Biophysical Chemistry 168–169 (2012) 1–9

Contents lists available at SciVerse ScienceDirect

Biophysical Chemistry

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

A mechanistic insight into the amyloidogenic structure of hIAPP peptide revealed from sequence analysis and molecular dynamics simulation

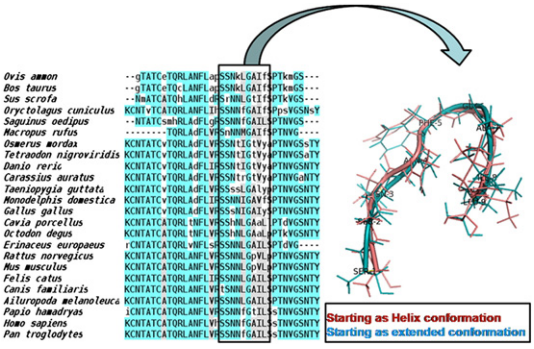
Sandipan Chakraborty [a,*,1], Barnali Chatterjee [b], Soumalee Basu [c,*]

[a] Saroj Mohan Institute of Technology, Guptipara, Hooghly, Pin 712512, India
[b] DOEACC Centre, Jadavpur University, Kolkata 700032, India
[c] School of Biotechnology & Biological Sciences, West Bengal University of Technology, BF-142, Salt Lake, Sector-1, Kolkata 700064, India

HIGHLIGHTS

- Our study unravels the species-specific variations of amylin sequences.
- 17th, 22nd and 23rd residues are found to be very crucial for amyloid formation.
- MD study illustrates a four residue turn spanning from residue 22 to 25 of hIAPP.
- Proline25 primarily dictates the observed non-amyloidogenecity in case of rodents.
- Proline29 is involved in the second turn and not in the second β-strand of hIAPP.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 9 March 2012
Received in revised form 10 May 2012
Accepted 25 May 2012
Available online 1 June 2012

Keywords:
IAPP
Aggregation
Sequence analysis
Molecular dynamics
Fibril
Type II diabetes

ABSTRACT

A collective approach of sequence analysis, phylogenetic tree and *in silico* prediction of amyloidogenecity using bioinformatics tools have been used to correlate the observed species-specific variations in IAPP sequences with the amyloid forming propensity. Observed substitution patterns indicate that probable changes in local hydrophobicity are instrumental in altering the aggregation propensity of the peptide. In particular, residues at 17th, 22nd and 23rd positions of the IAPP peptide are found to be crucial for amyloid formation. Proline25 primarily dictates the observed non-amyloidogenecity in rodents. Furthermore, extensive molecular dynamics simulation of 0.24 μs have been carried out with human IAPP (hIAPP) fragment 19–27, the portion showing maximum sequence variation across different species, to understand the native folding characteristic of this region. Principal component analysis in combination with free energy landscape analysis illustrates a four residue turn spanning from residue 22 to 25. The results provide a structural insight into the intramolecular β-sheet structure of amylin which probably is the template for nucleation of fibril formation and growth, a pathogenic feature of type II diabetes.

1. Introduction

Many human diseases, i.e., Alzheimer's disease, type II diabetes, spongiform encephalopathies, Creutzfeldt–Jakob's disease, Parkinson's disease, etc. are related with a common pathogenic process, known as "amyloidogenesis" [1–4]. In this process proteins or peptides undergo conformational changes becoming prone to form toxic soluble oligomers and insoluble fibrils. Interestingly, amyloidogenic proteins or

* Corresponding authors. Tel.: +91 33 23210731; fax: +91 33 23341030.
E-mail addresses: sandipanchakraborty.13@gmail.com (S. Chakraborty), soumalee@gmail.com (S. Basu).
[1] Present Address: Laboratory of Computational Biophysics & Bioengineering, Department of Physics, Tougaloo College, Tougaloo MS 39174, USA.

0301-4622/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2012.05.003

peptides in its native form belong to diverse structural folds and do not share any sequence homology. However, in fibrillar form they show a very similar topology of "cross β-strand" in which the β-sheets are parallel to the fibril axis and β-strands orientated perpendicular to the fibril axis [1]. A very similar hydrogen bonding pattern between neighboring β-strands is also evident from solid-state NMR study [5]. The common structural fold of the fibrillar form of the amyloidogenic proteins/peptides connotes a similar mechanism of amyloidogenesis in all those diseases [6]. Human type II diabetes is pathologically characterized by the presence of extra-cellular islet amyloid deposits which result in peripheral insulin resistance and impaired insulin response to increased glucose levels [7,8]. These deposits are believed to cause pancreatic β-cell dysfunction by direct cytotoxicity or by reducing beta cell mass [9]. Both small oligomers, as well as larger fibrils are thought to have cytotoxic properties [10]. The main constituent of islet amyloid is a 37 residue polypeptide called islet amyloid polypeptide (IAPP or amylin) which is co-synthesized and co-secreted with insulin by pancreatic β-cells [11,12]. Several functions have been associated with the soluble form of this hormone [13,14], including the control of hyperglycemia, inhibition of gastric emptying and pancreatic enzyme secretion. Amylin's function appears to be complement with insulin, which primarily controls glucose disposal. In addition, Amylin has the ability to lower plasma calcium (Ca^{+2}) concentration and inhibit osteoblast activity [15]. It is worth mentioning that diabetes-associated amyloid is found in primates [16] and cats but not in rodents [17].

The native structure of IAPP is unknown, although in aqueous buffer it seems to have a random-coil-like conformation, indicating that it may be a natively unfolded protein [18]. In amyloidogenic form, the peptide is generally believed to be composed of β-strands. Thus there is a dynamic transition which occurs in parallel with or as an alternative to physiological folding to generate aggregates. Amyloid formation can be interpreted with the nucleated polymerization model [19]. The rate-limiting step is the formation of a nucleus which is in equilibrium with monomers. This is evident from the observed lag phase in polymer growth, which is dependent on protein, ion concentrations and other experimental factors. The nucleus formation is followed by a rapid fibril growth [20]. Structural characterizations of these early conformational changes and oligomers are difficult to obtain experimentally as they are typically short lived. The aid of computational power has advanced our understanding of protein aggregation phenomenon. Using various simplified model to all atom representation and different simulation methodology including discontinuous molecular dynamics, simulated annealing to replica exchange dynamics reveal new insight into the aggregation processes [21–25]. Modeling such a competition between folding and aggregation still remains a challenge in current computational chemistry. Despite the significant advancement of computational power and simulation methodology, the mechanism of amyloidogenesis is elusive. All-atom MD simulation in the explicit solvent has been applied to study the stability of NFGAIL protomer and fibrils [26,27]. Simulations on dimers [28], trimers [29] and tetramers [30] in aggregated form provide insight into the thermodynamics and kinetics of aggregation.

Experimental studies indicate that there is more than one potential domain in the human IAPP sequence (hIAPP) capable of amyloid formation [18]. The presence of more than one potential β-strand in the peptide makes it an important macromolecule to analyze the interactions within a monomer during its folding to an amyloidogenic structure. Computational studies using simplified model have shown that variation in the ratio of intermolecular to intramolecular contacts decrease the aggregation rates. Melquiond et al. [31] interestingly showed that a rapid fibril growth seems to require a relatively small number of connections between the monomers using ART-OPEP simulation methodology of 12mer of hIAPP fragments 22–27 (NFGAIL). Although much effort has been devoted on the study of aggregated hIAPP fragments but the intrinsic folding profile of the monomer hIAPP is relatively unexplored. Recently, Andrews and Winter studied the critical differences between human and rat IAPP using MD simulation [32]. We are thus interested in studying the folding landscape of a hIAPP fragment which plays dominant role in the amyloidogenesis of hIAPP peptide to understand the very early conformational changes required for converting the peptide into the amyloidogenic form.

Here, a collective approach of sequence analysis and *in silico* prediction of amyloidogenecity using state of art bioinformatics tools have been used to delineate the observed species-specific variations in IAPP (islet amyloid polypeptide) sequences with the amyloid forming ability of the peptide. Furthermore, extensive molecular dynamics simulation has been carried out with human IAPP (hIAPP) fragment 19–27, the portion showing maximum sequence variation across different species, to understand the native folding characteristic of this region. To enhance the conformational sampling in MD simulation, we used an efficient sampling procedure where many replicas of the peptide with different initial velocity have been coupled together. Principal component analysis and free energy landscape analysis have been combined together to describe the folding landscape of the peptide. Our results present a theoretical insight into the structure of the assembly of the intramolecular β-sheets of hIAPP which probably is the template for nucleation of fibril formation and growth, the pathogenic feature of type II diabetes.

2. Materials and methods

2.1. Sequence analysis and amyloidogenecity prediction

Protein sequences were obtained from NCBI protein database. Redundant sequences were removed by using the CD-HIT program [33] with a cut-off 0.95. Multiple sequence alignment (MSA) was carried out using MUSCLE [34]. Phylogenetic analysis was performed using PhyML [35] and Phylogenetic tree was constructed using TreeDyn [36]. Amyloidogenic propensity of the sequences were analysed using AGGRESCAN [37] program. AGGRESCAN is a web-based program for the prediction of aggregation-prone regions in input protein sequences. The prediction method implemented in the program is based on an aggregation-propensity scale for natural amino acids derived from *in vivo* experiments. The program calculates aggregation-propensity values per amino acid (aaAV, or a^3v) followed by the a3v average (a^4v) over a sliding window of a given length and assigning the value to the central residue in the window. In the present study, we chose a window size of 5, as recommended by Sole et al. [37]. The program finally calculates the Na^4vSS which is the normalized a^4vSS score for 100 residues. We have referred the Na^4vSS value as AGGRESCAN score in the manuscript.

2.2. Molecular dynamics simulations

MD simulations of hIAPP19-27 fragment were performed using GROMACS 3.3 [38] with ffG53a6 force field. We have modelled the hIAPP19-27 peptide fragment both as a helix and an extended conformation due to lack of its native secondary structure preference information. Both the peptides were initially minimised *in vacuo* by 500 steps of conjugate gradient. Then the peptides were soaked in a cubic box of SPC water molecules with periodic boundary conditions. All the protein atoms were at a distance equal or greater than 1 nm from the box edges. The ionization state of the residues was consistent with neutral pH. Each peptide was then minimised by 500 steps of conjugate gradient followed by 100 ps position restrained dynamics where the peptide was kept fixed by adding restraining forces but water molecules were allowed to move. It was followed by 20 ps of NVT simulation at 300 K and by 20 ps of NPT simulation. Final production simulations were performed in isothermalisobaric (NPT) ensemble at 300 K, using an external bath with a coupling constant of 0.1 ps. Pressure was kept constant (1 bar) by using the time-constant for pressure coupling to 1 ps. The

LINCS algorithm was used to constrain bond-lengths, allowing the use of 2 fs time step. Electrostatic interactions were calculated using Reaction-Field method. Van der Waals and Coulomb interactions were truncated at 1.4 nm and conformations were stored every 50 ps.

To improve the conformational sampling, six independent 40 ns simulations (replicas), three starting from helix and three starting from extended conformation, were carried out initializing the MD runs with different initial atomic velocities. This resulted in a total of 0.24 μs production simulation. Analyses were carried out with the analysis tools of GROMACS and the secondary structure assignments were carried out with DSSP [39] module. The RMSD matrices were then computed on the concatenated trajectories by least square fitting on main-chain atoms and then processed using the GROMOS algorithm to extract clusters of similar conformations.

2.3. Principal component analysis (PCA)

Three independent replicas starting from a helix conformation with different seed values were concatenated. Likewise, three replicas starting from extended conformations were also concatenated. PCA was then performed on the concatenated trajectories. Mass-weighted covariance matrix of the atomic positional fluctuations was calculated on C_α atom and essential subspace was then described by the variance retained by the reduced representation defined by calculated eigenvectors. The free-energy landscape of the peptide was obtained from the conformational sampling by using the g_sham module implemented in GROMACS.

3. Results and discussions

3.1. Sequence and phylogenetic analysis

IAPP is expressed in pancreatic islets of almost all mammalian species so far examined but the potential to form amyloid is related to species-specific differences in the amino acid sequence of the peptide. Comparison of amylin sequences that are found in different eukaryotic lineages is shown in Fig. 1. Subsequently, species specific amyloidogenecity of these amylin sequences have been inferred by AGGRESCAN analysis (Fig. 2B). It has been already depicted during the validation of AGGRESCAN program that it successfully predicts the amyloidogenic regions of amylin and reproduces the experimentally observed effects of mutations on the aggregation propensity of amylin [37].

Results from multiple sequence analysis (MSA), phylogenetic tree and AGGRESCAN have been combined to delineate the observed

```
Ovis ammon                --gTATCeTQRLANFLapSSNkLGAIfSPTkmGS---
Bos taurus                --gTATCeTQcLANFLapSSNkLGAIfSPTkmGS---
Sus scrofa                --NmATCATQhLANFLdRSrNNLGtIfSPTkVGS---
Oryctolagus cuniculus     KCNTvTCATQRLANFLIhSSNNfGAIfSPpsVGSNsY
Saguinus oedipu           --NTATCsmhRLAdFLgRSSNNfGAILSPTNVGS---
Macropus rufus            --------TQRLAdFLVRSnNNMGAIfSPTNVG----
Osmerus mordax            KCNTATCvTQRLAdFLVRSSNtIGtVyaPTNVGSsTY
Tetraodon nigroviridis    KCNTATCvTQRLAdFLVRSSNtIGtVyaPTNVGSaTY
Danio rerio               KCNTATCvTQRLAdFLIRSSNtIGtVyaPTNVGSNTY
Carassius auratus         KCNTATCvTQRLAdFLVRSSNtrGtVyaPTNVGaNTY
Taeniopygia guttata       KCNTATCvTQRLAdFLVRSSssLGAlypPTNVGSNTY
Monodelphis domestica     KCNTATCvTQRLAdFLIRSSNNIGAVfSPTNVGSNTY
Gallus gallus             KCNTATCvTQRLAdFLVRSSsNIGAIySPTNVGSNTY
Cavia porcellus           KCNTATCATQRLtNFLVRSShNLGAaLlPTdVGSNTY
Octodon degus             KCNTATCATQRLtNFLVRSShNLGAaLpPTkVGSNTY
Erinaceus europaeus       rCNTATCATQRLvNFLsRSSNNLGAILSPTdVG----
Rattus norvegicus         KCNTATCATQRLANFLVRSSNNLGpVLpPTNVGSNTY
Mus musculus              KCNTATCATQRLANFLVRSSNNLGpVLpPTNVGSNTY
Felis catus               KCNTATCATQRLANFLIRSSNNLGAILSPTNVGSNTY
Canis familiaris          KCNTATCATQRLANFLVRtSNNLGAILSPTNVGSNTY
Ailuropoda melanoleuca    KCNTATCATQRLANFLVRSSNNLGAILSPTNVGSNTY
Papio hamadryas           iCNTATCATQRLANFLVRSSNNfGtILSsTNVGSNTY
Homo sapiens              KCNTATCATQRLANFLVhSSNNfGAILSsTNVGSNTY
Pan troglodytes           KCNTATCATQRLANFLVRSSNNfGAILSsTNVGSNTY
```

Fig. 1. Alignment of amylin sequences from different organisms.

amyloidogenecity/non-amyloidogenecity related to species specific sequence variations.

Calculation of amyloid forming propensity of individual amino acids by AGGRESCAN predicts two stretches of potential amyloid forming regions in amylin, namely residues 13–18 and 24–28 for most of the species. Since it is known that amylin sequences from *Rattus norvegicus, Mus musculus* and *Octodon degus* are non-amyloidogenic and their AGGRESCAN score lies between −8.80 and −11.00, other organisms like *Cavia porcellus* (−8.80), and *Sus scrofa* (−12.20) may be supposed to be non-amyloidogenic, based on the predicted AGGRESCAN score. *Tetraodon nigroviridis, Oryctolagus cuniculus* and *Osmerus mordax* are highly amyloidogenic as evident from positive AGGRESCAN score. It may be inferred that intermediate scores show moderate amyloidogenic propensity and with decreasing AGGRESCAN values, amyloidogenecity decreases. Close look into the phylogenetic tree (Fig. 2A) shows that non-amyloidogenic amylin sequences from rodents, *Cavia porcellus* and *Octodon degus* have clubbed together, while the amyloid forming amylins have given rise to a separate clade. The amylins from *Homo sapiens, Papio hamadryas* and *Pan troglodytes* sharing high level of sequence similarity group together in the phylogenetic tree. To relate the variation in the amyloidogenic propensities with sequence content, we have carried out an in-depth analysis of MSA with the amylin sequences (Fig. 1). The sequences show overall moderate conservation across different organisms with high conservation at the two termini. There are two highly conserved Cysteines at the 2nd and 7th position capable of forming di-sulphide linkage and probably are not involved in amyloid formation. This initial non-amyloidogenic region is followed by the 8th position, which is conserved with a sequence preference for Alanine/Valine. *Ovis ammon* and *Bos taurus* are exceptions where a non-familiar substitution of a hydrophobic residue occurs by a negatively charged residue (Glutamic acid). This substitution being on the very first residue of the first β-strand of the amyloid, might not be held responsible for disrupting the amyloid fibril. In *Saguinus Oedipus*, there is a Serine at this position followed by substitutions of highly conserved Threonine and Glutamine by Methionine and Histidine respectively. The substitution patterns at these positions explain the variability in the predicted amyloidogenicity of the protein in different organisms. It may be so that the residues at these positions influence the first out of the three β-strands that are believed to be involved in the amyloid formation by the peptide. At the 11th position, the highly conserved Arginine is substituted by Histidine in *Sus scrofa* and Cysteine in *Bos taurus*. The region 17–29 is significantly less conserved than the rest of the sequence. The 17th residue in most of the organisms is hydrophobic in nature. However, the presence of negatively charged Aspartate in *Sus scrofa*, might be disrupting the first β-strand thus explaining why AGGRESCAN predicts least amyloidogenecity for *Sus scrofa*. The 18th position is occupied by Arginine for most of the organisms with the exception of *Homo sapiens* and *Oryctolagus cuniculus* (Histidine). This similar substitution might not have any influence on the amyloidogenic propensity of amylin. But in case of *Ovis ammon* and *Bos taurus* a non-familiar substitution to β-sheet breaker Proline accounts for reduced amyloidogenecity predicted by AGGRESCAN. At the 21st position there is a basic residue Histidine instead of a highly conserved Asparagine, in case of *Cavia porcellus* and *Octodon degus*. This alteration may be responsible for the reduced predicted amyloidogenecity for the two organisms. In case of *Gallus gallus* and *Taeniopygia guttata*, a Serine substitution is observed. In addition to this, at the 22nd position there is a non-familiar substitution from Asparagine to Lysine in case of *Ovis ammon* and *Bos taurus*, to Serine for *Taeniopygia guttata* and to Threonine in case of *Osmerus mordax, Danio rerio, Carassius auratus* and *Tetraodon nigroviridis*. The substitution in *Ovis ammon* and *Bos taurus* drastically changes the hydrophobicity which results in reduced predicted amyloidogenecity. While, in *Taeniopygia guttata, Osmerus mordax, Danio rerio, Carassius auratus* and *Tetraodon nigroviridis* a similar substitution enhances the local hydrophobicity which is supported by higher AGGRESCAN score,

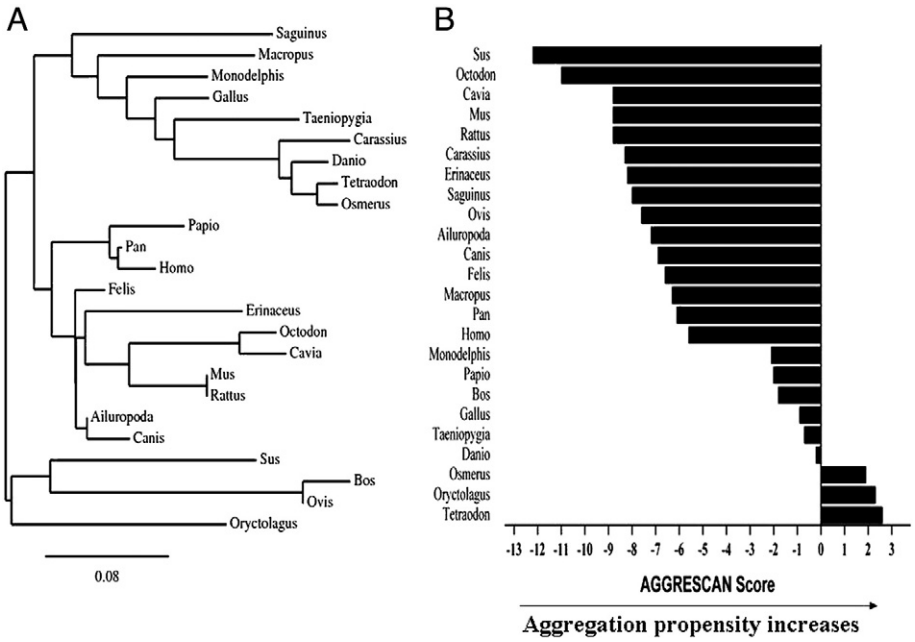


Fig. 2. A. Phylogenetic tree of amylin sequences. B. Amyloidogenic propensities (Na^4vSS values) of amylin sequences from different organisms calculated by using AGGRESCAN.

except for *Carassius auratus*. A highly polar hydrophilic Arginine at the 23rd position for *Carassius auratus* in place of a highly conserved hydrophobic residue is responsible for the loss of amyloidogenecity of the peptide shown by AGGRESCAN. However since the overall sequence similarity of the protein in the respective clade is high hence the amylin sequence of *Carassius auratus* has been clubbed with the amyloid forming amylins from other organisms in the phylogenetic tree.

Proline residue at the 29th position is almost conserved across different organisms which show diverse predicted amyloid propensity with *Homo sapiens, Papio hamadryas* and *Pan troglodytes* as exceptions. Hence the presence of proline at this position probably does not have any influence on the amyloid propensity. On the other hand for *Taenopaygia guttata*, the 28th residue is a Proline and AGGRESCAN predicts high amyloidogenecity of the amylin for that organism. Thus a particular Proline substitution at the 25th position may be primarily held responsible for the loss of amyloid formation ability of amylin in case of rodents. Though, it has been shown that proline at 28th position increases the helical propensity of rat IAPP peptide [32]. Presence of a conserved proline at the 29th position also questions the predicted second β-strand constituted by residues 24–29 [18]. This second β-strand should be shorter and the 29th residue should be involved in forming the turn that refolds the third β-strand on to the second one.

3.2. Molecular dynamics simulation of hIAPP19-27

Sequence comparisons of amylins from different organisms reveal that the region 19–27 is less conserved and substitutions in this region alter the amyloid forming ability of the amylin. Thus the intrinsic folding profile is worth studying to understand the mechanism of amyloid formation by amylin and also to understand the effect of sequence variations on the intrinsic folding profile and structural stability of amylin. We have chosen the hIAPP19-27 fragment for this purpose and carried out an extensive 240 ns MD simulation with parallel sampling method to understand the intrinsic folding profile of the peptide segment. Since key aspects in molecular dynamics simulations are adequate sampling and timescale of simulation, a simple parallel sampling method where many uncoupled copies of the same system with different initial conditions are coupled together have been used to improve the sampling during simulation, in thus study. This method was previously used to explore the conformational dynamics of much larger and complex protein myoglobin with comparable simulation time scale [40]. Comparing other studies [41–44], the time scale of MD simulation in our study is adequate enough to study the folding of peptide fragments of comparable or higher length.

3.2.1. Variation of RMSD and radius of gyration (R_g)

Fig. 3 depicts the variation of RMSD and R_g of the peptide as a function of time for three replicas starting with different initial atomic velocity distribution in helical and extended conformation. As is evident from the simulation, conformational subspace defined by each replica is significantly different from the others. It is believed that simulations starting from the same initial structure sample different direction of conformational space by using different initial velocity. Structural variations observed in simulations starting from extended conformation are higher than simulations starting from helical conformation. Simulation starting from helix conformation, i.e. the Helix1 simulation, is more stable, with C_α-RMSD fluctuating around 0.35 nm. In fact a closer look reveals that initial helical structure disappears very quickly within 500 ps thereafter stabilizing the RMSD fluctuations around 0.35 nm. Helix2 and helix3 simulations also contain the initial helix disruption signature. In Helix2, after the helix disruption, RMSD fluctuates around 0.55 nm and increases to 0.65 nm. While in the third replica, there is high fluctuation in RMSD till 10 ns, after which the RMSD fluctuation is around 0.25 nm. In contrast to the helix simulation, RMSD fluctuation in simulations starting from extended conformation is mostly around 0.6 nm in Sheet1 throughout the simulation. Sheet 2 and Sheet 3 simulations are more stable and fluctuate around 0.3 nm.

The radius of gyration (R_g) provides insight into the overall dimension of the peptide. The plot of R_g versus time starting with different initial velocity is shown in Fig. 3C and D. During initial periods of the simulation starting with the helix structure, it is being observed that the peptide adopts an elongated helical conformation. This is evident from the high R_g values during the first 15 ns of simulation for all the three replicas. As the simulation progresses, the peptide adopts a coil like conformation with a β-turn in the middle, evident from the progressive decrease of the R_g value during the remaining simulation period. For the simulation starting from the extended conformation the initial R_g value is around 0.8 nm which decreases to 0.6 nm with

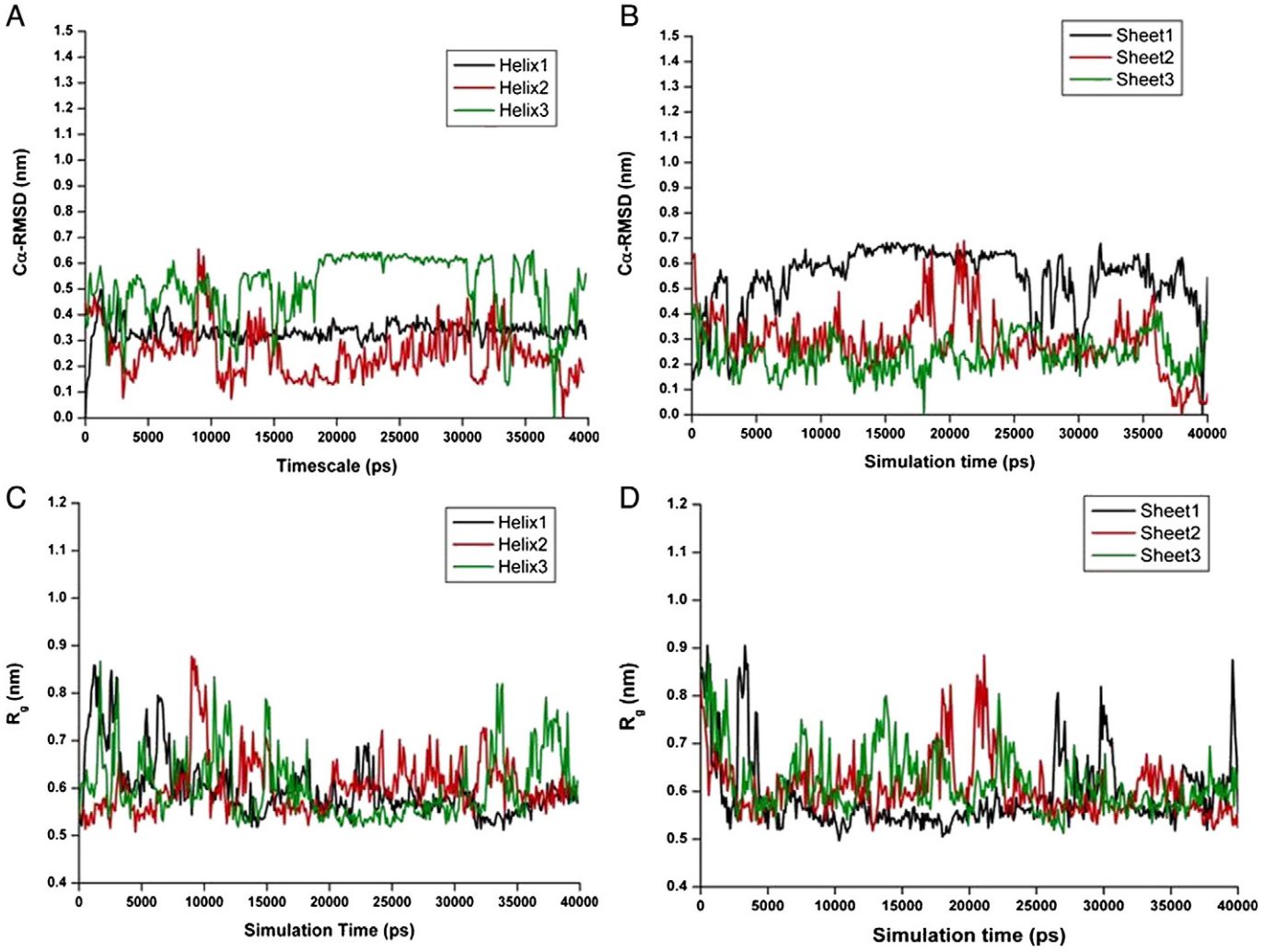


Fig. 3. A and B displays variations in Cα-RMSD of hIAPP fragments 19–27 with simulation time. C. and D. displays variations in radius of gyration (R_g) of hIAPP fragments 19–27 with simulation time.

the progress of the simulation. This observation can be interpreted by the folding of the peptide from an extended conformation to a bend like conformation. The mere appearance of spikes in the R_g curves can be explained in terms of partial unfolding of the peptide to an extended conformation.

3.2.2. Secondary structure analysis

Variation in $\Delta G_{solvation}$ with simulation time for each replica of the two (helix and sheet) simulations is very similar and it is ~32 kj/mol. Even the Solvent accessible surface area (SASA) of the peptide for the two simulations is very similar with a slight variation during 14–24 ns where helix simulation shows comparatively higher SASA than the simulation starting with extended conformation. Individual SASA components reveal a significant difference in the solvation dynamics of the hydrophilic residues account for the differences observed in both the simulations whereas those for the hydrophobic residues are very similar. This difference may be explained in terms of secondary structural elements (Fig. 4). For most of the simulation time the peptide adopts a bend structure specifically the four intermediate residues namely N, F, G and A (residue 22 to 25). An interesting observation is that there are three instances where the peptide adopts sheet-turn-sheet motif in replica3 of helix simulation in the time scale 19–26 ns, replica1 of the sheet simulation in the time scale 13-18 ns and the last 5 ns simulation of replica2 of the sheet simulation. This structure of the peptide is believed to be the structure that nucleates aggregation.

3.2.3. Principal component and free energy landscape analysis

Dimensionality of the essential subspace of the peptide is defined by the reduced subspace computed by principal component analysis. There are 261 eigenvectors needed to define the variations observed in the MD simulation. First few components are more informative as evident with high eigenvalues while the latter components carry very little information about the peptide dynamics (Fig. 5). In case of concatenated trajectories for helix and extended simulation, first three eigen vectors are most informative and can explain 49% and 47% of the total variance, respectively (Fig. 5). First principal component being the most informative explains 26% & 23% variance for helix and extended trajectories, respectively. We further analysed the cosine content of first 10 principal components for both the simulation types (Fig. 5, inset). Cosine content for the first 10 principal components is very low which signifies that the simulations represent the intrinsic dynamical behaviour of the peptide. It is to be mentioned that if the sampling of MD trajectories is insufficient, protein motions along the principal components appear indistinguishable from the dynamics of random diffusion and then the cosine content of the first few PCs is close to 1, a perfect cosine.

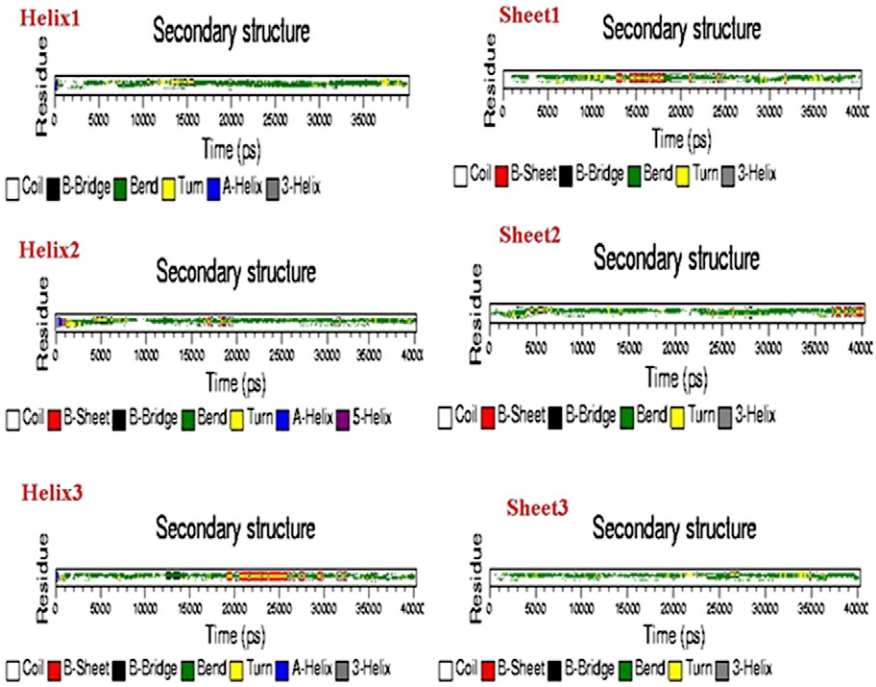


Fig. 4. Secondary structure evolution of hIAPP 19–27 of all the six replicas with simulation time.

To check whether the dynamical behaviour of the helix simulation is similar to the dynamics of the sheet simulation, we carried out the root mean square inner product (RMSIP) analysis. The calculated RMSIP value is 0.422 which indicates that the subspace defined by the concatenated helix trajectory is not identical to the defined subspace by the concatenated sheet trajectory. Thus consideration of both the concatenated trajectories is essential for more accurate representation of the folding energy landscape of the peptide. It is evident from Fig. 6 that for both the simulations, projection vectors are widely spread, reconfirming a good sampling of the peptide conformations during simulations. For both simulations, the first principal component (PC1) varies more broadly compared to the others. This is in accordance with the eigenvalue distribution plot where the first principal component has the highest eigen values in both the simulations. In case of helix simulation, projection of first two components reveal two highly populated clusters while simulation with extended conformation reveals mainly a single cluster of conformations. Further analysis of Gibbs free energy landscape (FEL) by Boltzmann inverting multidimensional histogram of first two principal components reveal two minima for helix simulation while for extended simulation, a single minimum is observed. The minima are designated by red dots while the orange dots signify meta-stable states of the peptides in the folding free energy landscape. We also have analysed the projection of first principal component of both the simulations along simulation time and found a multimodal distribution with two humps for helix simulation and a Gaussian distribution pattern with a single peak for extended simulation (Fig. S1). Thus the projection of first two PC from both simulation types also shows two distinct conformational sub-spaces for helix simulation while those components for extended simulation are confined within a single conformational sub-space. Cluster analyses based on RMSD matrices were used to provide further structural insight into the conformational dynamics of the peptide. Two distinct conformational sub-spaces observed from FEL analysis of concatenated helix trajectory represent two similar bend like structure with slightly different backbone orientation of the central turn residues. The average structure of the peptide obtained from the most populated cluster for helix and sheet simulation is very similar with an RMSD of 0.88 Å (Fig. 6C). Analysis of the projection vectors on the structure reveals that both the structure represents the observed minima in the free energy landscape. The convergence of parameters as cosine content, principal component combined with cluster analysis and the free-energy basins signifies a precise description of the conformational space has been achieved. Thus the minimum structure of the peptide obtained from MD simulation represents a native conformation of the peptide. Structural analyses reveal that the peptide adopts a broad bend like structure with a sharp turn comprising of Asn22, Phe23, Gly24 and Ala25 residues with the Alanine acting as a linker between the first turn and the second β-strand. This is in accordance with the recent molecular dynamics study of wild type hIAPP which predicts a significant turn propensity in the region of hIAPP 20–25 [45]. It is to be noted that the secondary structure prediction by using PSI-PRED reveals an overall coiled structure [46]. Interestingly, this region also

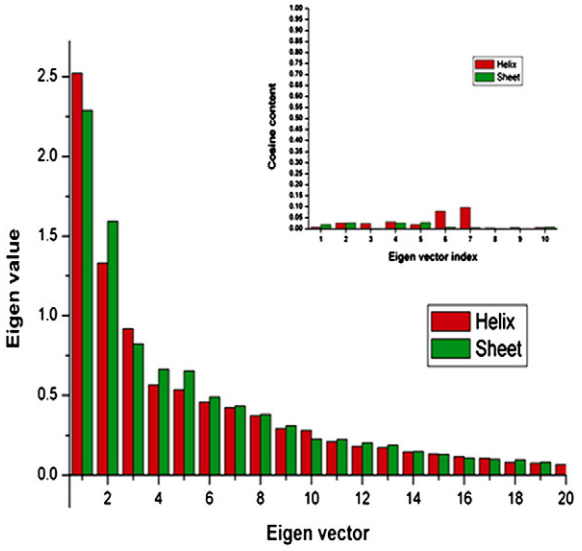


Fig. 5. Eigenvalue analysis of the calculated Eigenvectors of concatenated trajectory of helix (red) and sheet (green) simulation. Cosine content of first 10 eigenvectors of the respective simulation is shown in inset.

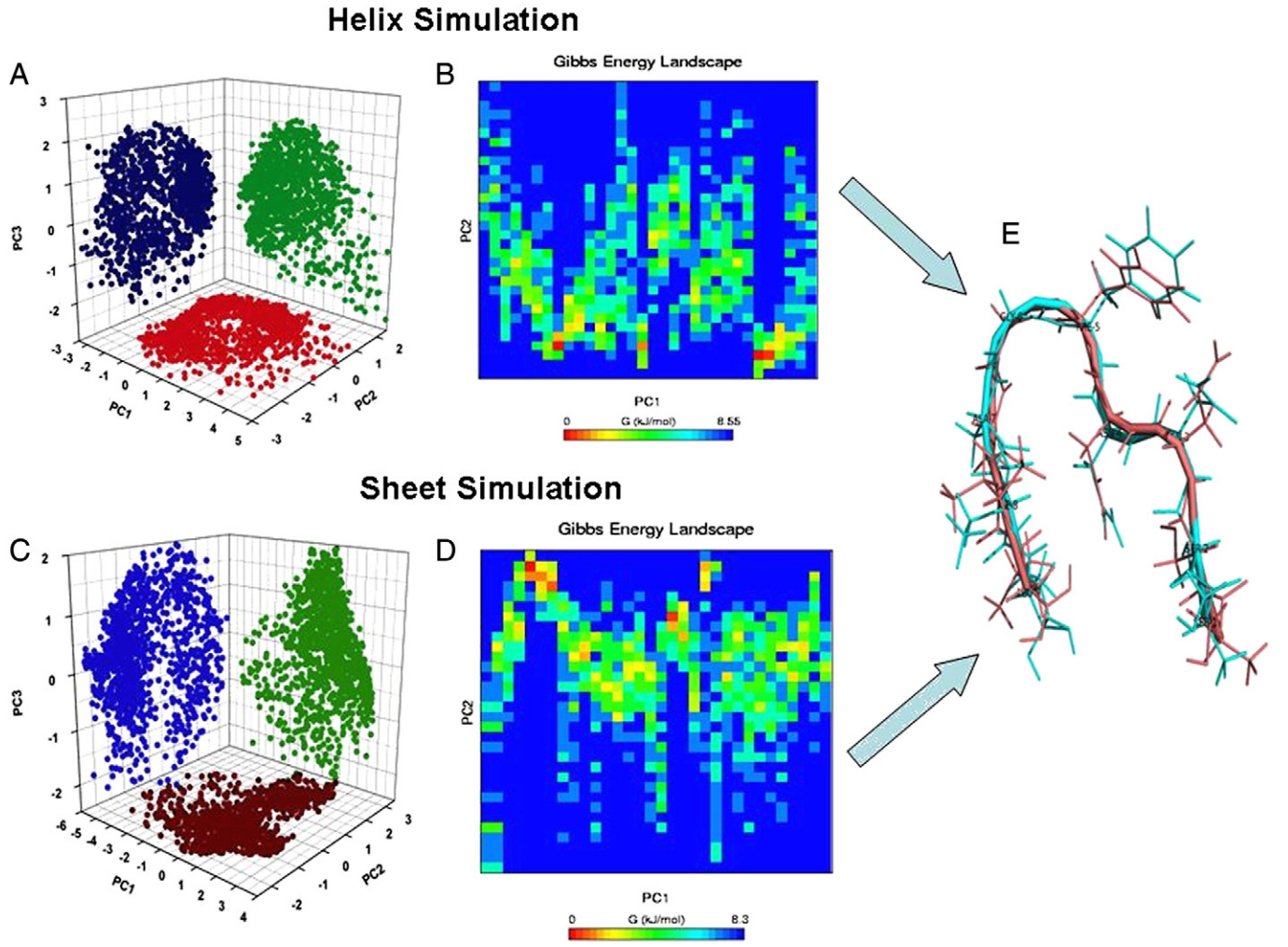


Fig. 6. 3-D projection of principal components (1, 2 and 3) on the concatenated trajectories of helix (A) and sheet (C) simulations and comparison of Free energy landscape of concatenated trajectories of helix (B) and sheet (D) simulations. E: 3-D structure of hIAPP 19–27 from the most populated cluster obtained from helix (cyan) and sheet (red) simulation.

adopts a similar bend structure in full length fibrillogenic hIAPP evident from isotope labelling and 2-D infrared spectroscopy [47]. Experimental data on the isolated and unmodified dimers of human (hIAPP) also suggests that this region adopts a wider bend like conformation linking two adjacent β-plate to form anti-parallel beta sheet [48]. IAPP being a natively unfolded protein exist in variety conformations [49,50]. Recent molecular dynamics and experimental studies reveals two main structural families of human IAPP monomer, a predominantly helix-coil conformation [50,32,45] and the second one is an extended β-hairpin conformation [50,48]. The extended conformation is topologically very similar to the fibril structure of the peptide. Intrinsically the region 22–25 of hIAPP adopts a bend/turn like structure and this structure is prevalent in both helix-coil and extended β-hairpin structure of hIAPP. Thus the region holds its structure during the initial pathogenic conversion of amylin to β-sheet structure in T2DM and might play a crucial role in these pathogenic folding events. This observation provides further credence to our prediction based on sequence analysis and AGGRESCAN that variation in hIAPP 19–27 region alters the aggregation propensity of the peptide and explain the species specific variation in amyloidogenecity.

4. Conclusions

Here, a combined approach of sequence analysis and MD simulation has been used to provide structural insight into the intramolecular β-sheet structure of amylin. Sequence analysis reveals specific substitutions at 17th, 22nd and 23rd positions to be very crucial for amyloid formation. This observation corroborates with the experimental facts that N22A and F23A mutation significantly influences the aggregation propensity of human amylin [51]. For rodents, presence of Proline at the 25th position solely dictates the observed non-amyloidogenecity. These detailed analyses of sequence variation and its effect on the aggregation propensity will be useful in designing various amylin variant that can be used as an anti-hyperglycaemic drug with lower aggregation property. We also propose that Val17Asp and Phe23Arg variant of human amylin can also be considered as amylinomimetic drug with reduced aggregation propensity. But further experimental investigations are needed to be carried out for confirmation.

0.24 μs MD simulation on hIAPP 19–27 fragment reveals this region to adopt a broad bend like conformation with a four residue turn spanning from residue 22 to 25. This result also can explain observed genetic variation found in Japanese and Chinese population where a Ser20Gly mutation increases T2DM significantly [52,53]. As evident from our simulated model, this region adopts a bend like conformation and S20G substitution stabilizes the bend conformation more compared to the native structure of the peptide and further stabilizes the intramolecular arrangement of the β-sheets of amylin, promoting fibril formation. Recent molecular dynamics study suggests that serine20glycine substitution drives hIAPP from helical to β-strands with higher extension and flexibility, which may promote the aggregation of hIAPP [45]. Based on

this study and pre-existing knowledge, we provide a structural insight into the amyloidogenic form of amylin. The proposed model can be summarized as: the monomer of this peptide in fibrillar form is believed to be composed of three β-strands [18]. The first one is a long β-strand from residues 8–20 followed by a bend (turn predicted from residues 22–25) and a second short β-strand (residues 25–28) which goes anti-parallel to the first β-strand. This region is followed by a tight turn (residues 29–31) and then again by a third β-strand from residue 32 to 37 which again goes antiparallel to the second β-strand. Sequence analysis reveals Proline29 to be conserved across most organisms with different amyloid forming propensity thus inferring the second β-strand to be shorter in comparison to the existing model. This intramolecular folding pattern of amylin converts the peptide from its native unfolded structure to the β-sheet conformation which might be the starting structure for fibril growth, the main pathological feature of type 2 diabetes. Our simulation results are consistent with the experimental observation and provide a detailed structural insight into the early amyloidogenic form of hIAPP.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bpc.2012.05.003.

References

[1] G. Merlini, V. Bellotti, Molecular mechanisms of amyloidosis, The New England Journal of Medicine 349 (2003) 583–596.
[2] C.M. Dobson, Protein misfolding, evolution and disease, Trends in Biochemical Sciences 24 (1999) 329–332.
[3] R.L. Nussbaum, C.E. Ellis, Alzheimer's disease and Parkinson's disease, The New England Journal of Medicine 348 (2003) 1356–1364.
[4] D. Thirumalai, D.K. Klimov, R.I. Dima, Emerging ideas on the molecular basis of protein and peptide aggregation, Current Opinion in Structural Biology 13 (2003) 146–159.
[5] C.P. Jaroniec, C.E. MacPhee, V.S. Bajaj, M.T. McMahon, C.M. Dobson, R.G. Griffin, High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy, Proceedings of the National Academy of Sciences of the United States of America 101 (2004) 711–716.
[6] R. Kayed, E. Head, J.L. Thompson, T.M. McIntire, S.C. Milton, C.W. Cotman, C.G. Glabe, Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis, Science 300 (2003) 486–489.
[7] G.M. Reaven, Banting lecture role of insulin resistance in human disease, Diabetes 37 (1988) 1595–1607.
[8] G.J.S. Cooper, A.C. Willis, A. Clark, R.C. Turner, R.B. Sim, K.B.M. Reid, Purification and characterization of a peptide from amyloid-rich pancreases of type 2 diabetic patients, Proceedings of the National Academy of Sciences of the United States of America 84 (1987) 8628–8632.
[9] A. Lorenzo, B. Razzoboni, G.C. Weir, B.A. Yankner, Pancreatic islet toxicity of amylin associated with type-2 diabetes mellitus, Nature 368 (1994) 756–760.
[10] A. Clark, M.R. Nilsson, Islet amyloid: a complication of islet dysfunction or an aetiological factor in Type 2 diabetes? Diabetologia 47 (2004) 157–169.
[11] A. Clark, G.J. Cooper, C.E. Lewis, J.F. Morris, A.C. Willis, K.B. Reid, R.C. Turner, Islet amyloid formed from diabetes associated peptide may be pathogenic in type-2 diabetes, Lancet 2 (1987) 231–234.
[12] T. Sanke, T. Hanabusa, Y. Nakano, C. Oki, K. Okai, S. Nishimura, M. Kondo, K. Nanjo, Plasma islet amyloid polypeptide (amylin) levels and their responses to oral glucose in type 2 (non-insulin-dependent) diabetic patients, Diabetologia 34 (1991) 129–132.
[13] R.A. Pittner, K. Albrandt, K. Beaumont, Molecular physiology of amylin, Journal of Cellular Biochemistry 55 (1994) 19–28.
[14] L. Haataja, T.C. Gurlo, J. Huang, P.C. Butler, Islet amyloid in type 2 diabetes and the toxic oligomer hypothesis, Endocrine Reviews 29 (2008) 303–316.
[15] M. Zaïdi, H.K. Datta, P.J. Bevis, S.J. Wimalawansa, I. MacIntyre, Amylin-amide: a new bone conserving peptide from pancreas, Experimental Physiology 75 (1990) 529–536.
[16] C. Howard, Insular amyloidosis and diabetes mellitus in *Macaca nigra*, Diabetes 27 (1978) 357–364.
[17] L.K. Phillips, M. Horowitz, Amylin, Current Opinion in Endocrinology, Diabetes, and Obesity 13 (2006) 191–198.
[18] E.T.A.S. Jaikaran, A. Clark, Islet amyloid and type 2 diabetes: from molecular misfolding to islet pathophysiology, Biochimica et Biophysica Acta 1537 (2001) 179–203.
[19] J.D. Harper, P.T. Lansbury Jr., Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins, Annual Review of Biochemistry 66 (1997) 385–407.
[20] J. Hardy, D.J. Selkoe, The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics, Science 297 (2002) 353–356.
[21] F. Ding, N.V. Dokholyan, S.V. Buldyrev, H.E. Stanley, E.I. Shakhnovich, Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism, Journal of Molecular Biology 324 (2002) 851–857.
[22] H.B. Jang, C.K. Hall, Y.Q. Zhou, Assembly and kinetic folding pathways of a tetrameric beta-sheet complex: molecular dynamics simulations on simplified off-lattice protein models, Biophysical Journal 86 (2004) 31–49.
[23] J. Gsponer, U. Haberthur, A. Caflisch, The role of side-chain interactions in the early steps of aggregation: molecular dynamics simulations of an amyloid-forming peptide from the yeast prion Sup35, Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 5154–5159.
[24] B. Urbanc, L. Cruz, S. Yun, S.V. Buldyrev, G. Bitan, D.B. Teplow, H.E. Stanley, In silico study of amyloid beta-protein folding and oligomerization, Proceedings of the National Academy of Sciences of the United States of America 101 (2004) 17345–17350.
[25] H.D. Nguyen, C.K. Hall, Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides, Proceedings of the National Academy of Sciences of the United States of America 101 (2004) 16180–16185.
[26] D. Zanuy, R. Nussinov, The sequence dependence of fiber organization. A comparative molecular dynamics study of the islet amyloid polypeptide segments 22–27 and 22–29, Journal of Molecular Biology 329 (2003) 565–584.
[27] W. Chun, L. Hongxing, Y. Duan, Formation of partially ordered oligomers of amyloidogenic hexapeptide (NFGAIL) in aqueous solution observed in molecular dynamics simulations, Biophysical Journal 87 (2004) 3000–3009.
[28] P.M. Harrison, H.S. Chan, S.B. Prusiner, F.E. Cohen, Conformational propagation with prion-like characteristics in a simple model of protein folding, Protein Science 10 (2001) 819–835.
[29] R.I. Dima, D. Thirumalai, Exploring protein aggregation and self-propagation using lattice models: phase diagram and kinetics, Protein Science 11 (2002) 1036–1049.
[30] T. Cellmer, D. Bratko, J.M. Prausnitz, H. Blanch, Protein-folding landscapes in multichain systems, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 11692–11697.
[31] A. Melquiond, J.C. Gelly, N. Mousseau, P. Derreumaux, Probing amyloid fibril formation of the NFGAIL peptide by computer simulations, Journal of Chemical Physics 126 (2007) 0651011–0651017.
[32] M.N. Andrews, R. Winter, Comparing the structural properties of human and rat islet amyloid polypepetide by MD computer simulations, Biophysical Chemistry 156 (2011) 43–50.
[33] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences, Bioinformatics 22 (2006) 1658–1659.
[34] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32 (2004) 1792–1797.
[35] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, Systematic Biology 59 (2010) 307–321.
[36] F. Chevenet, C. Brun, A.L. Banuls, B. Jacq, R. Chisten, TreeDyn: towards dynamic graphics and annotations for analyses of tree, BMC Bioinformatics 7 (2006) 439–447.
[37] O. Conchillo-Sole, N.S.D. Groot, F.X. Aviles, J. Vendrell, X. Daura, S. Ventura, AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides, BMC Bioinformatics 8 (8) (2007) 65–81.
[38] D.V.D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, Journal of Computational Chemistry 26 (2005) 1701–1718.
[39] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[40] E. Papaleo, P. Mereghetti, P. Fantucci, R. Grandori, L.D. Gioia, Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: the myoglobin case, Journal of Molecular Graphics and Modelling 27 (2009) 889–899.
[41] J.M. Borreguero, B. Urbanc, N.D. Lazo, S.V. Buldyrev, D.B. Teplow, H.E. Stanley, Folding events in the 21–30 region of amyloid β-protein (Aβ) studied in silico, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 6015–6020.
[42] L. Cruz, B. Urbanc, J.M. Borreguero, N.D. Lazo, D.B. Teplow, H.E. Stanley, Solvent and mutation effects on the nucleation of amyloid β-protein folding, Proceedings of the National Academy of Sciences of the United States of America 102 (2005) 18258–18263.
[43] P. Jiang, W. Xu, Y. Mu, Amyloidogenesis abolished by proline substitutions but enhanced by lipid binding, PLoS Computational Biology 5 (2009) e1000357.
[44] F.S. Legge, H. Treutlein, G.J. Howlett, I. Yarovsky, Molecular dynamics simulations of a fibrillogenic peptide derived from apolipoprotein C-II, Biophysical Chemistry 130 (2007) 102–113.
[45] M. Wang, J. Yang, J. Wang, X. Wang, Structural effects of L16Q, S20G, and L16Q-S20G mutations on hIAPP: a comparative molecular dynamics study, Chinese Journal of Chemistry 30 (2012) 241–248.
[46] C. Wu, H. Lei, Y. Duan, Formation of partially ordered oligomers of amyloidogenic hexapeptide (NFGAIL) in aqueous solution observed in molecular dynamics simulations, Biophysical Journal 87 (2004) 3000–3009.
[47] C.T. Middleton, P. Marek, P. Cao, C.C. Chiu, S. Singh, A.M. Woys, J.J.D. Pablo, D.P. Raleigh, M.T. Zanni, Two-dimensional infrared spectroscopy reveals the complex behaviour of an amyloid fibril inhibitor, Nature Chemistry 4 (2012) 355–360.
[48] N.F. Dupuis, C. Wu, J.-E. Shea, M.T. Bowers, The amyloid formation mechanism in human IAPP: dimers have β-strand monomer–monomer interfaces, Journal of the American Chemical Society 133 (2011) 7240–7243.
[49] R.D. Murphy, J. Conlon, T. Mansoor, S. Luca, S.M. Vaiana, N.V. Buchete, Conformational dynamics of human IAPP monomers, Biophysical Chemistry (2012), http://dx.doi.org/10.1016/j.bpc.2012.03.010.
[50] N.F. Dupuis, C. Wu, J.E. Shea, M.T. Bowers, Human islet amyloid polypeptide monomers form ordered β-hairpins: a possible direct amyloidogenic precursor, Journal of the American Chemical Society 131 (2009) 18283–18292.

[51] D. Zanuy, Y. Porat, E. Gazit, R. Nussinov, Peptide sequence and amyloid formation: molecular simulations and experimental study of a human islet amyloid polypeptide fragment and its analogs, Structure 12 (2004) 439–455.
[52] S. Sakagashira, T. Sanke, T. Hanabusa, H. Shimomura, S. Ohagi, K.Y. Kumagaye, K. Nakajima, K. Nanjo, Missense mutation of amylin gene (S20G) in Japanese NIDDM patients, Diabetes 45 (1996) 1279–1281.
[53] S.C. Lee, Y. Hashim, J.K.Y. Li, G.T.C. Ko, J.A.J.H. Critchley, C.S. Cockram, J.C.N. Chan, The islet amyloid polypeptide (amylin) gene S20G mutation in Chinese subjects: evidence for associations with type 2 diabetes and cholesterol levels, Clinical Endocrinology 54 (2001) 541–546.