DESIGN OF AN APTAMER BY SHAPE COMPLEMENTARITY MAXIMIZATION FOR INSULIN BINDING

A PREPRINT

Artem E. Kniga
M.Sc. in Biology
Lomonosov Moscow State University
artem.kniga@chemistry.msu.ru

June 21, 2023

ABSTRACT

In this paper, we propose to use parametrized secondary structure as low-dimensional space during optimization of aptamer sequence. We are testing the efficiency of a simple approach that directly maximizes shape complementarity on the intermolecular interface. From a library of symmetric parametrically designed structures, we identified a ssDNA sequence binding human insulin. Using molecular dynamics, we got an estimation of its binding affinity to be $K_d = 6 \mu M$.

Keywords  SELEX · aptamers · in silico

1 Introduction

The invention of the SELEX process allowed artificial evolution of nucleic acids in the direction of a targeted molecular recognition [Turk and Gold, 1990, Ellington and Szostak, 1990]. The growing number of discovered aptamers is used in analytical applications as compact sensors [Wu et al., 2019]. The understanding of physical interactions that govern aptamer-target complex formation is important, that’s why it may be of interest to rationalize the affinity of a newly discovered sequence showing binding in an experiment [Khrenova et al., 2022].

Aptamers are routinely chemically synthesised that makes them economically advantageous comparing to antibodies in developing diagnostic applications. There is a large corpus of literature today that describes results of SELEX experiments, but mostly without detailed molecular characterisation of aptamers interaction with their targets.

From another point of view, molecular design is a tool that can deeper our understanding of the molecular recognition and finally help us to produce better chemicals. The theoretical foundation for computational molecular design was laid by extensive efforts that concentrated on proteins and used a fast to compute and to optimize pairwise-decomposable score function that use implicit solvation model for easy parallelization [Leaver-Fay et al., 2011]. Different approach to study the problem of molecular recognition was laid to concentrate on physical properties of the molecular surface, such as lipophilicity [Pykov et al., 2009], or by calculating the electrostatic potential on molecular surface [Mironov et al., 2022]. In proteins, a large variety of available side-chains radicals make it generally more accessible to generate a fine-resolution surface complementarity of geometry and charge, as it is, for example, in antibodies. We speculate that the easiest way to assess contribution of enthalpy to the free energy binding on the stage of initial complex formation may be a geometric surface matching that is actually used in many docking studies [Lawrence and Colman, 1993].

This metric, shape complementarity, can be computed using the following formula:

$$Shape\, Complementarity = \sum_{i=1}^{k} \sum_{j=1}^{l} \left| \frac{n_i \cdot n_j}{d_{ij}} \right|$$

where $n$ is a normal vector for a vertex in a molecular (Connolly’s) surface meshes (i and j for vertex’s index of the interactor’s surface meshes going until k and l, number of vertices in first and second mesh correspondingly), and $d_{ij}$ is a matrix of pairwise distance between all pairs of vertices on the interactor’s surface meshes.
2 Methods

2.1 Library generation

We generated a library of aptamer sequences as solutions for the reverse RNA folding for strings of secondary structure with the RNAreverse program from ViennaRNA software suite [Lorenz et al., 2011]. Secondary structure strings were generated in combinatorial fashion using number of stems, their length, and length of loops as parameters. Code for the script used for this task is available on GitHub.

2.2 DNA structure modeling

A library of 3D ssRNA structures was generated using the rna_denovo application [Watkins et al., 2020]. For each secondary structure sequence, we generated a plausible primary sequence and then obtained a full-atom minimized structure. Secondary structures were checked using RNAfold application [Lorenz et al., 2011]. Sequences were converted from RNA to ssDNA using addition of O5’ to the first nucleotide in sequence with Pymol script, deletion of O3’ and HO31 and HO32 atoms, and subsequent minimization of obtained structures using AMBER force field in Gromacs.

2.3 Molecular docking

For molecular docking, we used PatchDock [Schneidman-Duhovny et al., 2005] software that computes surface similarity. The file of parameters is made publicly available in the GitHub repository.

2.4 Molecular dynamics

All molecular dynamics computations was conducted using Gromacs version 2022.2. AMBER14SB OL15 force field [Zgarbová et al., 2015] was used for all computations. Protein-aptamer system was solvated in a dodecahedron box with 1.5 nm distance, using a spc/e model of water. Ions (Na+) were added until neutralization of the system charge. MM/PBSA energies were computed using gmx_MMPBSA package [Valdés-Tresanco et al., 2021]. Steered molecular dynamics and umbrella sampling were adopted from the seminal paper [Lemkul and Bevan, 2010].

3 Results

3.1 Library generation

We use a library of simple symmetric multi-loop secondary structures, parametrized using a variable number of stems, length of stems, loops. We are operating within an induced fit theoretical framework of receptor-ligand binding using surface similarity as optimization function for screening a library for the best geometrically fit aptamer-insulin initial complexes. We choose the multi-stem topology because it is common in known aptamer structures, and easier to find primary sequence for than other topology types as pseudo-knots. The library consisted of 54 sequences, each described by 4 parameters: number of stems, length of the basal stem, length of stems, length of multi-loop.

3.2 Molecular docking (in silico screening)

After completing docking of insulin onto a library, we sorted DNA sequences according to the largest score of their geometric transformation. The name, the sequence and its corresponding secondary structure with the largest score:

2.3.5.4
5’-gcauauucgccggcauggcgcggacuggucaaguauaagugc-3’
(((....(((......))))...(((......)))....)))

3.3 Molecular Dynamics

3.3.1 MM/PD(GB)/SA

For 5 first docking conformation, we conducted refinement that allowed us to calculate energetic characteristics of the obtained complexes and rank them to pick the most stable conformation for the umbrella sampling study. For the best conformation we computed 100 ns of molecular dynamics, that showed the very dynamic nature of the multi-loops.
Design of an aptamer by shape complementarity maximization for insulin binding

A P

PREPRINT

Figure 1: Library of secondary structures used for screening.

connecting aptamer stems. We calculated MM/PB(GB)SA energy along the trajectory, that showed continued decline of association energy that showed progressive dissociation of the insulin from the aptamer².

3.3.2 Advanced sampling

After selecting one conformation with the most negative energy difference of association⁴. In order to obtain approximate binding affinity of the leading aptamer sequence, we calculated the potential of mean force across center-of-mass distance between aptamer and protein that gave the final free energy difference of ca. 30 kJ/mol³.

3.4 Molecular docking (in silico screening)

4 Discussion

Molecular design is an optimization problem that seeks to find an optimum of some function (e.g. Gibbs free energy of binding) given a space of possible molecular structures (in the case, there is a surjective mapping between molecular structures and the set of aptamer ssDNA sequences, which may be non-injective). In this study, we explore an approach for designing an aptamer by searching for a maximum geometric complementarity of interacting molecular interfaces across a library of 3D molecular models for ssDNA sequences and a human insulin. The latter is readily available from pharmacies, as it’s an important human protein. ssDNA for aptamer could also be readily synthesized by numerous
Design of an aptamer by shape complementarity maximization for insulin binding

In general, a good aptamer that binds with a high affinity should satisfy these two criteria, that just are mere consequences of the fundamental equation of thermodynamics:

- It has a high degree of geometrical and electrostatic shape complementarity to its target, so it maximizes enthalpic contribution to the free energy.
- It binds in a well-defined free energy minimum conformation that is stable on a timescale that is achieved in typical molecular dynamics studies. So it minimizes the entropic penalty on binding.

Developing a highly stable aptameric tertiary structure that will show a great enthalpic effect on binding is not an easy task.
Navigating low-dimensional space of secondary structures may be easier than randomly generating strings of primary sequences. Actually, this approach has the additional advantage that it generates sequences of different lengths that may be used for generating starting libraries for SELEX.

5 Conclusion

We discovered a ssDNA sequence binding human insulin using a new computational design approach, that we make publicly available to the molecular modeling community.

All code for reproduction of the results is available from:

https://github.com/le1ivre/aptamer/tree/main

References


