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

CHARMM27 force field [48]. The starting structure was the R^f conformational state obtained from the crystal structure of porcine insulin (PDB code 1ZNI [15]). The His residues were protonated, and both termini regions and titratable sidechains were charged, resulting in a neutral to acidic pH environment. This allowed for a more direct comparison with the NMR structure solved in a low pH environment [10], which is used for validation of our MD structures (discussed in the Results section).

2.1. Computational details

The molecule was immersed in solvent using the InsightII program (Accelrys, San Diego, CA, U.S.A.), with a total of 7002 water molecules of density of 1 g/cm^3. The solvated protein was placed in a cubic box with periodic boundary conditions and the system's energy was minimized to remove steric clashes. The NVT (constant number of particles, constant volume, and constant temperature) thermodynamic ensemble, followed by NPT (constant pressure) ensemble, was simulated during the equilibration of the system. The data collection stage was then performed in an NPT ensemble with a constant pressure of 1 bar for 20 ns with piston decay time of 40 ps. Bonds containing hydrogen atoms were constrained using the SHAKE algorithm to their energy minimized values, thus allowing a numerical integration time step of 2 fs to be used in MD simulation. Four simulations were performed at room temperature (300 K) changing the initial conditions by assigning random velocities according to Maxwell distribution to explore wider conformational space. An additional simulation was performed at an elevated temperature (400 K) to test the temperature effects on protein dynamics and accelerate conformational sampling. At equilibrium, the temperature of the system was maintained by coupling to an external bath [49]. The Particle Mesh Ewald (PME) summation [50] was applied to treat long range electrostatic interactions. Atom-based cutoff of 12 Å with switching at 10 Å was used for non-bonded van der Waals interactions. The trajectory data were saved for analysis every 2500 steps (corresponding to 5 ps) generating 3840 structures over approximately 20 ns.

2.2. Structural characterization

Visualization and manipulation of the conformations were performed using the program VMD [51]. The trajectories were analysed and classified using the PEPCAT program [52], a computational tool that enables not only the classification of conformational states and the relative population of each state, but also the transitions between each state to be identified and quantified. PEPCAT is particularly useful in monitoring changes in the specific elements of the secondary structure of proteins. The conformations from the MD trajectory are processed according to a set of specified geometric descriptors (e.g. atom–atom distances, ϕ and ψ dihedral angles), and are classified based on similarity in the descriptors. The conformational changes that occur during the MD simulation are reflected in the classified states, and can be visualised in conformation transition maps. This enables the dynamic pathways of the protein to be characterised and compared.

3. Results and discussion

The protein stability during the MD simulations was assessed by calculating the root mean square deviation (RMSD) between the structures generated from the simulations and the starting structure. As a means of validation, a comparison was then made with the NMR experimental data. The conformational dynamics were analysed by examining the secondary structure, conformational clustering, and monitoring the distance between residues over the simulation time course. Below we provide a detailed analysis of each insulin simulation at 300 and 400 K, labelled systems 300[1–4] and system 400, respectively.

3.1. Stability and validation

3.1.1. RMSD

The central helix residues (Ser 9 to Cys 19) in each simulation were aligned and compared to the minimized structure. The RMSD of the backbone of the central helix for the five systems is shown in Fig. 1. The plot shows a moving average of 250 ps to enable effective comparison of the overall trends between the systems. The RMSD was generally stable, even at the elevated temperature, however some interesting variations occurred. In the first 300 K simulation (system 300[1]), the system showed overall stability, although several sharp increases in RMSD reflect distinct conformational transitions (at approximately 3.0 ns the RMSD increases to 2.50 Å and at 11.5 ns to 3.25 Å). As expected, the RMSD at 400 K shows persistent variation, although within the upper and lower limits observed in the 300 K systems.

3.1.2. Comparison with NMR experimental results

The data derived from the NMR structure of bovine insulin chain B in solution [10] provide a useful benchmark for evaluating the conformations observed during the MD calculation. Hawkins et al. [10] derived nuclear Overhauser effect (NOE) constraints from a 250 ms NOESY spectrum of the oxidized chain B at 500 MHz, 300 K and pH of 2.2 to 2.5. As coordinates for the structure were not available, NOE distance restraints were used for comparison. Using the program XPLOR [53,54], the interproton distance restraints were used to calculate energy values for the MD conformations from the five trajectories. To evaluate our results, we compared these values with the NOE energy

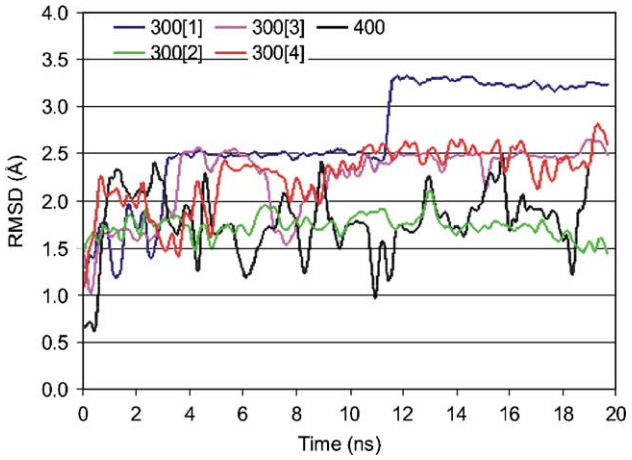

Fig. 1. RMSD of backbone atoms of residues 9 to 19 compared to the starting structure as a function of time for all systems. The plot shows a moving in average of 250 ps.

value of the minimized structure. The NOE energy plot of the structures generated from the five simulations is shown in Fig. 2. We found that the majority of the modelled structures satisfied the distance restraints with lower energy values compared to that of the minimized crystal structure. It is interesting to note the correlation at 11.5 ns between the decrease in NOE energy and the previously observed sharp increase in RMSD in system 300[1]. This indicates the changes in conformation which occurred at this time point had fewer NOE violations, i.e., the structures were in closer agreement with the experimental data. The exception to the generally low energy values occurred in one of the 300 K simulations (system 300[4]). In this simulation there was a noticeable increase in energy at about 8 ns indicating changes in conformation, which in this case, showed less agreement with experimental data. This observation stimulated further interest in the forces driving this divergence in structure, which will be discussed in later sections.

3.2. Structural analysis

3.2.1. Secondary structure

The secondary structure dynamics reveal details of conformational changes over the simulation time course. The secondary structure content as a function of time for the 300 (systems 1–4) and 400 K simulations is displayed in Fig. 3. The secondary structure was classified by VMD using the algorithm STRIDE which utilizes hydrogen bond energy and mainchain dihedral angles in addition to hydrogen bond distances [55].

The content of the secondary structure plot of each system shows variation, while they all share some common elements of secondary structure. All of the simulations showed an initial change in Cys 19 from the starting structure, most noticeable in systems 300[1] (Fig. 3(A)) and 300[3] (Fig. 3(C)). This can be expected because of the loss of the disulfide bonds with chain A. The helical content was maintained in each secondary structure plot although with some variation. The changes observed were most distinctive in the N-terminus of the central helix (residues 4 to 9). The secondary structure assignments for the first of the 300 K simulations system (Fig. 3(A)) showed a loosening of atom–atom contacts by 3 ns, resulting in the α-helix structure developing into π-helix structure and finally losing the hydrogen bonds of the helix altogether. Transitions between α-helix structure and π-helix have been previously reported in MD simulations. They are believed to be genuine physical transitions and an important stage in helix-to-coil transitions [56]. Conversely, a recent study has indicated that the occurrence of π-helix structure in MD may be an artefact of specific force fields [57]. However, this investigation of CHARMM, and other force fields, was performed using implicit solvent, which cannot necessarily be compared to simulations in explicit solvent, such as ours. The π-helix conformation is difficult to measure quantitatively by experiment. The conformation is rarely identified in crystal structures, although this could be attributed to the definition algorithm of helical structure. Modified π-helix definitions have shown that the π-helix is >10 times more prevalent than previously believed [58]. Despite the suggestion that the existence of π-helix is an artefact of the force field, there is sufficient evidence to suggest that the π-helix does occur in the protein's dynamic behaviour and plays an important role in the function of a protein [56,58].

Also of interest, was the change in structure at 8 ns, where the α-helix region between residues 4 and 7 reformed while accompanied by a change in structure of Gly 8. These results are consistent with experimental studies showing that Gly residues are often associated with perturbations in structure and are known as "helix-breakers" [59]. In addition, evidence of helical content in residues 4 to 7 was observed previously in the NMR solution structure of an insulin chain B mutant [27]. In our simulation, this trend was not observed in the 300[3] system where the helical

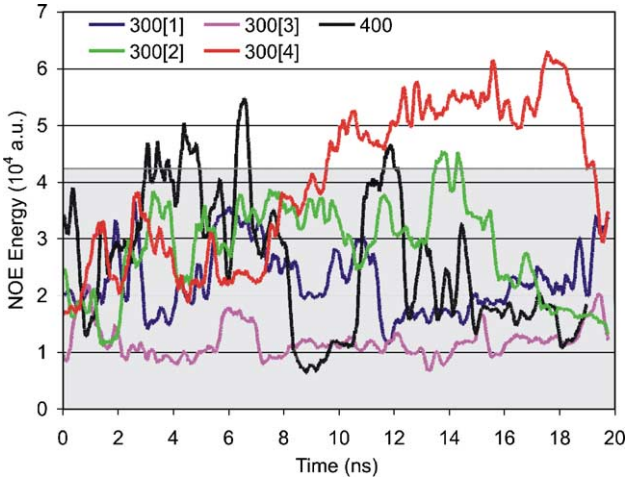

Fig. 2. The NOE energy derived from the distance restraints of the starting structure compared to all systems. The shaded region represents the MD structures with NOE energy less than the starting structure. The plot shows a moving average of 250 ps.

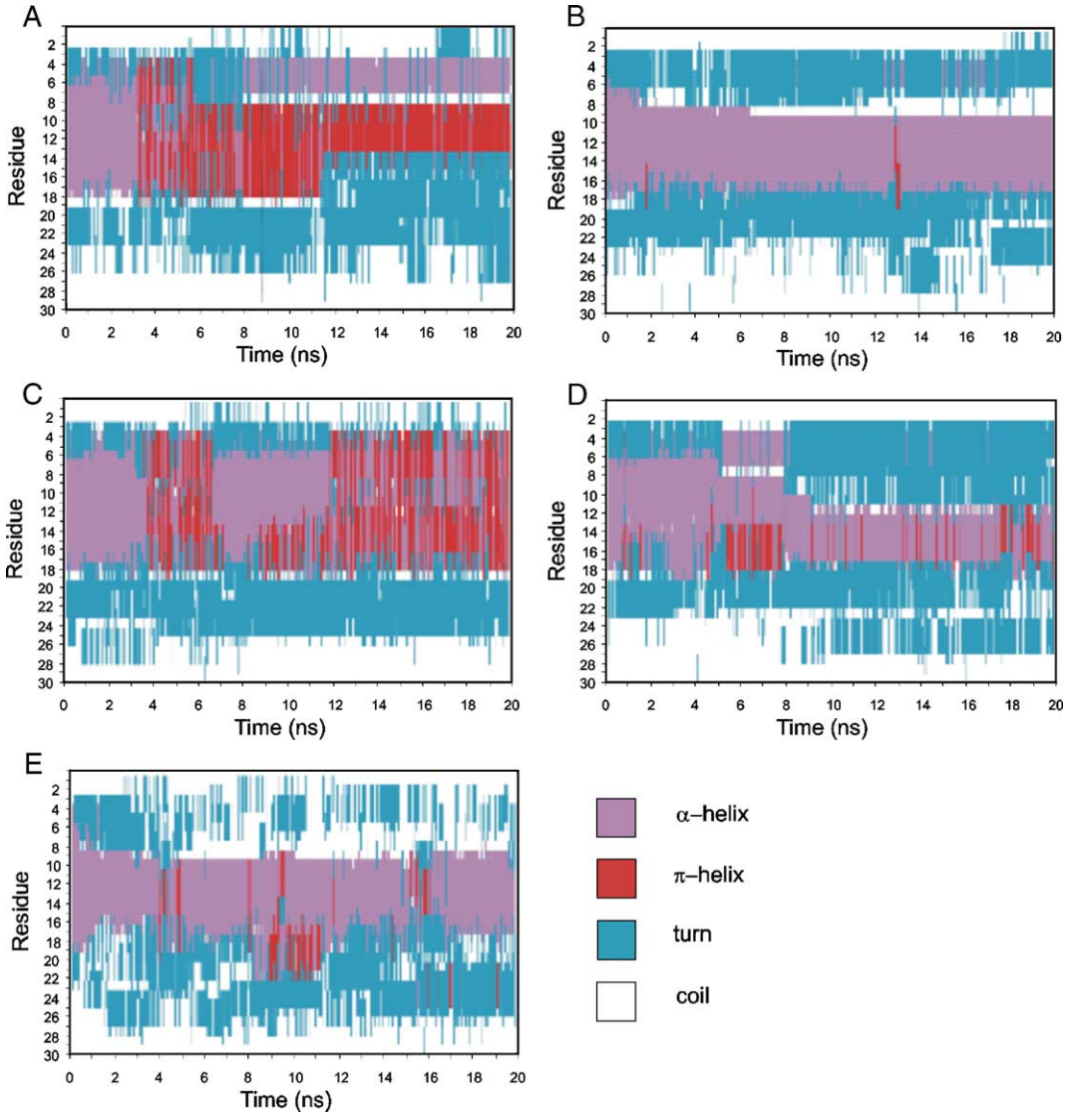

Fig. 3. Secondary structure as a function of time for all systems. (A) 300[1], (B) 300[2], (C) 300[3], (D) 300[4], (E) 400.

content was mostly maintained with a small amount of fraying at the N-terminus (Fig. 3(C)). The trend did occur, although to a lesser degree, in systems 300[2] and 300[4] (Fig. 3(B) and (D), respectively). The 300[4] system, however, showed variation where the simulation followed the same trend initially and then showed some distinctive differences. Between 5 and 8 ns, the secondary structure content showed strong similarities to the 300[1] system with a change in geometry of Gly 8 and the appearance of helical content in residues 4 to 7. However, at about 8 ns there was a noticeable loss of helical content in residues 8 to 12. It was at this time period in the simulation when a significant increase in the NOE energy was previously noted for this system (Fig. 2) suggesting the simulation may have become trapped in conformations corresponding to energetically higher local minima.

Another interesting secondary structure feature was the movement of the β-turn, again, most pronounced in the 300[1] system (Fig. 3(A)). During the simulation at about 11.5 ns, the turn moved from its initial position localized in residues 20 to 23 in the starting (crystal) structure to residues 17 to 20. This was in agreement with the turn-like structure defined in the oxidized NMR structure of chain B for residues 17 to 21 [10]. This movement was previously observed as an increase in RMSD (Fig. 1) accompanied by a sharp decrease in NOE energy (Fig. 2) confirming close agreement with experimental data.

In the 400 K simulation more secondary structure fluctuation was observed, however, the central helix was in general maintained (Fig. 3(E)). The helical content was reduced to residues 10 to 15, although there was a brief return to helical structure in the C-terminal end of the central helix. Notably, more movement was observed in the β-turn in the 400 K simulation, shifting in position between residues 16 to 27, thus reinforcing the view that the position of the turn is flexible.

3.2.2. Conformational transitions

The conformations sampled during the MD of each system were classified according to the structure of the N-terminus and central helix regions to identify stable states and transitions between them. The conformational states were defined by 17 geometric descriptors comprising the dihedral angles of residues 3 to 19. These residues are located in the central helix and the N-terminus (the first two residues were not included as their high flexibility would further complicate the analysis). The systems are classified sequentially according to the same set of descriptors resulting in the same rules defining the equivalent cluster in each system. The PEPCAT [52] conformational transition maps displaying the relative population of each conformation (states) and the transitions between the states for each system are shown in Fig. 4. The analysis of the first simulation (system 300[1]) in Fig. 4(A) revealed only a relatively small number of conformational states [10], where State 0 represents the initial R^f conformation. The most populated states are also shown. The sizeable number of transitions, which occurred between State 0 and State 7 at 5.5 to 8 ns, was the result of the change in conformation of Gly 8 noted previously in the secondary structure analysis. The conformations of typical structures representing States 0 and 7 shown in Fig. 5(A), illustrate the transition between the two states. It can be seen that this transition involved a change from the R^f-state (State 0) to a conformation that was very similar to the T-state conformation (State 7), which is believed to be the monomeric solution structure important for the hormone's activity. The State 7 structure, stabilized by a hydrogen bond between residues 5 and 13, will be discussed in more detail later.

The conformational analysis of the second 300 K simulation (system 300[2]) revealed an increased number of structures compared to that of the previous simulation (86). For clarity, only the major conformations are displayed in the transition map in Fig. 4(B) representing nearly 71% of the sampled structures. Conformational State 7 was only visited briefly and new pathways are followed through States 15, 32 and 48. The additional conformational states are largely structures with variations centred on Gly 8. Fig. 5(B) shows typical structures of States 15, 32 and 48. After a transition through State 15, States 32 and 48 are generally similar to the T-state conformations and are stabilized by hydrophobic interactions between the N-terminus and the central helix.

The stability observed in the secondary structure in the third 300 K simulation (system 300[3]) was reflected in the small number of conformational states [8] sampled. The transition map in Fig. 4(C) illustrates how the conformational changes did not progress much beyond State 0 (R^f-state). State 7, a key conformation observed in previous 300

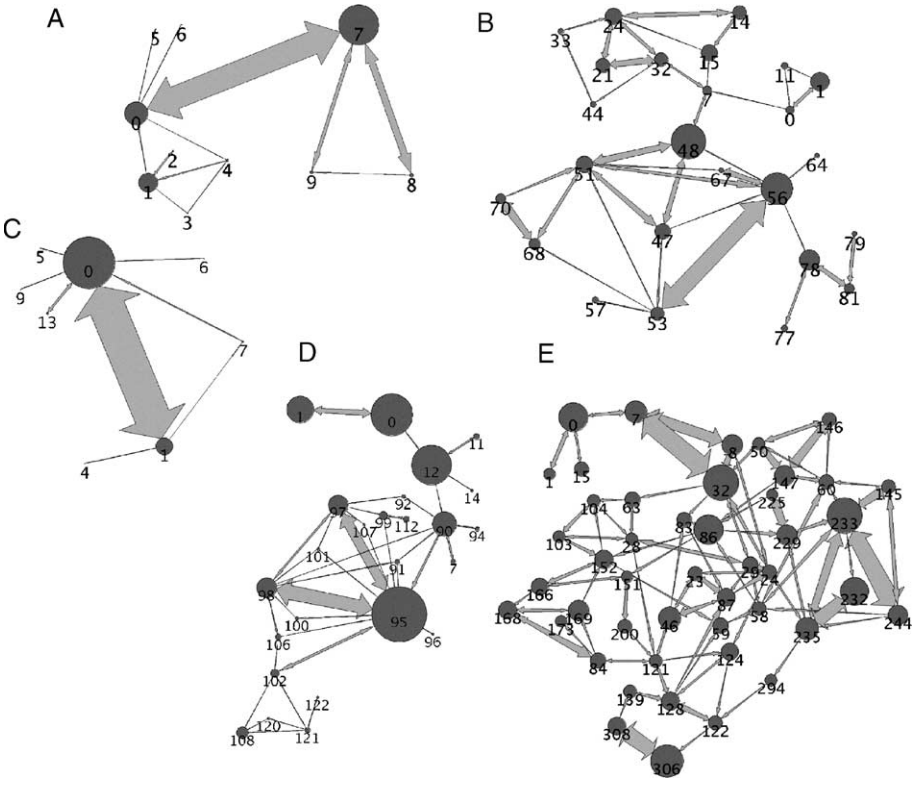

Fig. 4. Conformational transition map for all systems. (A) 300[1], (B) 300[2], (C) 300[3], (D) 300[4], (E) 400 K. The circles, labelled with the state identification number, represent conformational states visited during the simulation, and the size of the circle is proportional to the frequency. The lines between the circles indicate the conformational transitions and the width is proportional to the frequency. For classification of the states please refer to the text in the Results and discussion section.

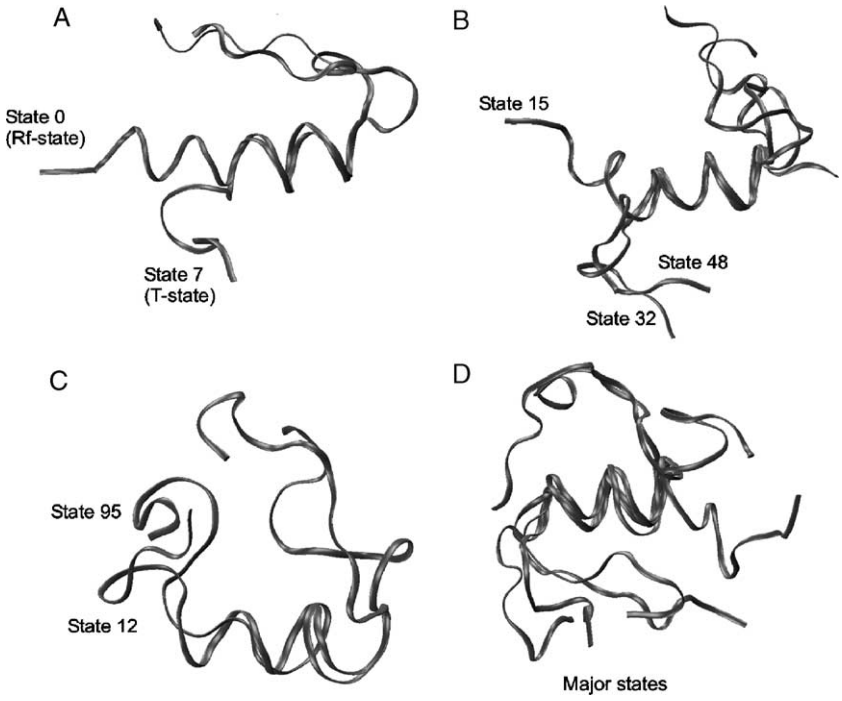

Fig. 5. Typical structures from the major conformational states for selected systems. (A) 300[1], (B) 300[2], (C) 300[4], (D) 400 K. The structures show the variation in the N-terminal region.

K simulations, was only briefly visited. The conformations generated from this simulation are low NOE energy structures (Fig. 2) and only represent a small subset of the already observed structures.

The conformational analysis in the fourth 300 K simulation (system 300[4]) revealed a moderately large number of structures classified [42]. The major conformational states representing 84% of the sampled structures are shown in the conformational transition map in Fig. 4(D). As indicated previously, there was evidence of an alternative conformational transition. Initially States 0 and 1 were well populated, however the conformations moved quickly along a pathway, only briefly visiting State 7, to a geometry which was largely centred on State 95. Fig. 5(C) shows typical structures of States 12 and 95, the most frequented conformations. It is worth noting the close proximity between the N- and C-termini in both States 12 and 95, indicating these structures are stabilized by hydrophobic interactions.

At 400 K there was a significant increase in the number of conformational states observed (310), as expected. The most populated states are shown in Fig. 4(E) representing 61% of the sampled structures. The purpose of increased temperature was to explore a broader region of conformational space in order to identify additional new states. Altogether, 46% of the states observed in the 300 K simulations were observed in the 400 K system (accounting for 34.5% of the total number of structures). The remaining states were classified as new. The conformations included many of the most populated ones previously observed at 300 K, for example, States 0, 7, 15 and 32. Many additional conformations were observed through the transition States 8 and 32, and the conformations were again, largely differentiated around the pivotal turning point of Gly 8 with increased mobility of the extended region of the N-terminus. The variations in the conformations of the major states are displayed in Fig. 5(D).

Even though our structural analysis showed a diverse range of conformations in all systems at 300 and 400 K, overall 86% of structures satisfied the experimental NMR distance restraints. A comparison of the structures and transitions shows similarities in the conformations generated in the time course of the simulations. Many of the experimentally known structural properties were reproduced in the simulations. The central α-helix (residues 9 to 19) was largely conserved in all the simulations apart from some fraying occurring at the ends. This would suggest that the central helix is the most stable region and possibly the folding nucleus of the molecule. Evidence of helical structure was also observed in residues 4 to 7 although this was far more transient in frequency. We observed the appearance of stabilizing interactions, such as between the residues 4 to 7 and the central helix. However, one simulation at 300 K showed some distinct differences where the conformational dynamics leads to structures with lower NOE energy. A comparison of the 300[1] and 300[4] systems revealed more detail of this divergence and is schematically illustrated in Fig. 6. In both simulations there was a change in geometry of Gly 8 between 5 and 6 ns. In the 300[1] system, the change in Gly 8 conformation leads to a change in the central helix characterized by a movement of the β-turn. This leads to a downward swing of the N-terminus with a corresponding outward movement of the C-

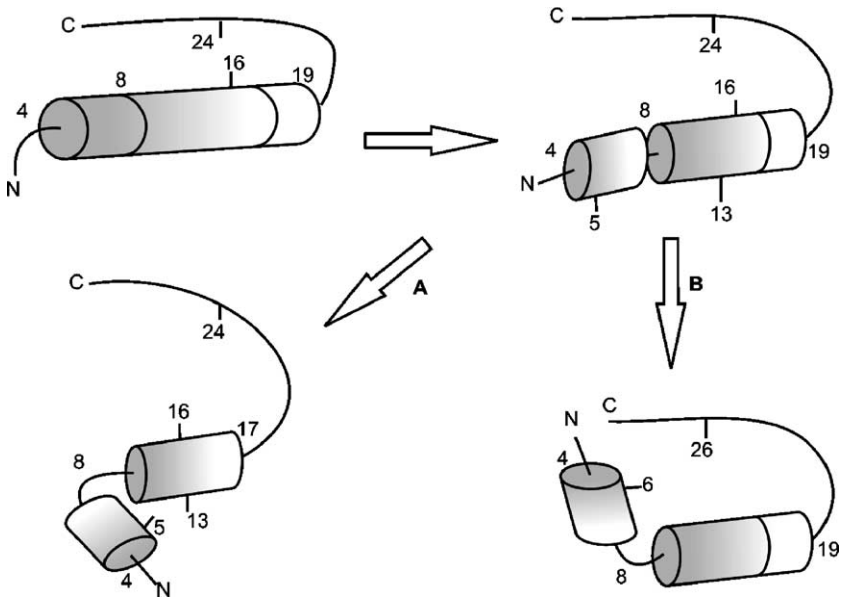

Fig. 6. Conformational changes of insulin chain B at 300 K. The initial change in structure of Gly 8 leads to two distinct conformations: Pathway A leads to T-state like conformation with low NOE energy in the 300 K[1] system; Pathway B leads to conformations trapped in local energy minima with high NOE energy in the 300 K[4] system. Both structures are stabilized by localized hydrophobic interactions.

terminus breaking the hydrophobic interactions with the central helix (illustrated in Figs. 5(A) and 6). This resulted in a T-state conformation. Conversely, in the 300[4] system (Fig. 5(C)) the hydrophobic interaction between the N- and C-termini resulted in the second helix (residues 4 to 7) being pulled up towards the C-terminus instead of down as in 300[1]. Both sets of conformations were effectively trapped in local energy minima and stabilized by hydrophobic interactions.

The correlation between the NOE energy structures and the distance between the N- and C-termini for the two systems is shown in Fig. 7. In the 300[1] system there was a marked drop in NOE energy associated with an increase in the distance between the two termini regions and the more T-state like conformation. In the 300[4] system there was a marked increase in NOE energy associated with a decrease in distance between the two regions. This reinforces the importance of the T-state conformation.

The T-state like conformation observed in our simulations is characterised by interactions between the N-terminal region and the central helix. This association is illustrated by the time history plot of the distance between residues 5 and 13 shown in Fig. 8. These two residues were important in the stabilization of the conformations in the 300[1] system by a hydrogen bond. The plot shows the strong interaction between these regions occurring in the 300[1] system. This was not evident in the 300[4] system where the plot showed the distance between these two residues markedly increased, a divergence occurring at approximately the same time in the simulation. This is a reflection of the different conformational pathways observed in these systems. Fig. 8 also shows the distance over time in the structures from the high temperature simulation (system 400). Although the distance between the residues was not as close as that observed in the 300[1] system, the high temperature system does appear to avoid

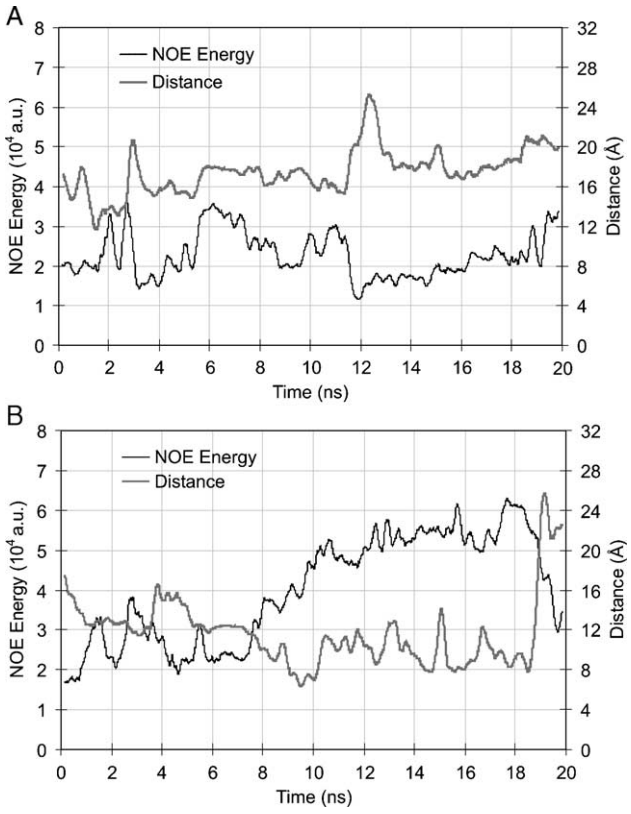

Fig. 7. Comparison of NOE energy and distance (Cα atoms) between N- and C-termini in selected systems (residues 6 to 26). (A) 300[1]. (B) 300[4]. The plot shows a moving average of 250 ps.

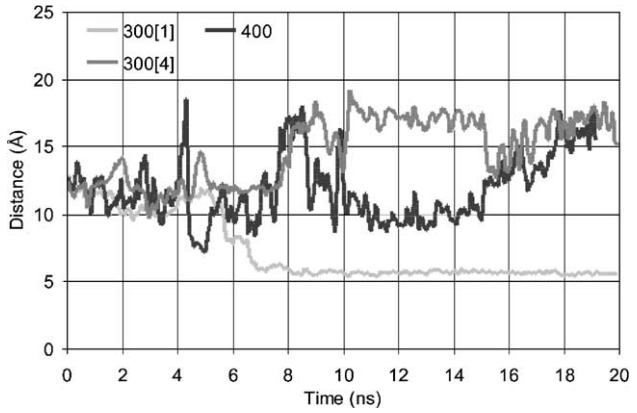

Fig. 8. Time dependence plot of the distance between residues 5 and 13 (Cα atoms) for 300[1], 300[4] and 400 systems. The plot shows a moving average of 250 ps.

the structural entrapment which occurred in the 300[4] system.

Whilst there are differences in the order and population of the conformational states observed in the 400 K system, the structures generally satisfied the NOE experimental constraints, and there was a significant amount of residual helical structure maintained. These reasons suggest high temperature simulations are a valid means of avoiding entrapment by local minima on the potential energy surface. High temperature exploration of a wider conformational space results in the production of viable conformations, and possibly an intimation of unfolding.

3.3. Biological implications

Our results show a high degree of flexibility in both termini regions of insulin chain B with many transitions between the R^f- and T-states. We have observed the consistent presence of a "T-state" type structure in our analysis, which can indicate the relevance of the T-state conformation to the biological function. Some of the trends observed in our simulations, including distinct conformational changes, may provide explanation for the protein's behaviour.

The central helix region has considerable impact on the dynamics of the protein and also plays a key role in the protein's activity. Most of this region has retained its integrity throughout the simulation, even at 400 K. Gly 8 is highly conserved in all species of insulin, with experimental studies showing a loss of activity if Gly 8 is replaced by other residues [60–62]. Gly is able to accommodate a range of dihedrals in the Ramachandran plot. Our results indicate that because of its unique geometric properties, Gly 8 acts as a pivotal point between the conformations necessary for activity. Examination of our simulated structures revealed evidence of a hydrogen bond commonly occurring between the sidechains of His 5 and Glu 13 thus having the effect of stabilizing the central helix. Fig. 9(A) shows the starting structure, and Fig. 9(B) shows the interaction between His 5 and Glu 13 in a low NOE energy structure from the 300[1] system. It must be noted that experimental results show that Glu 13 mutants result in decreased binding affinity [62], therefore confirming our findings of the important role of this residue in stabilizing the protein's structure.

The movement or flexibility of the β-turn from residues 20–23 (X-ray structure) to residues 17–20 resulted in the C-terminus moving away from the central helix thus exposing hydrophobic residues important for activity. The importance of the destabilization of the β-turn and C-terminus with respect to the receptor binding has been suggested in experimental studies [32]. Furthermore, this may also partly explain why alanine scanning mutagenesis experiments show enhanced binding affinity for Gly 20 mutants [62]. Flexibility in the position of the β-turn implies that Gly 20, unlike Gly 8, is not relied on for its unique geometric properties.

Phe 24 in chain B is of particular interest because this residue has been shown to play an integral role in hormone–receptor binding [29,36,63]. In the theory discussed by Weiss et al. [64], the receptor-binding surface of insulin is defined in terms of open and closed conformations. This theory suggests the role of Phe 24 in chain B. In the closed conformation Phe 24 anchors the conformation via hydrophobic interactions with residues 14 to 20. In the open conformation, Phe 24 moves away from the helix region, thus exposing hydrophobic residues in the central helix. The transition between

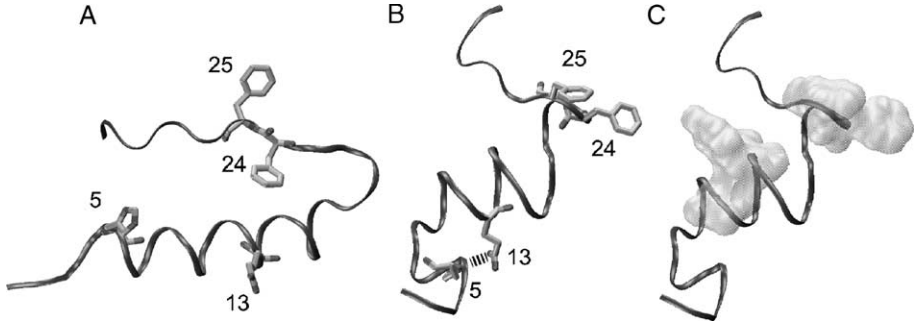

Fig. 9. Illustration of the role of individuals residues. (A) Starting structure. (B and C) Low NOE energy structure (300[1] system). A comparison between structures (A) and (B) shows the stabilization of the T-state conformation by a hydrogen bond interaction between the sidechains of His 5 and Glu 13; and the movement of the sidechains of Phe 24 and Phe 25 between the closed (A) and open (B) conformation. Structure (C) shows the two distinct binding sites of insulin to its receptor exposed with movement of the C-terminus.

the two conformations was observed during the simulations and is illustrated in Fig. 9. The starting structure in Fig. 9(A) represents the closed conformation and the simulated structure in Fig. 9(B) shows the position of sidechains of Phe 24 and Phe 25 in the open conformation.

Since insulin is cross-linked to the receptor dimer, it has been suggested that there are two distinct binding sites of insulin to its receptor [65]. Residues from both chain A and chain B are thought to be involved in the binding site [37,66]. The residues identified in chain B include residues 12, 15, 16 located in the central helix, and residues 23 to 25 in the C-terminus. The simulated movement of the C-terminus exposing the two distinct binding sites is displayed in Fig. 9(C).

4. Conclusion

Variation observed in the experimental structures of insulin raises the question of the biological implications of the inherent flexibility of insulin, including the conformation dependency of receptor binding. Our results have yielded information clarifying uncertainties about the structure and dynamics of insulin with respect to its behaviour. We carried out multiple simulations on insulin chain B and compared the conformational behaviour in each trajectory. Even though our sampling was limited, similar overall structural trends were observed. Gly 8, a residue known to commonly act as a helix-breaker, plays a key role in the structural changes over the simulation time course. Our calculations highlight the importance of packing interactions in determining the conformational behaviour of chain B of insulin, producing conformations stipulated by localized hydrophobic interactions. We were able to predict features observed in the NMR structure, including generally satisfying NMR experimental NOE distant restraints rather than the crystal structure. The high temperature simulations were found to be suitable for clarifying the details of protein dynamics at a smaller computational expense. This illustrates the potential of MD to represent the protein in solution, providing an important complement to fixed structure studies, such as crystal structures, that produce limited information about the dynamics of the molecule. MD can therefore be a viable tool to complement existing experimental methods in the investigation of protein flexibility in areas such as protein misfolding diseases. For example, the tendency of some proteins (including insulin) to misfold and form amyloid fibrils, a poorly understood mechanism forming the basis of many degenerative diseases.

Acknowledgements

The authors acknowledge the Australian Research Council (ARC) and Cytopia Pty. Ltd. for providing funding for this project, and the Australian Partnership for Advanced Computing (APAC) for the grant of computer time. We also thank Dr. David Craik for supplying the data from the NMR study of oxidized chain B of insulin and Mark O'Donohue for his advice on the computer program PEPCAT.

References

[1] C.M. Dobson, M. Karplus, The fundamentals of protein folding: bringing together theory and experiment, Curr. Opin. Struck. Biol. 9 (1999) 92–101.
[2] J. Ma, M. Karplus, Molecular switch in signal transduction: reaction paths of the conformational changes in ras p21, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 11905–11910.
[3] J. Apostolakis, P. Ferrara, A. Caflisch, Calculation of conformational transitions and barriers in solvated systems: application to the alanine dipeptide in water, J. Chem. Phys. 110 (1999) 2099–2108.
[4] P. Ferrara, A. Caflisch, Folding simulations of a three-stranded antiparallel beta-sheet peptide, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 10780–10785.
[5] A. Caflisch, M. Karplus, Acid and thermal denaturation of barnase investigated by molecular dynamics simulations, J. Mol. Biol. 252 (1995) 672–708.
[6] T. Lazaridis, M. Karplus, "New view" of protein folding reconciled with the old through multiple unfolding simulations, Science 278 (1997) 1928–1931.
[7] K.B. Wong, J. Clarke, C.J. Bond, J.L. Neira, S.M. Freund, A.R. Fersht, V. Daggett, Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding, J. Mol. Biol. 296 (2000) 1257–1282.
[8] Y.M. Rhee, E.J. Sorin, G. Jayachandran, E. Lindahl, V.S. Pande, Simulations of the role of water in the protein-folding mechanism, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 6456–6461.
[9] U. Derewenda, Z. Derewenda, G.G. Dodson, R.E. Hubbard, F. Korber, Molecular-structure of insulin — the insulin monomer and its assembly, Br. Med. Bull. 45 (1989) 4–18.
[10] B. Hawkins, K. Cross, D. Craik, Solution structure of the B-chain of insulin as determined by 1H NMR spectroscopy. Comparison with the crystal structure of the insulin hexamer and with the solution structure of the insulin monomer, Int. J. Pept. Protein Res. 46 (1995) 424–433.
[11] A. Budi, S. Legge, H. Treutlein, I. Yarovsky, Effect of external stresses on protein conformation: a computer modelling study, Eur. Biophys. J. 33 (2004) 121–129.
[12] V. Zoete, M. Meuwly, M. Karplus, A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer, J. Mol. Biol. 342 (2004) 913–929.
[13] E. Jacoby, P. Kruger, J. Schlitter, D. Roper, A. Wollmer, Simulation of a complex protein structural change: the T<–>R transition in the insulin hexamer, Protein Eng. 9 (1996) 113–125.
[14] J. Badger, M.R. Harris, C.D. Reynolds, A.C. Evans, E.J. Dodson, G.G. Dodson, A.C.T. North, Structure of the pig insulin dimer in the cubic-crystal, Acta Crystallogr., B 47 (1991) 127–136.
[15] G. Bentley, E. Dodson, G. Dodson, D. Hodgkin, D. Mercola, Structure of insulin in 4-zinc insulin, NAT 261 (1976) 166–168.
[16] E. Ciszak, J.M. Beals, B.H. Frank, J.C. Baker, N.D. Carter, G.D. Smith, Role of C-terminal B-chain residues in insulin assembly: the structure of hexameric LysB28ProB29-human insulin, Structure 3 (1995) 615–622.
[17] U. Derewenda, Z. Derewenda, E.J. Dodson, G.G. Dodson, C.D. Reynolds, G.D. Smith, C. Sparks, D. Swenson, Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer, NAT 338 (1989) 594–596.

[18] Z.P. Yao, Z.H. Zeng, H.M. Li, Y. Zhang, Y.M. Feng, D.C. Wang, Structure of an insulin dimer in an orthorhombic crystal: the structure analysis of a human insulin mutant (B9 Ser->Glu), Acta Crystallogr., D Biol. Crystallogr. 55 (1999) 1524–1532.
[19] Q.X. Hua, M.A. Weiss, Comparative 2D NMR studies of human insulin and des-pentapeptide insulin: sequential resonance assignment and implications for protein dynamics and receptor recognition, Biochem. 30 (1991) 5505–5515.
[20] H.B. Olsen, S. Ludvigsen, N.C. Kaarsholm, Solution structure of an engineered insulin monomer at neutral pH, Biochem. 35 (1996) 8836–8845.
[21] G. Dodson, D. Steiner, The role of assembly in insulin's biosynthesis, Curr. Opin. Struck. Biol. 8 (1998) 189–194.
[22] A.D. Kline, R.M. Justice Jr., Complete sequence-specific 1H NMR assignments for human insulin, Biochem. 29 (1990) 2906–2913.
[23] Q.X. Hua, S.E. Shoelson, K. Inouye, M.A. Weiss, Paradoxical structure and function in a mutant human insulin associated with diabetes mellitus, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 582–586.
[24] S. Ludvigsen, M. Roy, H. Thogersen, N.C. Kaarsholm, High-resolution structure of an engineered biologically potent insulin monomer, B16 Tyr–>His, as determined by nuclear magnetic resonance spectroscopy, Biochem. 33 (1994) 7998–8006.
[25] Y. Zhang, J.L. Whittingham, J.P. Turkenburg, E.J. Dodson, J. Brange, G.G. Dodson, Crystallization and preliminary crystallographic investigation of a low-pH native insulin monomer with flexible behaviour, Acta Crystallogr., D Biol. Crystallogr. 58 (2002) 186–187.
[26] Z.S. Qiao, C.Y. Min, Q.X. Hua, M.A. Weiss, Y.M. Feng, In vitro refolding of human proinsulin. Kinetic intermediates, putative disulfide-forming pathway folding initiation site, and potential role of C-peptide in folding process, J. Biol. Chem. 278 (2003) 17800–17809.
[27] F.Y. Dupradeau, T. Richard, G. Le Flem, H. Oulyadi, Y. Prigent, J.P. Monti, A new B-chain mutant of insulin: comparison with the insulin crystal structure and role of sulfonate groups in the B-chain structure, J. Pept. Res. 60 (2002) 56–64.
[28] R.Z. Luo, D.R. Beniac, A. Fernandes, C.C. Yip, F.P. Ottensmeyer, Quaternary structure of the insulin–insulin receptor complex, Science 285 (1999) 1077–1080.
[29] R.A. Pullen, D.G. Lindsay, S.P. Wood, I.J. Tickle, T.L. Blundell, A. Wollmer, G. Krail, D. Brandenburg, H. Zahn, J. Gliemann, S. Gammeltoft, Receptor-binding region of insulin, NAT 259 (1976) 369–373.
[30] R.G. Mirmira, S.H. Nakagawa, H.S. Tager, Importance of the character and configuration of residues B24, B25, and B26 in insulin–receptor interactions, J. Biol. Chem. 266 (1991) 1428–1436.
[31] U. Derewenda, Z. Derewenda, E.J. Dodson, G.G. Dodson, X. Bing, J. Markussen, X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue, J. Mol. Biol. 220 (1991) 425–433.
[32] Q.X. Hua, S.E. Shoelson, M. Kochoyan, M.A. Weiss, Receptor binding redefined by a structural switch in a mutant human insulin, NAT 354 (1991) 238–241.
[33] G.D. Smith, G.G. Dodson, The structure of a rhombohedral R6 insulin hexamer that binds phenol, Biopolymers 32 (1992) 441–445.
[34] E. Ciszak, G.D. Smith, Crystallographic evidence for dual coordination around zinc in the T(3)R(3) human insulin hexamer, Biochem. 33 (1994) 1512–1517.
[35] I.t. Pittman, H.S. Tager, A spectroscopic investigation of the conformational dynamics of insulin in solution, Biochem. 34 (1995) 10578–10590.
[36] S. Ludvigsen, H.B. Olsen, N.C. Kaarsholm, A structural switch in a mutant insulin exposes key residues for receptor binding, J. Mol. Biol. 279 (1998) 1–7.
[37] E.N. Baker, T.L. Blundell, J.F. Cutfield, S.M. Cutfield, E.J. Dodson, G.G. Dodson, D.M.C. Hodgkin, R.E. Hubbard, N.W. Isaacs, C.D. Reynolds, K. Sakabe, N. Sakabe, N.M. Vijayan, The structure of 2zn pig insulin crystals at 1.5-a resolution, Philos. Trans. R. Soc. Lond., B 319 (1988) 369–456.
[38] G.D. Smith, D.C. Swenson, E.J. Dodson, G.G. Dodson, C.D. Reynolds, Structural stability in the 4-zinc human insulin hexamer, Proc. Natl. Acad. Sci.-Biol. 81 (1984) 7093–7097.
[39] N.C. Kaarsholm, H.C. Ko, M.F. Dunn, Comparison of solution structural flexibility and zinc binding domains for insulin, proinsulin, and miniproinsulin, Biochem. 28 (1989) 4427–4435.
[40] M.J. Adams, E. Dodson, G.G. Dodson, M. Vijayan, E.N. Baker, Structure of rhombohedral 2 zinc insulin crystals, NAT 224 (1969) 491–495.
[41] I.S.G. Peking, Insulin's crystals structure at resolution, Peking Rev. 40 (1971) 11–16.
[42] G.A. Bentley, J. Brange, Z. Derewenda, E.J. Dodson, G.G. Dodson, J. Markussen, A.J. Wilkinson, A. Wollmer, B. Xiao, Role of B13 Glu in insulin assembly. The hexamer structure of recombinant mutant (B13 Glu–>Gln) insulin, J. Mol. Biol. 228 (1992) 1163–1176.
[43] M. Karplus, A. Sali, Theoretical studies of protein folding and unfolding, Curr. Opin. Struck. Biol. 5 (1995) 58–73.
[44] P. Kruger, W. Strassburger, A. Wollmer, W.F. van Gunsteren, G.G. Dodson, The simulated dynamics of the insulin monomer and their relationship to the molecule's structure, Eur. Biophys. J 14 (1987) 449–459.
[45] A.E. Mark, H.J. Berendsen, W.F. van Gunsteren, Conformational flexibility of aqueous monomeric and dimeric insulin: a molecular dynamics study, Biochem. 30 (1991) 10866–10872.
[46] M. Falconi, M.T. Cambria, A. Cambria, A. Desideri, Structure and stability of the insulin dimer investigated by molecular dynamics simulation, J. Biomol. Struct. Dyn. 18 (2001) 761–772.
[47] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, K. Schulten, NAMD2: greater scalability for parallel molecular dynamics, J. Comp. Phys. 151 (1999) 283–312.
[48] A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J. Phys. Chem., B 102 (1998) 3586–3616.
[49] H.J.C. Berendsen, J.P.M. Postma, W.F. Vangunsteren, A. Dinola, J.R. Haak, Molecular-dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[50] T.E. Cheatham, J.L. Miller, T. Fox, T.A. Darden, P.A. Kollman, Molecular-dynamics simulations on solvated biomolecular systems — the particle mesh ewald method leads to stable trajectories of DNA, Rna, and Proteins, J. Am. Chem. Soc. 117 (1995) 4193–4194.
[51] W. Humphrey, A. Dalke, K. Schulten, VMD — visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.
[52] M. O'Donohue, E. Minasian, S.J. Leach, A.W. Burgess, H.R. Treutlein, PEPCAT — a new tool for conformational analysis of peptides, J. Com. Chem. 21 (2000) 446–461.
[53] Brunger, Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy, Q. Rev. Biophys. 26 (1993) 49–125.
[54] Brunger, X-PLOR, A Aystem for Crystallography and NMR, Yale University, New Haven, Connecticut, 1992.
[55] D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment, Proteins 23 (1995) 566–579.
[56] R. Armen, D.O. Alonso, V. Daggett, The role of alpha-, 3(10)-, and pi-helix in helix–>coil transitions, Protein Sci. 12 (2003) 1145–1157.
[57] M. Feig, A.D. MacKerell, C.L. Brooks, Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations, J. Phys. Chem., B 107 (2003) 2831–2836.
[58] M.N. Fodje, S. Al-Karadaghi, Occurrence, conformational features and amino acid propensities for the pi-helix, Protein Eng. 15 (2002) 353–358.

[59] Y. Okamoto, U.H.E. Hansmann, T. Nakazawa, Alpha-helix propensities of amino-acids studied by multicanonical algorithm, Chem. Lett. (1995) 391–392.
[60] Z.Y. Guo, Y.H. Tang, S. Wang, Y.M. Feng, Contribution of the absolutely conserved B8Gly to the foldability of insulin, Biol. Chem. 384 (2003) 805–809.
[61] H. Zhang, Y.H. Tang, Y.M. Feng, Possible role of B8Gly in insulin structural motif, Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai) 32 (2000) 480–484.
[62] C. Kristensen, T. Kjeldsen, F.C. Wiberg, L. Schaffer, M. Hach, S. Havelund, J. Bass, D.F. Steiner, A.S. Andersen, Alanine scanning mutagenesis of insulin, J. Biol. Chem. 272 (1997) 12978–12983.
[63] E.J. Dodson, G.G. Dodson, R.E. Hubbard, C.D. Reynolds, Insulins structural behavior and its relation to activity, Biopolymers 22 (1983) 281–291.
[64] M.A. Weiss, Q.X. Hua, W. Jia, S.H. Nakagawa, Y.C. Chu, S.Q. Hu, P.G. Katsoyannis, Activities of monomeric insulin analogs at position A8 are uncorrelated with their thermodynamic stabilities, J. Biol. Chem. 276 (2001) 40018–40024.
[65] U. Schell, M. Grun, R. Hilgenfeld, Binding of insulin to its receptor: towards an understanding in three dimensions, ChemBiochem 1 (2000) 37–40.
[66] S. Gammeltoft, Insulin receptors: binding kinetics and structure–function relationship of insulin, Physiol. Rev. 64 (1984) 1321–1378.