ELSEVIER

Biophysical Chemistry 119 (2006) 146 – 157

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Protein flexibility: Multiple molecular dynamics simulations of insulin chain B

F.S. Legge [a], A. Budi [a], H. Treutlein [b], I. Yarovsky [a,*]

[a] *Applied Physics, School of Applied Sciences, RMIT University, GPO Box 2476V, Melbourne, Victoria 3001, Australia*
[b] *Cytopia Research Pty. Ltd., PO Box 6492, St. Kilda Road Central, Melbourne, Victoria 8008, Australia*

Received 12 July 2005; received in revised form 3 August 2005; accepted 3 August 2005
Available online 29 August 2005

Abstract

Multiple molecular dynamics simulations totaling more than 100 ns were performed on chain B of insulin in explicit solvent at 300 K and 400 K. Despite some individual variations, a comparison of the protein dynamics of each simulation showed similar trends and most structures were consistent with NMR experimental values, even at the elevated temperature. The importance of packing interactions in determining the conformational transitions of the protein was observed, sometimes resulting in conformations induced by localized hydrophobic interactions. The high temperature simulation generated a more diverse range of structures with similar elements of secondary structure and populated conformations to the simulations at room temperature. A broad sampling of the conformational space of insulin chain B illustrated a wide range of conformational states with many transitions at room temperature in addition to the conformational states observed experimentally. The T-state conformation associated with insulin activity was consistently present and a possible mechanism of behavior was suggested.

Keywords: Molecular dynamics; Multiple simulations; Conformational flexibility; Insulin; Chain B; T-state; R^f-state

1. Introduction

Theoretical modelling is a useful complement to experimental studies in providing insights into protein folding [1]. The advent of faster computers allows an opportunity to broaden the scope of molecular dynamics (MD) simulations and apply more sophisticated analysis of the ensemble of conformations sampled. However, it is still unclear how to apply the increase in computing power most effectively. Even with improved resources, it is still only possible to simulate on a relatively short biological timescale, often falling short of timescales of relevant occurrences, thus limiting the effectiveness of simulations. MD simulations commonly examine the conformational changes of one molecule over time, whereas, in real life many molecules are undergoing conformational changes cooperatively. This is reflected in studies that have shown the existence of different pathways with similar free-energy barriers in the unfolding of a protein [2,3]. For these reasons, the use of multiple simulations with different initial conditions may provide more reliable and statistically relevant information about protein dynamics. Multiple MD simulations have been used previously in a number of studies to investigate protein dynamics [4–8]. These studies have produced useful information relating to the diversity of pathways a protein undergoes in the process of denaturation and to the principles of protein folding. Depending on the number of simulations and their time length, the total simulation time can vary in length from nanoseconds to several microseconds. Probably the largest of these studies are those carried out using the global distributed computing network (Stanford Folding@Home group) [8]. In the present paper we use the results of four MD simulations to study the conformational behaviour of the prototype protein insulin which has been intensively studied previously, both

* Corresponding author. Tel.: +61 3 9925 2571; fax: +61 3 9925 5290.
E-mail address: Irene.yarovsky@rmit.edu.au (I. Yarovsky).

0301-4622/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2005.08.002

computationally and experimentally [9–12], thus sufficient data are available for comparison. The use of multiple long-term simulations can help sample conformational phase space more effectively and efficiently by minimizing the impact of entrapment in local minima. Each of the simulations is of 20 ns duration, which has been shown to be of sufficient time-scale to sample biologically relevant conformational changes in insulin [13].

Insulin is of high physiological importance. A diverse range of conformations has been observed in dimeric and hexameric crystal forms [14–18]. In solution insulin exists as a monomer [19,20], with hexameric insulin stored in the pancreas for future use. The active form of insulin is the monomeric state [21], however, because of a tendency for self-association, a crystal structure of the monomer has yet to be determined. Strategies employed to avoid dimerisation, such as, low pH conditions and chemical modifications, suggest a significant degree of molecular flexibility [22–25].

Insulin is a small globular protein composed of two chains, A and B, containing 21 and 30 amino acids, respectively, and linked together by disulfide bonds. Chain B of insulin is believed to retain much of its structure independently of chain A [10,26,27]. The significance of observed flexibility of chain B in terms of insulin's activity is not well understood. There are many structures currently available of insulin, although only the general binding mode of the insulin-receptor complex is known [28]. Structure–activity studies of insulin indicate the C-terminus of chain B as integral to receptor interaction [29–32], and also suggest that the conformation of the C-terminus is influenced by the structure of the N-terminus [33–35]. Various studies of insulin [9,36], including a preliminary crystallographic structure of the native insulin monomer at a low pH [25] and also solution structures of isolated chain B determined by NMR spectroscopy [10,27], confirm the termini region's mobility. The secondary structure characteristics of chain B of insulin are generally defined as the N-terminal residues (1 to 8), a central α-helix (9 to 19), a characteristic chain B fold (β-turn from 20 to 23) and the extended C-terminus (24 to 30). There are several known conformational states for the N-terminus of chain B [17,33,34,37,38]. The principal conformations have been designated as the R-state and T-state [39], although additional conformations have also been identified [18]. The T-state is associated with activity as it is believed to be the monomeric solution state [19]. In the T-state, chain B consists of the α-helix (9 to 19) and two extended N- and C-termini chain regions [40,41]. In the R-state, the N-terminal residues (1 to 8) form a helix, joining with the central helix [15,42]. Another variation is the "frayed" or R^f-state, where the α-helix is only present for part of the N-terminal residues (4 to 8) in addition to the central helix.

While both X-ray and NMR techniques have produced useful information on the protein's structure, each method has its limitations. For example, NMR spectroscopy is useful because the structures are obtained in solution without crystal packing constraints, but because insulin has a tendency towards aggregation, either low concentrations or specific pH conditions must be used. This is of some concern, as changes in pH can also affect the structure. Alternatively, X-ray structures are static glimpses of structures often restrained by crystallographic environment. Therefore, it is not surprising that the structures of insulin produced by these two methods show notable variation, particularly in the flexible N-terminus of chain B. Molecular dynamics is therefore a useful complementary technique to experimental methods, in that it enables a broad sampling of the conformational states of a protein under controlled conditions and therefore delivers additional understanding of the protein's dynamics [43]. Previous molecular dynamics studies of insulin show the conformational flexibility of insulin chain B, reporting a high degree of movement in aqueous solution of both the monomer and dimer [44–46]. The most recent of these studies, simulations of 5–10 ns length of monomeric insulin in the T-state conformation, describes the flexibility of the N- and C-termini regions of the chain B [12]. This inherent flexibility was also observed in a simulation of insulin chain B in solution in our study of possible thermal and chemical effects on protein dynamics [11].

It is always difficult to assess the statistical weight that should be applied to MD results. With only a small number of trajectories, which dynamic features are relevant to explain the macroscopic behaviour? There are approaches that may be used to evaluate the results. The frequency of occurrence of a particular event in the conformational dynamics may represent stable states. Furthermore, the results can be statistically verified by comparison with experimental results. In this paper we report multiple simulations of insulin chain B with the view of gaining statistically relevant insights into protein dynamics. Chain B was selected as it is believed to fold independently of chain A [26]. This stability has also been observed in NMR structures of the isolated chain in solution [10,27]. Apart from its biological relevance, we have chosen this protein as a model system to investigate because the availability of extensive experimental and theoretical results enables comparison with our simulations.

In this work, we analysed the conformations sampled and the conformational transitions occurring in a series of 20 ns simulations. We also performed a simulation at 400 K to investigate the usefulness of high temperature as a means of "speeding up" the conformational sampling. We compared our results with the NMR data on isolated chain B in a low pH solution.

2. Methods

The molecular dynamics simulations were performed using the program NAMD [47] in combination with the