Molecular Modelling of Peptide Folding, Misfolding and Aggregation Phenomena

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Nevena Todorova


School of Applied Sciences
College of Science, Engineering and Health
RMIT University
August 2009
Declaration of Candidature

I certify that except where due acknowledgment has been made, the work is that of the candidate alone. This body of work has not been submitted previously, in whole or in part, to qualify for any other academic award. The content of this thesis is the results of work which has been carried out since the official commencement date of the approved research program. Any editorial work, paid or unpaid, carried out by a thirst part is acknowledged.

____________________
Nevena Todorova
Acknowledgements

This project would not have existed without the contribution, support and guidance of many people who deserve to be acknowledged.

I wish to thank my supervisor, Prof. Irene Yarovski, for giving me the opportunity to work on this project. Her invaluable guidance and assistance throughout these years have made this research possible and also very enjoyable.

Дорогая Ирина,
Я никогда не забуду наши разговоры, смех, танцы и упражнения, это было замечательное время. Вы были не просто моим учителем, но и хорошим, близким другом. Спасибо за всё. С любовью, Невена.

I would also like to say a special thank you to my other supervisors, Dr. Sue Legge and Dr. Andrew Hung for their support, guidance and help during my research project. Your comments and suggestions that have made this thesis be a better piece of work.

Sincere thank you to Dr. Akin Budi (a.k.a Akinski) for accepting to read this thesis to the smallest detail as humanly possible. The hilarious comic printouts have made me look forward to your comments, thanks! :)

I would also like to acknowledge, Dr. Stefano Piana (Curtin University, Perth) and Fabrizio Marinelli (SISSA, Trieste) for their guidance and contribution to the work on the folding of insulin.

Geoff Howlett and group from the Bio21 Institute, Melbourne, thank you for the collaboration and experimental support you have provided to this project.

I would like to thank Dr. Herbert Treutlein (Cytopia Research Pty Ltd) for his contribution to the force field work.

Acknowledgments to the Australian Research Council and Cytopia Research Pty Ltd. for providing financial support on this project.

For providing all the computational facilities, I would like to acknowledge the Australian Partnership for Advanced Computing (APAC) and the Victorian Partnership for Advance Computing (VPAC).

To everyone in the Molecular Modelling and Simulations Group, thank you for making this project be a very enjoyable experience.
On a personal note, I would like to thank my dear friends for their support, understanding and encouragements over the past few years, with a special thank you to Zana for always being an e-mail or phone call away. Starbucks will never be the same!

I would like to say a special thank you to my partner Paul, for his love, support and understanding through these years, especially for his resilience over the last few months.

A very special thank you goes to my beautiful Sparky, you have brought so much joy and happiness to my life, and you are always there to cheer me up.

I would also like to thank my dearest grandma, who might be far away but never far from my heart. Your picture at my desk has given me strength to never give up.

And last but definitely not the least, I would like to thank and dedicate this thesis to my parents, for their love, understanding and never-ending support they always provide me. You are by backbone, my everything. I love and thank you greatly from the bottom of my heart. Бе сакам премногу, Невена.


Publication List

Peer-reviewed Publications


“Computational investigation on the effects of mutation of the amyloidogenic apoC-II(60-70) monomers” (2009), in preparation.

“Lipid concentration effects of the amyloidogenic apoC-II(60-70) peptide: Computational study” (2009), in preparation.
Conference Presentations

[1] **N. Todorova**
“Computational studies of peptide folding, misfolding and aggregation phenomena.”
The 12th Australian Molecular Modelling Conference, Gold Coast, Australia, 2009.

“Molecular modelling studies of peptide folding, misfolding and aggregation phenomena”
The 8th Melbourne Protein Group Symposium, Melbourne, Australia, 2009.

“Computational study on the effect of different lipid concentrations on the amyloidogenic peptide derived from apolipoprotein C-II.”

“Exploring the folding free energy landscape of insulin”
Biomolecular Interactions and Dynamics Symposium, Melbourne, 2009.
*Award received for best Poster Presentation.*

“Simulation study on the effect on stability of oxidation and mutation of amyloidogenic peptides derived from apolipoprotein C-II. Part 2: apoC-II (60-70)”

“BE-META: A new method for protein folding prediction, application to insulin”

[7] **N. Todorova, F. Marinelli, S. Piana, I. Yarovsky**
“Exploring the folding free energy landscape of insulin”
*Award received for best Poster Presentation.*

“Computational methods for studying proteins”
Workshop: Interaction Energies and the Structure of Surfaces and Nano-Structures, RMIT, Melbourne, Australia, 2007

“Different force fields, one protein, do they give the same answer?”
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D-IR</td>
<td>Two-Dimensional Infrared Spectroscopy</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic Force Microscopy</td>
</tr>
<tr>
<td>APAC</td>
<td>Australian Partnership for Advanced Computing</td>
</tr>
<tr>
<td>ApoC-II</td>
<td>Apolipoprotein C-II</td>
</tr>
<tr>
<td>BE-META</td>
<td>Bias Exchange Metadynamics</td>
</tr>
<tr>
<td>CD</td>
<td>Circular Dichroism</td>
</tr>
<tr>
<td>D5PC</td>
<td>Dipentanoylphosphatidylcholine</td>
</tr>
<tr>
<td>DMD</td>
<td>Discrete Molecular Dynamics</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DPI</td>
<td>Despentapeptide insulin</td>
</tr>
<tr>
<td>FRET</td>
<td>Fluorescence Resonance Energy Transfer</td>
</tr>
<tr>
<td>GROMACS</td>
<td>GROningen MAchine for Chemical Simulation</td>
</tr>
<tr>
<td>KMC</td>
<td>Kinetic Monte Carlo</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>MCL</td>
<td>Markov Cluster analysis method</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>NAMD</td>
<td>Not Another Molecular Dynamics</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>NOESY</td>
<td>Nuclear Overhauser Effect Spectroscopy</td>
</tr>
<tr>
<td>NPT</td>
<td>Constants number of particles, pressure and temperature</td>
</tr>
</tbody>
</table>
NVT ........................................ Constant number of particles, volume and temperature
PBC ................................................................. Periodic boundary condition
PDB ............................................................... Protein Data Bank
RDF ................................................................. Radial Distribution Function
PEPCAT ......................................................... Peptide Conformational Analysis Tool
REMD ............................................................ Replica Exchange Molecular Dynamics
RMSD ............................................................. Root Mean Square Displacement
RESP ............................................................. Restrainted Electostatic Potential
SASA .............................................................. Solvent Accessible Surface Area
TFE ................................................................. Trifluoroethanol
ThT ................................................................. Theoflavin T
VMD ............................................................... Visual Molecular Dynamics
VPAC ............................................................. Victorian Partnership for Advanced Computing
WHAM .......................................................... Weighted Histogram Analysis Method
XRD ............................................................... X-ray Diffraction
Table of Contents

Declaration of Candidature .................................................................................................... II
Acknowledgements ................................................................................................................ III
Publication List ........................................................................................................................ V
Abbreviations ........................................................................................................................ VII
Table of Contents ................................................................................................................... IX
List of Tables ........................................................................................................................ X
List of Figures ........................................................................................................................ XI
Abstract ................................................................................................................................ 1

1. Computational studies of protein folding and aggregation .............................................. 3
   1.1 Overview ....................................................................................................................... 3
   1.2 Protein structure .......................................................................................................... 4
   1.3 Protein folding .............................................................................................................. 11
       1.3.1 Studies of protein folding mechanisms ............................................................... 11
       1.3.2 Protein folding kinetics ....................................................................................... 15
       1.3.3 Transition / intermediate states in protein folding .............................................. 16
       1.3.4 Unfolded or denatured states ............................................................................. 18
   1.4 Protein self-association and aggregation .................................................................. 20
       1.4.1 Early history of amyloid diseases ......................................................................... 20
       1.4.2 Protein aggregation related diseases .................................................................. 21
       1.4.3 Structure of amyloid fibrils ................................................................................ 23
       1.4.4 Mechanisms of fibril formation .......................................................................... 24
       1.4.5 Sequence and mutation effects on fibril formations ............................................ 28
       1.4.6 Lipid effects on fibril formation ........................................................................... 29
       1.4.7 Oligomer stability ............................................................................................... 30

2. Computational techniques for protein studies ................................................................. 32
   2.1 Overview ....................................................................................................................... 32
   2.2 Molecular modelling .................................................................................................... 34
2.3 Molecular mechanics: Empirical force field models ..............................................................36
  2.3.1 Potential energy function .............................................................................................36
  2.3.2 Force field parameterisation ......................................................................................39
  2.3.3 Treatment of solvation ..............................................................................................39
  2.3.4 Force field validation studies ....................................................................................40
2.4 Molecular dynamics: Theory and application to protein simulations ..............................41
  2.4.1 Equations of motion ..................................................................................................42
  2.4.2 Periodic boundary conditions ..................................................................................43
  2.4.3 Neighbour search and nonbonded interactions ......................................................45
  2.4.4 Thermodynamic ensembles ......................................................................................46
  2.4.5 Temperature coupling ..............................................................................................46
  2.4.6 Pressure coupling .....................................................................................................48
  2.4.7 Bond constraint algorithms ......................................................................................50
2.5 Conformational sampling problem ..................................................................................51
2.6 Umbrella sampling .........................................................................................................52
2.7 Replica exchange molecular dynamics ...........................................................................53
2.8 Metadynamics ................................................................................................................54
2.9 Bias-exchange metadynamics ........................................................................................55
2.10 Simulation procedures ..................................................................................................57
3. Proteins investigated: Insulin & ApoC-II ...........................................................................59
  3.1 Overview .......................................................................................................................59
  3.2 Insulin ..........................................................................................................................60
    3.2.1 Insulin structure ......................................................................................................60
    3.2.2 Insulin folding and aggregation studies ..................................................................64
  3.3 Apolipoprotein C-II .....................................................................................................67
    3.3.1 ApoC-II structure ...................................................................................................68
    3.3.2 ApoC-II misfolding and aggregation studies .........................................................70
  3.4 Project aims ...................................................................................................................73
4. Systematic comparison of empirical force fields for molecular dynamic simulation of insulin ...........................................................................................................74
  4.1 Overview .......................................................................................................................74
  4.2 Introduction ...................................................................................................................75
  4.3 Computational Details ..................................................................................................77
  4.4 Results and Discussions ...............................................................................................79
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.1</td>
<td>Secondary structure evolution</td>
<td>79</td>
</tr>
<tr>
<td>4.4.2</td>
<td>NOE energies and violations</td>
<td>84</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Root mean square deviation (RMSD)</td>
<td>91</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Radius of gyration</td>
<td>93</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Solvent accessible surface area (SASA)</td>
<td>94</td>
</tr>
<tr>
<td>4.4.6</td>
<td>Cross-force field simulations</td>
<td>95</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions</td>
<td>96</td>
</tr>
<tr>
<td>5.</td>
<td>Exploring the folding free energy landscape of insulin</td>
<td>98</td>
</tr>
<tr>
<td>5.1</td>
<td>Overview</td>
<td>98</td>
</tr>
<tr>
<td>5.2</td>
<td>Introduction</td>
<td>99</td>
</tr>
<tr>
<td>5.3</td>
<td>Computational Details</td>
<td>100</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Collective variables</td>
<td>102</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Analytical method for biased statistics: Kinetic model construction</td>
<td>103</td>
</tr>
<tr>
<td>5.4</td>
<td>Results and Discussions</td>
<td>104</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Neutral replica analysis</td>
<td>105</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Kinetic model construction</td>
<td>108</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Structural analysis of the folded state basin</td>
<td>114</td>
</tr>
<tr>
<td>5.4.4</td>
<td>Folding pathway of chain B of insulin</td>
<td>117</td>
</tr>
<tr>
<td>5.5</td>
<td>Conclusions</td>
<td>119</td>
</tr>
<tr>
<td>6.</td>
<td>Investigation on the effects of mutation of the amyloidogenic apoC-II(60-70) peptide</td>
<td>121</td>
</tr>
<tr>
<td>6.1</td>
<td>Overview</td>
<td>121</td>
</tr>
<tr>
<td>6.2</td>
<td>Introduction</td>
<td>122</td>
</tr>
<tr>
<td>6.3</td>
<td>Computational Details</td>
<td>125</td>
</tr>
<tr>
<td>6.4</td>
<td>Results and Discussions</td>
<td>127</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Structural evolution analysis</td>
<td>127</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Cluster analysis</td>
<td>130</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Root mean square deviation</td>
<td>132</td>
</tr>
<tr>
<td>6.4.4</td>
<td>Radius of gyration</td>
<td>133</td>
</tr>
<tr>
<td>6.4.5</td>
<td>Side-chain interactions</td>
<td>135</td>
</tr>
<tr>
<td>6.4.6</td>
<td>Aromatic side-chain orientation</td>
<td>137</td>
</tr>
<tr>
<td>6.4.7</td>
<td>Solvation properties</td>
<td>139</td>
</tr>
<tr>
<td>6.5</td>
<td>Conclusions</td>
<td>142</td>
</tr>
</tbody>
</table>
7. **Lipid concentration effects on the amyloidogenic apoC-II(60-70) peptide** ..........144

7.1 Overview ..................................................................................................................144
7.2 Introduction ...............................................................................................................145
7.3 Computational Details .............................................................................................147
7.4 Results and Discussions ..........................................................................................150
  7.4.1 Molecular dynamics: Lipid-peptide interactions .............................................150
  7.4.2 Molecular dynamics: Secondary structure evolution .....................................162
  7.4.3 Umbrella sampling: Free energy profiles ......................................................166
  7.4.4 Umbrella sampling: Lipid-peptide interactions .............................................167
7.5 Conclusions ..............................................................................................................170

8. **Oligomers of apoC-II(60-70): stability & effect of mutation** ....................173

8.1 Overview ..................................................................................................................173
8.2 Introduction ...............................................................................................................174
8.3 Computational Details .............................................................................................175
8.4 Results and Discussions ..........................................................................................178
  8.4.1 Root mean square deviation and clustering analysis of wild-type apoC-II(60-
    70) oligomers ...........................................................................................................178
  8.4.2 Root mean square deviations of mutated apoC-II(60-70) oligomers ..........185
  8.4.3 Free energy profiles of wild-type and mutated apoC-II(60-70) dimers .....189
8.5 Conclusions ..............................................................................................................191

9. **Future work** ..........................................................................................................192

9.1 Computational techniques and force field development ........................................192
9.2 Theoretical investigation of fibril microcrystal structures ....................................193
9.3 Lipids and apolipoprotein interactions in heart disease .......................................193
9.4 Protein self-assembly on surfaces, interfaces and nanoparticles ..........................195

Appendix .........................................................................................................................197

References .......................................................................................................................208
List of Tables

Table 1.1 A list of diseases associated with the aggregation of proteins, adopted from the review article by Chiti and Dobson [101]. The name of the disease, the pathological protein, together with the size and native structure are specified.................................................................23

Table 3.1 Primary sequence of chain A and B of porcine insulin (PDB code 1ZNI). .................61
Table 3.2 Specific sequence differences between human, porcine and bovine insulin.........62
Table 3.3 Primary sequence of apoC-II (PDB code 1SOH). The sequence coloured in grey represents the residues with undefined structure...........................................69

Table 4.1 Average NOE interproton distance violations calculated for each force field simulation. The total average number of violations, as well as the severe violations larger than 1 Å from the upper bound value of the classified cross-peak intensities by Hawkins et al. [232] are indicated.................................86

Table 4.2 Average radius of gyration (Rg) and its standard deviation, calculated for each simulation. .................................................................................................................93

Table 4.3 Average solvent accessible surface area (SASA) and standard deviations calculated for each simulation. ..............................................................................94

Table 5.1 Thermodynamic properties of the basins found using the MCL algorithm with \(p = 1.08\). All thermodynamic properties were calculated with respect to the native-like basins. ...........................................................................................................111

Table 7.1 The nomenclature for the individual system, as well as their simulated times using: a) Molecular dynamics and b) Umbrella sampling method. The two digits in the label describe the ratio of lipid to peptide, and the letters \(c\) and \(h\) describe coil and helix starting structure, respectively........................................150

Table 8.1 System set-up listing the names given for each simulation, together with the simulation box dimensions, number of water molecules, counterions and total simulation time. ...........................................................................................177
List of Figures

Figure 1.1 The general structure of an amino acid, with the *amino* group on the left, the *carboxyl* group on the right and the unique *side-chain* (R) bonded to the Cα atom. ........................................................................................................................4

Figure 1.2 The condensation of two amino acids to form a peptide bond. ..............................................5

Figure 1.3 The amino acid sequence of a protein is called its primary structure. Different regions of the sequence form local regular secondary structures, such as α-helices or β-strands. The tertiary structure is formed by packing such structural elements into a globular unit or domain. The final conformation of a protein may contain several polypeptide chains arranged in a quaternary structure. ..................................................................................................................7

Figure 1.4 Ramachandran diagram showing the three sterically allowed regions for common secondary structures of a protein. Lighter (yellow/green) colours represent the more favourable combinations of ψ-φ angle values. .........................8

Figure 1.5 Illustration of the hydrogen bonding patterns represented by dotted red lines in a) parallel; and b) anti-parallel β-sheets. .................................................................9

Figure 1.6 The β-hairpin structure; a) Inter-strand hydrogen bonding shown as black dotted lines (backbone shown only); b) Cartoon representation of β-hairpin. ......10

Figure 1.7 Cartoon representation of the different energy landscapes from a denatured conformation to the native conformation N, adopted from reference [16]. (a) A smooth energy landscape for fast folding protein; (b) A moat landscape, where folding must go through an obligatory intermediate; (c) A rugged energy landscape with kinetic traps; and (d) A golf course energy landscape in which folding is dominated by diffusional conformational search. ..................13

Figure 1.8 Illustration of the free energy profile of a simple protein folding event, showing the unfolded (U), transition state (TS) and folded or native structure (N). ...........................................................................................................17

Figure 1.9 Experimentally determined crystal structures of Aβ1-42 and Sup35 derived GNNQQNY peptides.................................................................24

Figure 1.10 A schematic representation of some of the many conformational states adopted by polypeptide chains during folding, misfolding and aggregation.
The figure was reproduced from the detailed review presented by Chiti and Dobson [101].

Figure 2.1 Diagram of the accessible system size and time scale for different types of simulation studies.

Figure 2.2 A two-dimensional illustration of periodic boundary conditions.

Figure 2.3 Three-dimensional representation of a typical simulation box with solvent and protein model shown explicitly.

Figure 2.4 The real distribution of a system along a reaction coordinate “Distance” can be calculated using several umbrella sampling windows (red); after correcting for the bias potentials the free energy along the selected reaction coordinate is identified (purple).

Figure 2.5 In replica exchange molecular dynamics simulation several replica of the same system are simulated in parallel at different temperatures. Exchanges between neighbouring replicas are performed at selected times.

Figure 3.1 Representation from different angles of the porcine insulin monomer. Chain A is coloured blue, chain B is coloured red and the disulfide bonds between the Cys residues are coloured yellow.

Figure 3.2 An overlay of X-ray structures, showing experimentally observed states of insulin chain B.

Figure 3.3 The microcrystal structure of fibril forming peptide segment, VEALYL of insulin chain B (PDB code 2OMQ); a) A two-sheet motif of the smallest unit-cell of the VEALYL fibril structure, depicted in projection down the needle crystal axis; b) Cartoon representation of the crystalline seed; c) Longitudinal view of two-sheet fibril formation extending from the core seed structure represented in colour.

Figure 3.4 Ribbon representation of apoC-II(13-79) showing secondary structure and side-chains.

Figure 3.5 Amyloidogenic peptide ApoC-II(60-70).

Figure 4.1 Secondary structure evolution of insulin chain B. Simulations for each force field are labelled as described in the *Computational Details* section of this chapter. The secondary structure colour codes: magenta – α-helix, red – π-helix, cyan – turn, white – coil, yellow – extended conformation and green – bridge.
Figure 4.2 The reference structure of the two most populated clusters of each force field simulation, labelled as described in the Computational Details section of this chapter. The colouring of each protein is by its residue type, blue – basic, red – acidic, green – polar and silver – nonpolar. The hydrogen bonds of structure d2 are coloured purple.

Figure 4.3 NOE energy derived from the distance restraints of the sampled structures. The region below the line at $\sim 41.5 \times 10^3$ a.u. represents the MD structures with NOE energy less than that of the starting structure. The plot shows a moving average of 250 ps.

Figure 4.4 Important protein conformations and their NOE distance violations. The satisfied distance restraints are shown in yellow and the violated distances in black. The colour scheme is the same as in Figure 4.1.

Figure 4.5 RMSD of the helical region of chain B, from residues Ser9 to Cys19, compared to the starting structure as a function of time for all systems. The plot shows a moving average of 250 ps.

Figure 4.6 Secondary structure evolution of insulin chain B; a) Starting structure taken from the last frame of simulation using GROMOSA6 force field was simulated using AMBER03; b) Starting structure taken from the last frame of simulation using CHARMM force field was simulated using OPLS.

Figure 5.1 The starting structure selected for this study on the folding mechanisms of insulin chain B.

Figure 5.2 The free energy of the 20 most populated clusters. The native-like cluster is shown as an inset of the figure. The structures showing the most typical features are represented on the right of the graph. The clusters are numbered in the order of free energy with ascending cluster ID.

Figure 5.3 Free energy profiles calculated from the BE-META simulations of chain B of insulin. The position of the lowest energy state of the neutral replica is marked by a blue square on each free energy curve. A typical lowest energy structure is also shown.

Figure 5.4 Representative low free energy cluster of the three basins determined using $p = 1.08$. Molten-globule 1 Most stable cluster of the basin, where all charged residues interact with the N- and C-termini of the protein, giving rise to strong electrostatic contacts.
Figure 5.5 Molten-globule 2  a) Most stable cluster of the basin, where Phe25 (green) is burried in a hydrophobic pocket, while Phe24 (yellow) is partially exposed to the solvent. b) This structure is only 0.9 kcal/mol from the most stable cluster and shows the presence of a hydrophobic core in which Phe24 (yellow) is present. Phe25 (green) is partially solvent exposed in this basin. c) In this cluster the Tyr16 (orange) residue and partially Phe24 (yellow) are buried in a hydrophobic pocket. The hydrophilic part of Tyr16 is directed toward the Arg11, Gln21, and the terminal salt bridges........................................110

Figure 5.6 Native-like a) Lowest free energy cluster in the folded state basin. b) Cluster present between the native-like and molten-globule 2 basin having the N-terminal α-helix unfolded. ...........................................................................................110

Figure 5.7 Occupancy (%) of the three basins found for $p = 1.08$ as a function of temperature. A simple linear extrapolation [65] from the calculated cluster enthalpy and entropy at 300 K has been used to obtain the basin occupancy versus temperature........................................................................................................112

Figure 5.8 Metastable sets (cluster basins) detected by MCL using $p = 1.12$, are shown as coloured contours, with the coloured sphere corresponding to the lowest free energy cluster of each basin. The respective structures are presented above with the same colour code and ordered based on their free energy (1 has the lowest, 8 has the highest free energy). .....................................................113

Figure 5.9 The lowest energy structure (blue) from the folded state basin superimposed with the X-ray crystallographic structure of chain B, PDB code 1ZNI (pink)....116

Figure 5.10 Schematic representation of the basin dynamics. Transitions between the basins are represented by an arrow along with the corresponding transition time. Transitions that occur on a time scale longer than 1 µs are represented as dashed arrows.................................................................118

Figure 6.1 Structure representation of apoC-II(60-70) monomeric peptide showing mutated residues and simulation times.................................................................126

Figure 6.2 Secondary structure evolution plot for each mutation of apoC-II(60-70) peptide; a) Wild-type; b) oxi-Met; c) Met60Val; d) Met60Gln. The secondary structure colour codes: magenta – α helix, red – π helix, cyan – turn, white – coil, yellow – extended conformation and green – hydrogen bridge...............................................................129
Figure 6.3 The number of clusters identified using an RMSD cutoff of 0.15 nm for each individual simulation is shown in the left of the figure. Structural representation of the two most populated clusters and their population are presented on the right.................................................................131

Figure 6.4 Radius of gyration (nm) of each simulation, represented by a moving average of 5 ns; a) Wild-type; b) oxi-Met; c) Met60Val; d) Met60Gln. .................................................134

Figure 6.5 Contact maps representing the average contact distance between all atoms of each residue for the structures contained in the most populated cluster of apoC-II(60-70) peptide; a) Wild-type; b) oxi-Met; c) Met60Val; and d) Met60Gln. A contact is defined by a distance of < 0.5 nm. .................................................136

Figure 6.6 Histograms of Tyr63 ring orientation with respect to Phe67 (x-axis), obtained from total simulated trajectories for oxidised Methionine [285] (red), Met60Val (green), Met60Gln (blue) and wild-type (black) peptides. Y-axis indicates the percentage of structures for a specific orientation angle range exhibited during the trajectory. Forty bins were used over the range 0 – 180°. Each data point represents the percentage (%) of structures with orientation angles between x and x + 4.5°. Angle < 90° indicates both rings are on one side of the hairpin, while > 90° indicates rings exist on opposite sides. Structures illustrating the relative ring orientations are represented in insets.................................................................138

Figure 6.7 Average values calculated per residue over the structures contained in the most populated cluster of the oxidised Methionine (red), Met60Val (green), Met60Gln (blue) and wild-type (black) simulation. a) Solvent accessible surface area (nm²); b) The average number of water molecules within a 3 Å shell of each residue.........................................................................................141

Figure 7.1 Structural representation of D5PC lipid molecule, where the carbon atoms are represented in cyan, hydrogen in red and the phosphate in green. The apoC-II(60-70) starting conformations, coil and helix are also shown together with the residue sequence of the peptide...............................................148

Figure 7.2 Typical radial distribution functions of the lipid head/s and tail/s with respect to hydrophobic and hydrophilic residues of the apoC-II(60-70) peptide. The RDFs for the 4lh system are shown as an example, while for all the other systems the RDF plots are presented in the Appendix of the thesis..........................151
Figure 7.3 Schematic of the micellar-like formation of lipids surrounding the peptide. The hydrophobic residues are represented in blue, hydrophilic in red and the lipid in black. ..........................................................152

Figure 7.4 Histograms representing the contacts formed between the lipid head and tail with individual residues of apoC-II(60-70) for systems 1lc and 6lh. Percentages of the times in contact are shown in the red/blue histograms, while the percentages of total number of contacts are shown in the grey/green histograms of each system. ..........................................................154

Figure 7.5 Average solvent accessible surface area (SASA) per residue and corresponding error ........................................................................................................156

Figure 7.6 Histogram representation of the % of lipids that were in contact with the peptide for every simulation. The bars of the histogram plots are coloured as described in the Computational Details section of this chapter. .........................159

Figure 7.7 Water radial distribution functions (RDF) with respect to apoC-II(60-70) peptide at various lipid concentrations; a) Water RDF profiles of the simulations starting from the coil conformation; b) Water RDF profiles of the simulations starting from the helix conformation. The region of the first RDF peak is shown as an inset. The RDF profiles are coloured as described in the Computational Details of this chapter. ........................................161

Figure 7.8 Structural models of several lipid-peptide complexes. The surface of the peptide is coloured yellow, whereas the surface of the lipids is coloured red. .164

Figure 7.9 Secondary structure evolution of apoC-II(60-70) at various lipid concentrations. Each simulation is labelled as described in the Computational Details section of this chapter ........................................165

Figure 7.10 Potential of mean force (PMF) profiles as a function of centre of mass separation (COM) of the N- and C-termini of apoC-II(60-70) peptide at various lipid concentrations. Magnified representations of selected regions are shown as an inset of the figure. Each PMF profile is coloured and labelled as described in the Computational Details section of this chapter. ......166

Figure 7.11 Average solvent accessible surface area (SASA) per residue. The calculations were performed on the data collected using umbrella sampling. Each plot is labelled as described in the Computational Details section of this chapter. ........................................................................................................169
Figure 7.12 Water radial distribution functions (RDF) with respect to apoC-II(60-70) peptide at various lipid concentrations calculated from the simulations performed using umbrella sampling at the lowest free energy region. Magnified representation of the region of the first peak is shown as inset. The RDF profiles are coloured as described in the Computational Details of this chapter.............................................................................................................170

Figure 8.1 Illustration of the starting arrangement and size of apoC-II(60-70) oligomeric composites. ......................................................................................................................................................176

Figure 8.2 Root mean square deviation (RMSD) plots of the trajectories of dimeric apoC-II(60-70) oligomers. The structures of the most populated clusters of every system and their population are also presented. ........................................179

Figure 8.3 Contact maps representing the average contact distances between the backbone and side-chain atoms between each residue of each dimeric system of apoC-II(60-70). The residues of monomer 1 are shown on the horizontal axis, whereas the residues of monomer 2 are shown on the vertical axis........................................................................................................................................180

Figure 8.4 Root mean square deviation (RMSD) plots of the trajectories of trimeric apoC-II(60-70) oligomers. The structures of the most populated clusters of every system and their population are also presented. ........................................182

Figure 8.5 Root mean square deviation (RMSD) plots of the trajectories of tetrameric apoC-II(60-70) oligomers. The structures of the most populated clusters of every system and their population are also presented. ........................................184

Figure 8.6 Root mean square deviation (RMSD) plots of wild-type and mutated, four stranded anti-parallel oligomer of apoC-II(60-70). .........................................................187

Figure 8.7 Structures representing the typical conformations observed during the simulations of mutated tetrameric apoC-II(60-70) oligomers; a) 1-4-APC-oxiMet (red), b) 1-4-APC-M60V (green) and c) 1-4-APC-M60Q (blue). In some structures the aromatic side-chains are also shown, as well as the surface of the hydrophobic residues is represented as a green mesh..............188

Figure 8.8 Free energy profiles as a function of the center-of-mass (COM) separation of the wild-type and mutated (oxiMet60, Met60Val and Met60Gln) apoC-II(60-70) dimers with a) Charged and b) Neutral termini. .........................................................190
Abstract

In this thesis computer modelling studies were conducted to investigate protein behavior in various environments causing their folding, unfolding and aggregation. An introduction to the principles of protein structure and function, along with a concise literature review on some of the latest discoveries in the area of protein folding and aggregation are presented in Chapter 1.

Classical Molecular Dynamics techniques and their derivative methods such as umbrella sampling and bias-exchange metadynamics (BE-META) were employed and are described in Chapter 2. Applications related to two important proteins – insulin and apolipoprotein C-II (ApoC-II) are presented. The current knowledge of the structure and behaviour of these proteins is discussed in Chapter 3.

The use of atomistic simulation methodologies based on empirical force fields has enhanced our understanding of many physical processes governing protein structure and dynamics. However, the force fields used in classical modelling studies are often designed for a particular class of proteins and rely on continuous improvement and validation by comparison of simulations with experimental data. In Chapter 4 a comprehensive comparison of five popular force fields for simulation of insulin is presented. The effect of each force field on the conformational evolution and structural properties of the protein is analysed in detail and compared with available experimental data.

A fundamental phenomenon in nature is the ability of proteins to fold \textit{ab initio} to their functional native conformation, also known as their biologically active state. Due to the heterogeneity and dimensionality of the systems involved, it is necessary to employ methodologies capable of accelerating rare events, specifically, configurational changes that involve the crossing of large free energy barriers. In Chapter 5, using the recently developed method BE-META the structural transitions and possible folding pathways of insulin were identified.

Another interesting phenomenon is the misfolding of proteins causing their aggregation, that may lead to formation of either amorphous compounds or structures of elongated-unbranched morphology known as amyloid fibrils. The deposition of amyloid fibrils in the human body may cause many debilitating diseases such as Alzheimer’s and
variant Creutzfeldt-Jakob diseases, thus making this field of research important and urgent. Due to the insoluble and non-crystalline nature of amyloid fibrils, experimental techniques are unable to elucidate the molecular mechanisms of fibril formation, in particular the initial stages of self-association. Thus computational methods are suitable for the investigation of these early fibril forming events and can give atomistic details of the initial peptide aggregation mechanisms. The human plasma protein apoC-II serves important roles in lipid transport, and it has been shown to form amyloid-like aggregates in solution. Recently, it has been demonstrated experimentally that oxidation of Met60 in the region of apoC-II(60-70) results in inhibition of fibril formation. Computational studies were performed to investigate the effect of mutations, such as Met oxidation and the residue substitutions to hydrophobic Val and hydrophilic Gln, on dynamics of apoC-II(60-70) peptide. The conformation features relevant to the amyloidogentic propensities of the peptide were identified and presented in Chapter 6.

The involvement of lipids at the various stages of development of amyloid diseases is becoming more evident in recent research efforts. In particular, micellar and sub-micellar concentrations have been shown to have different effect on fibril growth and kinetics of native apoC-II and derived peptides. In Chapter 7, investigation on the influences of phospholipids at various concentrations on the structure of apoC-II(60-70) using MD and umbrella sampling methods was performed. The molecular mechanisms of lipid effects on the peptide conformation and dynamics were identified.

In Chapter 8 preliminary results on the structural stability of pre-formed oligomeric composites of apoC-II(60-70) peptide of different sizes (dimer, trimer and tetramer) and arrangements (parallel and anti-parallel) were also presented. The most stable oligomer formation was a tetramer with the β-strands arranged in an anti-parallel conformation. The effects of mutation (oxidised Met, Met60Val and Met60Gln) on the most stable cluster were also investigated.

To conclude, several ideas for continuation of research in the protein folding and aggregation field are discussed in the Future Work section of this thesis.