embedding scheme, the QM region is polarized by MM charges and thus the QM atoms

feel the electric potential due to the MM atoms which provide a more accurate description of electrostatic interactions between QM and MM regions compared to the mechanical embedding scheme. Therefore, the V_{QM-MM}^{el} term in Eq. 2.28 can be expressed as

$$V_{QM-MM}^{el} = - \sum_{i,m} \frac{q_m}{R_{im}} + \sum_{\alpha,m} \frac{Z_\alpha q_m}{R_{\alpha m}} \quad (2.30)$$

where, q_m are the charges of MM atoms, R_{im} is a distance between an MM external charge of atom m and a QM electron i , Z_α are QM nuclei and $R_{\alpha m}$ represents the distance between QM atom α to MM atom m .

2

2.3.3. Polarized Embedding

The next refinement is to introduce the polarization of the MM region by the QM region, called polarized embedding. In order to calculate the total energy in the polarized embedding scheme, the MM polarizations need to be determined in a self-consistent fashion. This scheme of course needs a polarizable force field to model the MM polarizability by the QM region which increases the computational cost.

It should be mentioned that we applied the electrostatic embedding scheme for our QM/MM simulations as described in chapter 5.

2.4. Coarse-Graining

Coarse-graining (CG) models have been used extensively in biomolecular simulations to study larger systems at longer time scales by reducing the number of degrees of freedom in the system⁸⁰⁻⁸². By grouping a few atoms into virtual particles, often called beads, the number of interactions reduces, resulting in smoothening of the energy landscape, thus speeding up the CG simulations compared to the corresponding atomistic simulations. A wide range of approaches have been developed to coarse-graining which can be classified into two approaches: systematic and building block⁸¹⁻⁸³. The systematic or bottom-up approach focuses on the parametrization of the interactions based on the atomistic structural details. However, building block approaches which follow a top-down approach, take macroscopic features (e.g., thermodynamic data) as the main target of their parametrization. Top-down CG models are more transferable and cheaper compared to the bottom-up CG models. Many successful CG force fields combine both approaches^{81,82}.

The most famous model of the building block approach is the Martini model. Here we will briefly discuss this CG model which will be used throughout this thesis. The Martini

force field^{84,85} is one of the most widely used CG models suited for simulations of biomolecular systems. The Martini force field applies a chemical building block mapping of two to four non-hydrogen atoms to beads as the interaction centers (see Figure 2.3). Four main types of beads have been defined: polar (P), non-polar (N), apolar (C) and charged (Q). These bead types have a number of subtypes based on the hydrogen-bonding capabilities (d = donor, a = acceptor, da = both, 0 = none) or the degree of polarity (from 1, low polarity, to 5, high polarity), giving a total of 18 different bead types. In Martini force field, bonded interactions which include bonds, angles and dihedrals are optimized based on atomistic simulations in a bottom-up approach. The non-bonded interactions are described by using a Lennard-Jones potential, targeting partitioning free energies of solutes between water and organic solvents, and also the densities of liquids in a top-down approach. In addition, the electrostatic interactions between charged beads are described by using a Coulomb potential. The building block mapping approach makes the Martini easily extensible and different classes of molecules compatible with each other. The Martini model was originally developed for lipid simulations^{85,86} and then created for proteins^{87,88}, polymers⁸⁹⁻⁹¹, carbohydrates⁹², carbon nanoparticles⁹³, DNA⁹⁴ and other molecules.

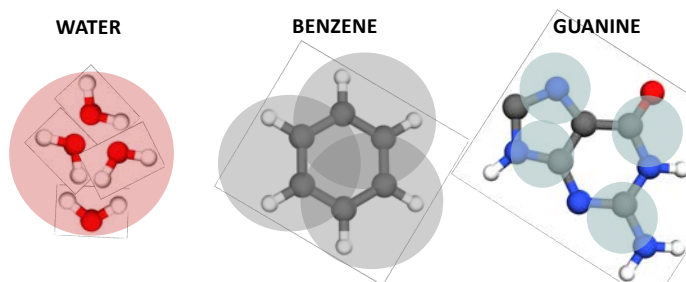


Figure 2.3 | Representations of atomistic chemical structures and their CG mapping of Martini model for water, benzene and guanine. The CG beads are depicted as transparent vdW spheres.