Journal of Molecular Liquids 201 (2015) 68–76

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq

Calculation of binding free energy of short double stranded oligonucleotides using MM/3D-RISM-KH approach

Joshy P. Yesudas [a,c], Nikolay Blinov [b,c], Steven K. Dew [a], Andriy Kovalenko [b,c,*]

[a] Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada
[b] Department of Mechanical Engineering, University of Alberta, Edmonton, AB T6G 2G8, Canada
[c] National Institute of Nanotechnology, Edmonton, AB T6G 2M9, Canada

ARTICLE INFO

Article history:
Received 28 February 2014
Received in revised form 25 October 2014
Accepted 8 November 2014
Available online 11 November 2014

Keywords:
DNA
Solvation free energy
Binding free energy
Molecular dynamics simulation
3D-RISM-KH molecular theory of solvation

ABSTRACT

We calculated the binding free energy of short oligonucleotides (9–20-mers DNA) by using molecular dynamics, followed with post-processing based on molecular theory of solvation (MM/3D-RISM-KH approach). A comparison with the PBSA and GBSA continuum solvation models was performed to identify the approach in the best agreement with experimental results for the binding free energy. Compared to the PBSA or GBSA methods, the 3D-RISM-KH molecular theory of solvation provides a more accurate description of the nonpolar contribution to the solvation free energy from the first principles of statistical mechanics. The binding free energy was calculated by using separate trajectories for the DNA complex and its two strands, as well as based on a single trajectory for the complex, both with and without account for explicit counter ions in post-processing of molecular dynamics trajectories to calculate the binding free energy of oligonucleotides. Overall, both GBSA and 3D-RISM-KH predict the binding free energy obtained from "separate trajectory" calculations with implicit account for counter ions in a good agreement with experiment, the latter showing a better performance for larger oligonucleotides.

1. Introduction

The molecular recognition properties of base pairing make nucleic acids potential candidate for novel biotechnology applications [1–3]. The oligonucleotides are increasingly being used for many analytical and diagnostic techniques such as microarray technology [4,5] and deoxyribonucleic acid (DNA) based molecular electronics [6–8]. The highly specific self-assembling nature of DNA, together with its properties such as high chemical stability and ease in synthesis, aids in self-assembly of nanoparticles [9–12]. The higher binding free energy [13] of complementary DNA strands compared to the binding of mismatched bases provides selectivity of DNA assisted integration of micro- or nano-electronic devices. Careful control of DNA hybridization kinetics can expand the achieved selectivity. Stability of a DNA duplex structure can be predicted starting from DNA sequence by using thermodynamic parameters such as binding affinities of nucleic acids [14–16]. Methods (for example, the nearest neighbor model) [15,17,18] developed to estimate thermodynamics of DNA hybridization reaction cannot describe modified systems with potentially reduced number of degrees of freedom (such as DNA attached to the substrate in bionanotechnological applications).

Since most biological processes take place in aqueous environment, it is important to assess the effects of solvents on the thermodynamics of reactions involved in these processes. Molecular dynamics (MD) simulations with explicit solvent can adequately represent experimental conditions [19–25]; however, they are computationally time-consuming and sometimes unfeasible. On the other hand, continuum solvation models [26–29] involve empirical parameterization and may not properly treat the nonpolar contributions to the solvation free energy [30,31].

Statistical–mechanical, Ornstein–Zernike (OZ) integral equation theory of liquids contracts degrees of freedom of individual solvent molecules down to the density correlation functions [32]. The OZ integral equation for the total and direct correlation functions has to be complemented with another relation called "closure" which has an exact diagrammatic representation in terms of infinite series of multiple integrals. In practice it is replaced with amenable closure approximations [32]. The three-dimensional reference interaction site model (3D-RISM molecular theory of solvation) represents molecular solvation structure in terms of 3D maps of solvent site densities around a solute macromolecule or a supramolecular assembly of arbitrary shape [33–41]. An important component enabling utility of 3D-RISM method for complex chemical and biomolecular systems in solution has been the closure relation proposed by Kovalenko and Hirata (KH approximation) [37,40,41]. For simple and complex solvents and solutions of a given composition, including buffers, salts, polymers, ligands and other cofactors at a finite concentration, the 3D-RISM-KH molecular

* Corresponding author at: National Institute for Nanotechnology, 11421 Saskatchewan Drive, Edmonton, AB T6G 2M9, Canada.
E-mail address: andriy.kovalenko@nrc-cnrc.gc.ca (A. Kovalenko).

http://dx.doi.org/10.1016/j.molliq.2014.11.017
0167-7322/

theory of solvation properly accounts for chemical functionalities by representing in a single formalism both electrostatic and non-polar features of solvation, such as hydrogen bonding, structural solvent molecules, salt bridges, solvophobicity, and other electrochemical, associative, and steric effects [37–54]. With less computational effort, the 3D-RISM-KH theory analytically accounts for the solvation effects from the first principles of statistical mechanics in full molecular details, with accuracy comparable to that of explicit solvation molecular dynamics approaches [42,43]. Further, it yields the solvation thermodynamics, including the energy–entropy decomposition of the solvation free energy and volumetrics [40,41].

The robustness of the 3D-RISM approach was demonstrated recently by comparing the solvation structure of DNA calculated by using 3D-RISM-KH with that from MD simulations [44]. The salt-induced conformation change of DNA from its canonical B-form to the left handed Z-form was also studied using the 3D-RISM-KH theory [45]. In addition, other studies used 3D-RISM-KH to understand ions and solvent distribution around DNA [46,47], as well as solvent and salt effects on structural stability of DNA [48].

In this work, we calculated the binding free energy of 9- to 20-mer DNA oligonucleotides in a double stranded complex by post-processing the MD trajectories with the 3D-RISM-KH molecular theory of solvation. The solvation free energy contribution to the binding free energy calculated by post-processing using 3D-RISM-KH is compared with those using the Poisson–Boltzmann and Generalized Born solvent accessible area (PBSA and GBSA) continuum solvation models. The PB(GB)SA methods empirically describe the nonpolar terms in the solvation free energy, whereas the 3D-RISM-KH molecular theory of solvation performs statistical mechanical evaluation of the solvation free energy, including its nonpolar part.

The binding free energy calculated using these theoretical methods is compared to the experimental results. To the best of our knowledge, this is the first application of the 3D-RISM-KH theory to study the solvation thermodynamics of DNA assembly.

2. Material and methods

2.1. Binding energy calculations

Hybridization of a single strand oligonucleotide (SS1) and its complementary strand (SS2) to form a double stranded (DS) oligonucleotide can be represented by the following equation:

$$SS1 + SS2 \leftrightarrow DS. \tag{1}$$

Accordingly, the free energy difference between bound and unbound states of two complementary strands of ss-DNA molecules in solution can be calculated as

$$\Delta G_{bind} = \langle G_{ds}\rangle - \langle G_{ss1}\rangle - \langle G_{ss2}\rangle, \tag{2}$$

where the angular brackets denote average over MD trajectories sampled in the canonical distribution. The binding free energy ΔG_{bind} is also decomposed into the gas-phase contribution (including both the energetic and conformational entropic parts) and the solvation free energy:

$$\Delta G_{bind} = \langle \Delta E_{MM}\rangle + \langle \Delta G_{sol}\rangle - T\langle \Delta S_{MM}\rangle. \tag{3}$$

In the current study, the total gas phase energy E_{MM}, which is composed of the internal energy E_{int} due to the bonded interactions, the electrostatic energy E_{ele}, and the van der Waals energy E_{vdW} of the DNA strands, is obtained by using AMBER 10 molecular modeling package [55]. The solvation free energy, $\langle \Delta G_{sol}\rangle$, includes both the energetic and entropic parts, while the gas phase contribution to the binding free energy is explicitly split into the energetic contribution $\langle \Delta E_{MM}\rangle$ and the entropic one $-T\langle \Delta S_{MM}\rangle$.

The free energy of solvation consists of polar (ΔG_{pol}) and nonpolar (ΔG_{np}) contributions:

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{np}. \tag{4}$$

The solvation free energy ΔG_{sol} is calculated by using the 3D-RISM-KH theory as well as with the PBSA and GBSA continuum solvation methods. The latter provide the electrostatic potentials, the solvation free energies and forces by modeling solvent as structureless dielectric continuum, and mobile ions as continuous distribution of charge [30]. The nonpolar part G_{np} of the solvation free energy in the PBSA and GBSA methods is computed from the solvent-accessible surface area (SASA) term. The SASA method cannot accurately describe nonpolar solvation forces at the atomic length scale and may not be accurate or transferable for high resolution modeling studies of biomolecule folding and binding [31].

The conformational entropy part of the binding free energy, $-T\Delta S_{MM}$, comprises the translational, vibrational, and rotational contributions. In this study, it is obtained from the normal mode analysis based on the *nmode* module of the AMBER 10 package [56].

We calculated all the terms in Eq. (2) from three separate simulations for ds-DNA and two ss-DNA, as well as from simulating only the complex. In the latter case, the conformations of the strands were obtained by splitting the conformations of the complex along the trajectory into those of two individual strands. Such an approach has proven to provide a reasonable estimate for the binding free energy despite the fact the conformational changes upon binding are neglected [49]. We will refer these two ways of calculating the free energy as "3-trajectory" and "1-trajectory" calculations, respectively. The results obtained with these two approaches are compared to test accuracy of the assumption that the conformation of DNA strands does not change significantly upon binding. As compared to the "3-trajectory" approach, the "1-trajectory" approach is computationally less demanding, since it requires only a single MD simulation of the complex molecule instead of three separate MD simulations for the complex and the individual strand to generate conformational ensemble of the molecules for post-processing with the 3D-RISM-KH molecular solvation theory or with the PB(GB)SA continuum solvation models.

2.2. 3D-RISM-KH molecular theory of solvation

The 3D-RISM-KH theory yields the solvation structure of a solute macromolecule or a supramolecular assembly of arbitrary shape in terms of the 3D maps of solvent site density distributions around the solute, and then readily the solvation thermodynamics of the system, including the entropy–enthalpy decomposition and the solute partial molar volume [37,40,41].

The 3D-RISM integral equation [34–41] can be derived from the six-dimensional molecular OZ equation for the solute–solvent correlations [32] by orientational averaging of solvent molecular orientations thus reduced to representation with 3D solvent site correlations [37,40], and has the form

$$h_\gamma(\mathbf{r}) = \sum_\alpha \int d\mathbf{r}'\, c_\alpha(\mathbf{r}-\mathbf{r}')\, \chi_{\alpha\gamma}(r') \ , \tag{4}$$

where $h_\gamma(\mathbf{r})$ is the total correlation function of interaction site of solvent molecules at 3D space position $\mathbf{r}$, which is related to the 3D site distribution function of solvent around the solute, $g_\gamma(\mathbf{r}) = h_\gamma(\mathbf{r}) + 1$. The 3D site direct correlation function $c_\gamma(\mathbf{r})$ has the asymptotics of the 3D interaction potential of solvent site γ around the solute molecule. The site–site susceptibility $\chi_{\alpha\gamma}(r)$ of pure solvent splits up into the intra- and inter-molecular terms as

$$\chi_{\alpha\gamma}(r) = \omega_{\alpha\gamma}(r) + \rho_\alpha h_{\alpha\gamma}(r), \tag{5}$$