COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 111 (2013) 755–762

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Quantitative thermodynamic predication of interactions between nucleic acid and non-nucleic acid species using Microsoft excel

*Jiaqi Zou, Na Li**

Department of Mechanical and Aerospace Engineering, University of Miami, Coral Gables, FL, USA

ARTICLE INFO

Article history:
Received 15 March 2012
Received in revised form 5 June 2013
Accepted 14 June 2013

Keywords:
Thermodynamics
Equilibrium analysis
Excel
Sequence design
Nucleic acid probes

ABSTRACT

Proper design of nucleic acid sequences is crucial for many applications. We have previously established a thermodynamics-based quantitative model to help design aptamer-based nucleic acid probes by predicting equilibrium concentrations of all interacting species. To facilitate customization of this thermodynamic model for different applications, here we present a generic and easy-to-use platform to implement the algorithm of the model with Microsoft® Excel formulas and VBA (Visual Basic for Applications) macros. Two Excel spreadsheets have been developed: one for the applications involving only nucleic acid species, the other for the applications involving both nucleic acid and non-nucleic acid species. The spreadsheets take the nucleic acid sequences and the initial concentrations of all species as input, guide the user to retrieve the necessary thermodynamic constants, and finally calculate equilibrium concentrations for all species in various bound and unbound conformations. The validity of both spreadsheets has been verified by comparing the modeling results with the experimental results on nucleic acid sequences reported in the literature. This Excel-based platform described here will allow biomedical researchers to rationalize the sequence design of nucleic acid probes using the thermodynamics-based modeling even without relevant theoretical and computational skills.

1. Introduction

Nucleic acid probes are widely used in many PCR [1–3] and non-PCR based applications. For example, linear [4–7] and stem-loop (i.e., molecular beacons [8,9]) nucleic acid probes are generally used to detect nucleic acid targets; functional nucleic acids [10,11] (e.g., aptamers [12–14] and DNA/RNAzymes [15–17]) have been used to detect non-nucleic acid targets. Proper sequence design is critical for the performance of these nucleic acid probes.

Sequences of nucleic acid probes are often designed empirically and then optimized with a trial-and-error experimental approach. Because pure experimental approaches tend to be cost- and time-consuming, thermodynamics-based computational models have been utilized to help rationalize this process. Many of these models are based on free energy ranking to provide qualitative comparison between different sequence candidates [18–21]. Recently, several quantitative computational models have also been reported [22–27]. The quantitative models calculate equilibrium concentrations of all interacting species from initial concentrations and

* *Corresponding author at*: Department of Mechanical and Aerospace Engineering, University of Miami, 1251 Memorial Drive, McArthur Building, Room 222, Coral Gables, FL 33124-0624, USA. Tel.: +1 305 284 3316; fax: +1 305 284 2580.
E-mail address: n.li1@miami.edu (N. Li).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.06.008

thermodynamic parameters, thus providing a quantitative estimate on the extent of various interactions between the interacting species. Particularly, we have previously established a model that considers various interactions between nucleic acid and non-nucleic acid species, as well as formation of various nucleic acid self-structures [28]. The validity of this model as a quantitative guide in optimizing aptamer-based competitive nucleic acid probes has been experimentally verified. This model could be easily adapted to other applications that involve interactions between multiple nucleic acid and non-nucleic acid species.

However, implementation of the above-described thermodynamic model could be difficult for some biomedical researchers who are not familiar with relevant theoretical and computational skills. Herein we present a generic and user-friendly platform that allows users to apply the algorithm of the model for their own applications without the need to know the theoretical and computational details. Based on Microsoft® Excel formulas and VBA (Visual Basic for Applications) macros [21,29,30], two Excel spreadsheets have been developed: one for the applications involving only nucleic acid species, the other for the applications involving both nucleic acid and non-nucleic acid species.

2. Computational methods

For any nucleic acid species i, five different types of conformations are considered here (Fig. 1): random coil (D_i), secondary structure ($D_{i,S}$), self-dimer (D_iD_i), duplex formed with another nucleic acid species j ($D_iD_{j(j\neq i)}$), and complex formed with a non-nucleic acid species k ($D_i{}^*P_k$). Formation of $D_{i,S}$, D_iD_i, and D_iD_j is considered by Eqs. (1)–(3), respectively:

$$D_i \leftrightarrow D_{i,S}: \quad \frac{[D_i]}{[D_{i,S}]} = K_{D_{i,S}} \tag{1}$$

$$D_i + D_i \leftrightarrow D_iD_i: \quad \frac{[D_i]\cdot[D_i]}{[D_iD_i]} = K_{D_iD_i} \tag{2}$$

$$D_i + D_j \leftrightarrow D_iD_j: \quad \frac{[D_i]\cdot[D_j]}{[D_iD_j]} = K_{D_iD_j} \tag{3}$$

where [] represents equilibrium concentration, and K is dissociation constant which can be determined by:

$$K = e^{\Delta G/RT} \tag{4}$$

where R is the universal gas constant, T is the absolute temperature, and ΔG is the change in Gibbs free energy. ΔG for interactions between nucleic acid species can be accurately estimated based on nearest neighbor model [31] with web servers (e.g., mfold [32–34]) or standalone software packages (e.g., RNAstructure [35]).

For any non-nucleic acid species k, only two types of conformations are considered here: unbound (P_k) and complex formed with a nucleic acid species i ($D_i \times P_k$). In contrast to the interactions between nucleic acid species that can be described with their intrinsic dissociation constants as in Eqs. (1)–(3), the interactions between nucleic acid and non-nucleic acid species usually need to be described using experimentally

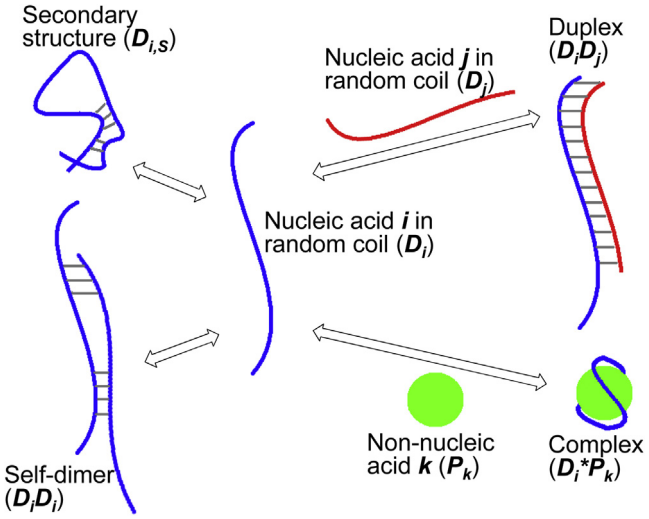

Fig. 1 – Illustration of different conformations considered for each nucleic acid species in the thermodynamic model: random coil D_i, secondary structure $D_{i,S}$, self-dimer D_iD_i, duplex formed with another nucleic acid species D_iD_j, and complex formed with a non nucleic acid species $D_i{}^*P_k$.

determined apparent dissociation constant $K^{App}_{D_i^*P_k}$ as follows:

$$D_i^* + P_k \leftrightarrow D_i^*P_k: \quad \frac{[D_i^*]\cdot[P_k]}{[D_i^*P_k]} = \frac{([D_i]+[D_{i,S}]+2[D_iD_i])\cdot[P_k]}{[D_i^*P_k]} = K^{App}_{D_i\times P_k} \tag{5}$$

where D_i^* is the nucleic acid species i in one of the unbound conformations, which could be either random coil D_i, secondary structure $D_{i,S}$, or self-dimer D_iD_i.

In a system that contains N number of nucleic acid species and M number of non-nucleic acid species, mass conservation for nucleic acid i and non-nucleic acid k results in Eqs. (6) and (7), respectively:

$$[D_i] + [D_iD_i] + [D_{i,S}] + \sum_{j=1}^{N}[D_iD_j] + \sum_{k=1}^{M}[D_i^*P_k] = [D_i^{initial}] \tag{6}$$

$$[P_k] + \sum_{i=1}^{N}[D_i^*P_k] = [P_k^{initial}] \tag{7}$$

where $[D_i^{initial}]$ and $[P_k^{initial}]$ are the initial (i.e., total) concentrations of nucleic acid species i and non-nucleic acid species k, respectively.

For such a system that contains N number of nucleic acid species and M number of non-nucleic acid species, there will be totally $(N+M)$ number of mass conservation equations, and $(3N/2+N^2/2+N\cdot M)$ number of equilibrium equations (N for formation of nucleic acid secondary structures, $N\cdot(N+1)/2$ for formation of nucleic acid duplexes, and $N\cdot M$ for formation of nucleic acid & non-nucleic acid complexes). This yields a closed set of equations about equilibrium concentrations of all nucleic acid and non-nucleic acid species in various bound