Biosensor Design through Molecular Dynamics Simulation

Wenjun Zhang, Yunqing Du, Steven W. Cranford, Ming L. Wang

Abstract—The beginning of 21st century has witnessed new advancements in the design and use of new materials for biosensing applications, from nano to macro, protein to tissue. Traditional analytical methods lack a complete toolset to describe the complexities introduced by living systems, pathological relations, discrete hierarchical materials, cross-phase interactions, and structure-property dependencies. Materionics – via systematic molecular dynamics (MD) simulation – can provide structure-process-property relations by using a materials science approach linking mechanisms across scales and enables oriented biosensor design. With this approach, DNA biosensors can be utilized to detect disease biomarkers present in individuals’ breath such as acetone for diabetes. Our wireless sensor array based on single-stranded DNA (ssDNA)-decorated single-walled carbon nanotubes (SWNT) has successfully detectedtrace amount of various chemicals in vapor differentiated by pattern recognition. Here, we present how MD simulation can revolutionize the way of design and screening of DNA aptamers for targeting biomarkers related to oral diseases and oral health monitoring. It demonstrates great potential to be utilized to build a library of DNDA sequences for reliable detection of several biomarkers of one specific disease, and as well provides a new methodology of creating, designing, and applying of biosensors.

Keywords—Biosensor, design, DNA, molecular dynamics simulation.

I. INTRODUCTION

The field of biomaterials research has been very successful over the past fifty years with many innovations and developments affecting people’s lives. It has witnessed the introduction of a number of sensational devices that had relied on coordinated materials development, such as heart valves, hip implants and contact lenses. Another example can be the development of advanced biomaterials, such as heart valves, medical implants development, disease diagnostics, research and lots of biomedical applications such as drug design, medical implants development, disease diagnostics, and health surveillance [22]-[25].

Recent developments in biosensing and device-level integration with nanomaterials have taken such an approach, exploiting DNA, RNA, and protein-based materials. Indeed, electronic biosensing and detection represents the most developed area of bio-nanoelectronics. One reason is due to the potential to control the molecular sequence of the structure (e.g., via peptide or nucleobase). DNA, for example, has a limited set of four building blocks (four nucleobases). Yet, even with such a simple base set, sequencing of DNA leads to the genetic diversity we see across Nature. DNA consists two long polynucleotide chains which run in the opposite directions and are twisted around each other right-handedly [26]. Each strand of the double helix is a linear chain with a backbone made of sugars and phosphate groups joined by ester bonds. Attached to each sugar is one of the four types of bases, including the purines: adenine (A) and guanine (G), and the pyrimidines: cytosine (C) and thymine (T). DNA is well suited for bio-sensing applications because of their specific and robust base-pairing interactions between complementary sequences [27], [28]. DNA sensors have been used to detect DNA [29], [30], proteins [31], [32] and even small molecules/ions [33], [34] in the form of optical [29], [35], [36], electrochemical [37]-[39], or mass-sensitive [40], [41] for a variety of biological applications. Most recently, DNA microarray technology has emerged, offering remarkable high-throughput screening properties and reliable biomedical
diagnostics applications [42]-[44]. It provides a discovery platform of functional genomics [45], [46] and a revolutionizing way of drug design and disease diagnostics [47], [48]. DNA sensors are envisioned to be valuable, easy, inexpensive, fast, and specific techniques in many applications such as medical diagnostics, genetic screening, drug design, food and agricultural analysis, environmental monitoring and health surveillance. In addition, DNA has been shown to be compatible with other emerging nanomaterials such as carbon nanotubes [49]-[51], enabling the potential exploitation of the benefits of both materials on simple devices.

For sensing, rather than screen thousands of potential material candidates, it would be pragmatic to optimize variations of well-known materials. DNAs (through nucleobases) have their own “programming language” to explore, with a finite number of sequence combinations. Exploration of synthetic DNA sequences has enable the development of DNA-origami as well as bioimaging and functional scaffolds [52], [53]. The flexibility of DNA sequencing allows for the incorporation of multiple ligands, labels for bioimaging, antibodies, hormones and so forth that might be used for efficient and site-specific drug delivery and release [52], [54]. Here, we discuss the variation of sequence with oral biomarker interaction. Ultimately, a library of DNA sequences can be developed, to encompass all of the nucleobase sequence combinations within the “sight” of the biomarker (a function of molecular size and atomistic interaction distance).

While great, the number of sequences is finite in size and thus computational tractable. The key challenge will be the development of a robust computational protocol to assess the biomarker interaction of a suite of target molecule and DNA chemistries. Minimal physical effort is required to successfully model the materials, enabling the potential to automate the approach. Atomistic interactions can then be used to interpret and guide experimental efforts, which reciprocally feed computational models. This feedback loop provides unprecedented insight into the behavior of complex material systems [55], [56].

II. METHOD AND RESULTS

The interactions between DNA and small molecules have been largely applied to build biochemical sensors for disease diagnosis [57]-[61] and detection of explosives [49], [62]. The first step is to rank DNA sequences with specific biomarkers to demonstrate potential molecular specificity. According to our knowledge, there is no known protocol to design DNA sequences to achieve the best detection results for particular molecules. It has opened up a tremendous possibility to map an array of DNA sequences for reliable detection of several particular biomarkers of one specific disease, and provides a new paradigm of design, development, and application of advanced engineering material systems, combining computational approaches, optimization methods, and DNA informatics.

Clearly, there is a multitude of diseases and biomarkers to potentially explore. Focus here is given to a single affliction: diabetes. Diabetes has known biomarkers throughout the body. The most common of which is glucose levels in the blood (which defines the disease), requiring blood screening as definitive diagnosis. However, blood sampling, by definition, requires breaking the epidermis – i.e., a needle prick – that is adversarial to many patients, especially to children or older people. Thus, our hope is to develop sensing and screening technologies which can noninvasively detect diabetic oral biomarkers from breath or saliva [14], [15], among which we will focus on the detection of diabetic ‘signature’ VOC (Volatile Organic Compounds) components from breath using DNA sensors.

A recent preliminary investigation explored the variation of DNA sequence with diabetes biomarker interaction to demonstrate proof-of-concept screening approaches (Fig. 1). Short sequences of both single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) were modeled interacting with a particular diabetes biomarker molecule using full atomistic MD simulation. The breath biomarker of diabetes selected for the study was acetone [63]-[65]. Acetone is reported to be less than a few hundred ppb (by volume) in the breath of healthy individuals [66] while for diabetic patients, acetone concentration can reach 560 ppm or even > 1000 ppm [67]. The interaction between four single DNA nucleotides (A, G, C, and T) on both ssDNA and dsDNA with acetone was demonstrated via Steered Molecular Dynamics (SMD) [56] (Fig. 1 (b)) applying the well-utilized CHARMM and CVFF potentials. Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS), an open-source molecular dynamics software package is used to perform all MD simulations [68], [69].

SMD is a novel approach to study the dynamics of binding or unbinding events in biomolecular systems [70], revealing the details of molecular interactions in the coarse of unbinding [71], [72] and providing important insights of the mechanisms underlying these processes. The primary advantage of nonequilibrium SMD over conventional equilibrium MD methods is the possibility of inducing relatively large conformational changes in molecules within the nanoscale time scales accessible to simulation. Computationally, the SMD method applies a moving spring force so that the molecule can behave in a manner not obtained by either force or displacement loading alone, allowing induced conformational changes in a system along a prescribed reaction vector. SMD can be thought of the simulation counterpart to popular AFM experimental assays.

For each simulation, the DNA molecule, either single-stranded or double-stranded, was set at one end of a solvation box (explicit water), and the SMD force was applied at the geographical center atom of the biomarker. The small molecules were pulled towards the middle of one particular nucleotide. Total force and the potential mean force (PMF) during the SMD simulations was then plotted against the distance between the biomarker and DNA, enabling analysis and comparison of the interaction pathways.
Despite plentiful modeling methods for the interactions between a biomarker and DNA [73], little is known a priori about processes of binding and unbinding, limiting any predictive (or design) power. An SMD simulation is a non-equilibrium process, which accepts irreversibility, ceding for the present time accurate evaluation of binding affinities and PMFs, but gaining access to biologically relevant information related to non-covalent bonding. PMF can be equated to the free energy profile along the reaction path. To properly capture the free energy describing the conformational space of the binding event, the SMD simulations typically include a very large statistical sample of multiple initial conditions and multiple directions of the binding vector. For ranking purposes, however, such a degree of accuracy is unnecessary. Only a few (one to three) approaches/trajectories per biomarker/nucleobase pairing is sufficient for a preliminary assessment.

The interaction between four single DNA nucleotides (A, G, C, and T) on both ssDNA and dsDNA with acetone was studied. The mechanical work of pulling it forwards (forward pulling path) and backwards (reverse pulling path) at a number of points was measured during this process. By sampling these forward and reverse paths, the free-energy profiles of the eight aforesaid systems for acetone could be assessed (Fig. 2 (a)). Four DNA nucleotides on dsDNA were found to react differently to the targeted molecules than on ssDNA, requiring significant higher energy to move the molecule close to DNA than the later. Comparing the PMF values of the different systems, we obtained the optimal DNA nucleotide for the detection of acetone is Adenine for acetone (Fig. 2 (b)), which is in good agreement with experimental sensing results using DNA sensors.

Beyond single nucleotide, the simulation process can be automated to assess a whole library of possible DNA sequences to select the optimal arrangement in order to optimize the entire DNA sequence for one specific biomarker interaction. This necessitates modifying the MD code to swap sequences on-the-fly based on the determined critical metrics.

![Fig. 1](image1.png)

**Fig. 1** (a) Full atomistic model of 24 nucleotide-DNA strand-acetone; (b) Schematic of SMD simulation. General constant velocity SMD approach where macromolecule is connected with harmonic spring with defined stiffness, $k_{spring}$, and a fixed velocity, $v_{constant}$, towards a target coordinate (x, y, z); in this case, the target is a single nucleobase (A, C, G, T) of a 24-based ssDNA/dsDNA

![Fig. 2](image2.png)

**Fig. 2** SMD simulation results of A, T, C, and G nucleotides in ssDNA-acetone system at $k_{spring} = 6.95$ N/m with pulling speed at 10 m/s: (a) energy profiles indicated by accumulated PMF; and (b) ranking of acetone interaction with A, T, C, G nucleotides on ssDNA
ACKNOWLEDGMENT

We acknowledge fruitful discussions with E. Goluch, V. Godoy-Carter, and R. Birken regarding this project. Portions of our research work were conducted using the Discovery Cluster (discovery.neu.edu) at Northeastern University.

REFERENCES

Digital Open Science Index, Biotechnology and Bioengineering Vol:10, No:1, 2016


