Journal of Molecular Graphics and Modelling 84 (2018) 43–53

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Molecular dynamics simulation for rational protein engineering: Present and future prospectus

Maryam Rouhani [a], Farnaz Khodabakhsh [b], Dariush Norouzian [a], Reza Ahangari Cohan [a, **], Vahideh Valizadeh [a, *]

[a] Nano-Biotechnology Department, New Technologies Research Group, Pasteur Institute of Iran, Tehran, Iran
[b] Department of Genetics and Advanced Medical Technology, Medical Biotechnology Research Center, Faculty of Medicine, AJA University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Article history:
Received 17 April 2018
Received in revised form
5 June 2018
Accepted 8 June 2018
Available online 9 June 2018

Keywords:
Protein engineering
Computational modeling
Glycoengineering
PASylation
PEGylation
Protein stability
Contents

ABSTRACT

Recently protein engineering has been used as a pivotal tool for designing proteins with improved characteristics. While the experimental methods might be laborious and time-consuming, *in silico* protein design is a time and cost-effective approach. Moreover, in some cases, protein modeling might be the only way to obtain structural information where the experimental techniques are inapplicable. Molecular dynamics (MD) simulation is a method that allows the motion of protein to be simulated in defined conditions on the basis of classical molecular dynamics. MD simulation could widely be used when protein design needs accurate modeling of the target protein dynamics and also descriptions of the relation between conformational changes and function of protein at the atomic level. In this review, the effectiveness and the power of MD simulation in designing proteins with improved characteristics will be discussed.

1. Introduction

Proteins are one of the fundamental building blocks of living organisms which are able to form a distinct structure through spatial arrangement. The molecular evolution has revealed that variations occurred in protein sequences, by mutagenesis or recombination, can alter proteins characteristics which is due to generation of new structures [1]. An impressive challenge in structural biology is to design and engineer proteins in order to exhibit new or desired functionalities. Protein engineering is a technology through which novel proteins with preferred or improved properties can be developed. It is one of the dynamically developing disciplines which can be used in bio-industries. In last three decades, the protein engineers have successfully tailored wide ranges of proteins specified to use in industry and medicine [2]. This could be achieved by developing novel experimental technologies such as recombinant DNA, high-throughput screening, deep sequencing, directed evolution, fluorescence-based screening, and gene synthesis [3]. In addition to experimental methodology and rational design strategies, computational methods have been successfully employed for more facile designing and engineering the proteins [4].

2. Computation in molecular biology

Nowadays, computer is a crucial device in most studies, particularly in molecular biology. Computational modeling principally is performed by two different methods; (i) a subjective computer graphic and (ii) an objective computational analysis on the basis of mathematical equations and biophysical properties of structural energies. Translocation, rotation, zooming in real space, creating complicated molecular models, simulation of molecules at atomic levels, analysis of molecular surface, least-squares superposition of molecules, and analysis of large datasets can be accomplished only by computers. Computational methods are mainly grouped into three categories: (i) bioinformatics analysis of primary sequences, (ii) computer modeling of tertiary structures,

* Corresponding author.
** Corresponding author.
E-mail addresses: cohan_r@pasteur.ac.ir (R.A. Cohan), v_valizadeh@pasteur.ac.ir (V. Valizadeh).

https://doi.org/10.1016/j.jmgm.2018.06.009
1093-3263/

known as molecular modeling, and (iii) prediction of new structures by *de novo* design [5].

Molecular modeling studies are often combined with several *in silico* methods such as bioinformatics analysis and quantitative structure/activity relationship (QSAR) to predict effectiveness of every change(s) made in the protein. However, prediction of the conformational behavior of amino acid residues requires accurate estimation of binding energies and assessing the reaction activation barrier changes. In spite of significant progress in quantum mechanics, in the long-term simulation of complex molecules, MD simulation is still preferred due to the greatly reduced computing process [3,6]. MD simulation is a powerful method to study the dynamic feature of a protein at atomic levels [7]. Additionally, it provides correlation between structure and dynamics considering conformational energy landscape accessible to protein molecules [8–10]. Today, modern computers allow molecular simulations ranging from nanoseconds to microseconds which are enough time to determine conformational changes at atomic level. Therefore, MD simulation is an attractive method which identifies flexible regions in the protein serving as proper targets for stability increment or achievements in other protein engineering objectives [7].

3. Molecular dynamics simulation for protein engineering

Earlier, protein modeling software tasks dealt with chain closure, constructing molecules from building elements, and examining the conformational space by manual changes of torsional angles [5]. In 1972, Katz and Levinthal studied the hardware and software aspects of molecular structure presentation, manipulation, and structure fitting into electron density contours. Beside many aspects of protein structural science, computer hardware progress and evolution of methods revealed direct impact in this field. For instance, development of computer graphic was often driven by technical requirements of protein crystallography techniques [11].

One of the early papers published on computer modeling of protein was the study of Levitt and Marshal (1975). They described the computer modeling of a protein folding on the basis of a new and simple presentation of a protein conformation plus energy minimization and thermalization. The method successfully described the renaturing of bovine pancreatic trypsin inhibitor from open-chain conformation to the folded state which is similar to the native protein under defined conditions [12]. Later, Sherga and Kuntz published their studies with similar themes [13–15]. However, in a primitive modeling system, despite the significant simplification of the protein molecule, the molecular mechanic force fields were not selected properly. Because the potential energy of protein in a vacuum is not a good approximation for the free energy of a biological system [5]. Computer modeling should finally face the biophysical and thermodynamic characteristics of a protein in an aqueous solution. A decade later, a 210-ps simulation of understudied protein in water was reported [16].

Later, a significant increase in computing power brought about routinely simulation of larger proteins which are 1000–10000 times longer than the original primitive simulation in an aqueous environment containing ions. Additionally, significant improvements in the potential function of a protein with respect to enhancing the stability have been achieved [17]. This potential was attained by using more defined/accurate force fields along with coordinates [18]. In theory, the force field is employed for parameterizing protein energy. However, considering the structural complexity of protein, protein force fields are divided into different terms. Regarding variety of methods existing to develop a model system and parameterize surface energy of a protein, different force fields are available [19]. Currently, the most commonly used force fields are chemistry at HARvard molecular mechanics (CHARMM) [20], assisted model building and energy refinement (AMBER) [21], Optimized potential for liquid simulations (OPLS) [22] and groningen molecular simulation (GROMOS) [23]. These force fields are generally available in particular modeling packages which frequently can be used to simulate the macromolecules.

The ultimate goal of protein modeling is the accurate prediction of a protein structure from its primary sequence which is comparable to that of experimentally obtained results [24]. This will allow the investigators to safely use easy generated *in silico* protein models which can be used in all contexts, instead of experimental examination. Such approach could be very supportive in structure-based drug design, analysis of protein function, rational design of proteins with enhanced stability or increased *in vivo* half-life, decrement in immunogenicity, and in some cases, achieving novel proteins with new functionalities [25–27]. *In silico* design is an alternative time saving and cost effective method which sometimes is the only approach to obtain the structural information of the protein when the experimental procedures are failed [24]. Regarding this viewpoint, in this manuscript, the attempts will be made to review the application of MD simulation in protein engineering. In the following, a number of protein modification approaches will be discussed in which molecular dynamics was used for simulating engineered proteins that led to a general understanding of mechanism or developing a molecule with enhanced or novel properties.

3.1. Molecular dynamics simulation for protein glycoengineering

A large number of therapeutic proteins have been developed for the treatment of different diseases, but some drawbacks, like loss of activity or rapid clearance from the circulation, limit their clinical applications [28]. Novel strategies are in use to design new drugs with higher activity and longer *in vivo* half-life. Glycoengineering which means a change in carbohydrate moiety of a protein, causes alterations in pharmacokinetics characteristics of the target protein [29,30]. Carbohydrate chain addition to the protein can lead to a significant reduction in protein aggregation by increasing solubility through masking the hydrophobic patches on the protein surface [31].

Naturally, protein glycosylation involves covalent binding of glycan to proteins through amino acid side chains of asparagine (N-linked), and serine/threonine (O-linked) [32]. The N-glycosidic linkages occur between the carbohydrate moiety of beta-N-acetyl glucosamine and the side chain of asparagine residue wherein the amino acid is embedded in tripeptide sequence N-X-S/T [33,34]. On the contrary, the O-glycosidic linkage occurs between different glycan moieties and the residues of serine/threonine. Unlike the particular position of an asparagine residue in tripeptide sequence, serine/threonine involved in O-linked glycosylation does not show any specific amino acid sequence preferences.

Introduction of new glycosylation sites in the protein structure leads to formation of proteins with high carbohydrate content. Nonetheless, introducing new glycosylation sites affects the folding, three-dimensional (3D) structure, and activity of target protein [29]. Moreover, the surface accessibility of asparagine residue and the possibility of enzymatic glycosylation should be considered. Therefore, rational selection of proper positions for the introduction of new glycosylation site before the experimental approach is important due to being cost-effective and time-saving.

Samoudi et al. (2015) used recombinant human β interferon (rhuIFN-β) as a model protein to identify the suitable positions for introducing new N-glycosylation sites [35]. They employed a computational strategy to predict the structural distortion and function of the target protein which might be caused by the