ON THE EFFECTS OF EXTERNAL STRESSES ON PROTEIN CONFORMATION

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

B.H. (AKIN) BUDI
B. APP. SCI (APP. PHYS.) HONS.

Applied Physics
School of Applied Sciences
Science, Engineering and Technology Portfolio
RMIT University
March 2006
Declaration of the Candidate

I certify that except where due acknowledgement 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 result 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 third party is acknowledged.

B.H. (Akin) Budi
19 July 2006
I may not have gone where I intended to go, but I think I have ended up where I needed to be.

_Douglas Adams (1952–2001)_
Acknowledgements

I am typing this in the early evening of the day before my thesis submission. This section is the last thing I have to do before printing my thesis. I am sitting here in the corner of the lab, typing on this laptop at my messy desk, listening to Dire Straits, thinking of something witty or creative to write. Unfortunately, the lights are on but nobody’s home upstairs. So I think I’ll stick to the usual routine.

As the number of people to thank grows exponentially with the time scale of a project, I’d like to cover myself by thanking everybody. I owe you all my gratitude. Yes, that means you.

Now that I took care of the baseline and everybody is happy, I wish to thank a few people in particular. My supervisors: Irene, Herbert, and Ian for their guidance. Sue for a lot of help and feedback in the project. Sunnie for agreeing to read my thesis at a very short notice and giving his feedback. Yoke and John for help, feedback, and experimental support. David McKenzie for the initial idea of the project and the occasional meetings. RMIT University for the chance to do PhD and the scholarship covering my tuition fees. The Department of Applied Physics for the financial support and for being a great place to work in. I also thank APAC and VPAC where I ran all my simulations on. I thank the state library of Victoria for the wonderful Redmond Berry reading room \(^1\), where I spent a lot of my final weeks of PhD, writing myself silly. Thanks to my friends here, there, and everywhere (you know who you are!\(^2\)) for being such good chumps and chumpettes.

\(^1\)It’s the perfect place to do long writing sessions. I encourage people who are writing up to go spend some time there.

\(^2\)I know... I’m cheating by not naming you one by one. But my excuse is, if I did that, this acknowledgement page would become the a chapter instead.
Finally, I’d like to thank my parents for supporting me and letting me pursue my interests.

NAMD and VMD were developed by the Theoretical Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. PEPCAT was developed by Mark O’Donohue et al..

I wish to conclude with an excerpt from one of my favourite Leunig cartoons that I think sums up the PhD process quite nicely.

How To Get There

How to get there.

Go to the end of the path until you get to the gate.
Go through the gate and head straight out towards the horizon.
Keep going towards the horizon.
Sit down and have a rest every now and again.
But keep on going.
Just keep on with it.
Keep on going as far as you can.
That’s how you get there.

Akin Budi

March 30, 2006

This thesis was brought to you today by the letters B and σ, and by the number 42.
Abbreviations

APAC ................................. Australian Partnership for Advanced Computing
BPTI ........................................ Bovine Pancreatic Trypsin Inhibitor
CI2 ........................................... Chymotrypsin Inhibitor 2
DNA ............................................. Deoxyribonucleic acid
DPI .............................................. Despentapeptide insulin
DTT ............................................... dithiotreitol
MC ................................................ Monte Carlo
MD ................................................ Molecular dynamics
MSD ............................................. Mean square displacement
MSI ............................................. Molecular Simulations, Inc.
NAMD ........................................ Not Another Molecular Dynamics
NMR ............................................ Nuclear Magnetic Resonance
NOE ............................................ Nuclear Overhauser Effect
NPT .............................................. Constant number of particles, pressure, and temperature
NVT .............................................. Constant number of particles, volume, and temperature
PBC ................................................................. Periodic boundary condition
PCA ................................................................. Principal component analysis
PDB ................................................................. Protein Data Bank
PEPCAT ............................................................ Peptide Conformational Analysis Tool
REMD ............................................................... Replica Exchange Molecular Dynamics
RMS ................................................................. Root mean square
RMSD .............................................................. Root mean square displacement/deviation
SAR ................................................................. Specific absorption rate
SASA ............................................................... Solvent accessible surface area
sHSP ............................................................... small Heat Shock Protein
VAF ................................................................. Velocity autocorrelation function
vCJD ............................................................... variant Creutzfeldt-Jakob disease
VMD ................................................................. Visual Molecular Dynamics
VPAC .............................................................. Victorian Partnership for Advanced Computing
XRD ................................................................. X-ray diffraction
Publication List

The following publications have resulted from this study:

**Peer-reviewed publications**


**Conference proceedings**


**Journal articles in preparation**

1. A. Budi, S. Legge, H. Treutlein, I. Yarovsky, “Electric Field Effects on Monomeric Insulin Chain-B Conformation” *in preparation*

2. A. Budi, S. Legge, H. Treutlein, I. Yarovsky, “Frequency-Dependent Effects of Oscillating Electric Field on Insulin Chain-B Conformation” *in preparation*
Abstract

The use of electromagnetic devices such as microwave ovens and mobile phones has certainly brought convenience to our lives. At the same time, the proliferation of said devices has increased public awareness of the potential health hazards. It is generally assumed that there is little or no risk associated with the use of electromagnetic devices, based on the small amount of power associated with those devices. However, case studies on animals indicate that the risk cannot be entirely ruled out.

It has long been known that proteins are sensitive to stress, arising from various sources such as temperature, chemical, pressure, and changes in pH condition. In all of these cases, the protein exhibits clear signs of damage and distress, which range from slight unfolding to complete loss of structure. Frequently, the damage to the protein is alleviated by refolding, either by itself or by the aid of molecular chaperones. However, if the damage to the protein is too great, the protein will generally undergo proteolysis.

Opinion has been divided over the implication of prolonged use of electromagnetic devices to human health. Studies conducted on animals so far have given conflicting results. The studies on the separate components, electric and magnetic fields, also give inconclusive results. This indicates that our understanding on how electric and magnetic fields interact with biological matter is incomplete.

In this project, we use molecular dynamics to explore the behaviour of two forms of insulin chain-B, isolated and monomeric (in the presence of chain-A with all disulfide bonds intact), at ambient conditions and under the influence of various stress. Specifically, we focus our attention to thermal stress and electric field stress. The electric field stress
considered in this study takes several forms: static and oscillating with three different frequencies. These fields have strength ranging from $1806 \text{ V/m}$ to $10^9 \text{ V/m}$.

By performing molecular dynamics simulations totalling over 500 ns, we have gained valuable insights into the effects of elevated temperature and electric field on insulin chain-B. We observed differences in the damage mechanisms by the application of static electric field and oscillating field. The application of static fields restricts the conformational freedom of a protein, whereas the application of oscillating fields increases the mobility and flexibility of the protein, similar to the effect of thermal stress. Both of these interfere with the normal behaviour of a protein. We have also observed frequency-dependent effects, with low frequency fields having static field-like characteristics in damage mechanism.
Summary

This project utilises classical molecular dynamics simulation techniques to perform an all-atomistic study of the response of insulin chain-B under several forms of stress. A total of over 500 ns of simulation has been performed under various environments and stress conditions. The extent of the simulations has enabled this thesis to be a comprehensive study of the structural and conformational behaviour of the insulin chain-B peptide under various stresses, including thermal, chemical, and electric field stress.

Chapter 1 provides a review of the current state of knowledge in the field of protein dynamics and highlights the protein folding problem as one of the currently unsolved problems, despite decades of intensive study. The chapter also traces the progression from early attempts in the understanding of the protein dynamics and the protein folding problem, to the challenges that currently face the researchers in this field, especially in dealing with protein stress response.

Chapter 2 gives a review of the current methodologies and applications in protein simulation studies, along with their strengths and weaknesses. A broad array of techniques are discussed, including the simplified lattice models, the all-atomistic molecular dynamics models, and the highly accurate quantum mechanics/molecular mechanics hybrid models. In this chapter, an overview of the simulation and analysis techniques used in this project is also discussed.

In Chapter 3, the history of insulin is discussed, including the discovery process of insulin and the efforts involved in purifying and sequencing the insulin hormone. The structural motifs and variations of insulin are described, along with the differences be-
tween insulin produced by different species. A review of both experimental and computer simulation studies on insulin is also discussed. The results from simulations of insulin chain-B at ambient and thermal stress conditions are presented for two atomistic environments: isolated insulin chain-B, and chain-B in the presence of chain-A with all disulfide bonds intact. The results show that we could reliably replicate the features observed in the NMR structure, including generally satisfying NMR experimental NOE distance restraints better than the crystal structure. The application of thermal stress was observed to increase the conformational dynamics of both isolated and monomeric insulin chain-B by increasing the mobility and flexibility of the termini regions, with minimal structural change to the central B9–B19 helix region.

Chapter 4 contains the results of the simulation studies of insulin chain-B under static electric field stress which ranges from 1806 V/m to $10^9$ V/m. Just like Chapter 3, two atomistic environments were applied: isolated insulin chain-B, and chain-B in the presence of chain-A with all disulfide bonds intact. The results indicate that the application of static field aligns the proteins along their dipole moment axis, leading to a directional preference which can be exploited as a simulation technique to align a group of proteins. Restricted mobility was observed for proteins with a few, long helical regions, due to the interaction between the applied electric field with the dipole moment of that region. However, the presence of several helical regions (e.g. insulin chain-B in the vicinity of chain-A) was found to contribute to the total dipole moment, causing the protein to be susceptible to damage by a much weaker field strength.

Chapter 5 presents the results of the simulation studies of insulin under oscillating electric field of frequency 2.45 GHz. A methodology was developed to directly compare each of the oscillating field systems to the static field systems of equivalent electric field strength. Both isolated chain-B and chain-B in the presence of chain-A with all disulfide bonds intact were studied. The oscillating fields were observed to damage the secondary structure of the isolated chain-B systems at a lower field strength compared to the application of static fields of the same strength, due to a different mechanism of damage. However, in the presence of chain-A, the oscillating fields yield less damage due to the
interference of the cooperative effect of the dipole moments observed in the equivalent system under static field stress.

Chapter 6 serves as an extension of Chapter 5, where two different oscillating field frequencies (1.225 GHz and 4.9 GHz) were used. The frequencies correspond to half and double the oscillating field frequency used in the previous chapter. The results were then compared to the results presented in Chapter 5. They indicate that the effect of the oscillating fields is invariant of frequency at field strengths lower than $10^8$ V/m. However, at field strength of $10^8$ V/m or higher, the dipole moment of the protein was observed to interact more strongly with the applied field of frequency 1.225 GHz compared to the field of frequency 4.9 GHz. This is because at a lower oscillating field frequency, protein is exposed to longer periods of instantaneous high strength electric field as the oscillation reaches its maximum. Consequently, some of the systems showed a similar restriction to the protein structure to that observed in the static field simulations.

Chapter 7 contains the conclusion of the studies presented in this thesis, along with further work and potential applications of the presented findings. For example, a strong, intermittent static electric field can be applied to a group of simple helical proteins in order to align them and induce aggregation.
# Table Of Contents

Declaration of the Candidate .................................................. ii

List of Tables ........................................................................ xxiii

List of Figures ........................................................................... xxxiv

1 Introduction ............................................................................. 1

1.1 Overview ............................................................................. 1

1.2 Protein structure ................................................................. 2

1.2.1 Primary structure ............................................................ 2

1.2.2 Secondary structure ......................................................... 4

1.2.3 Tertiary and quaternary structure ........................................ 7

1.3 Protein folding problem ....................................................... 9

1.3.1 Energy landscape ............................................................. 9

1.3.2 Reversibility ................................................................. 11

1.3.3 Helices and folding mechanisms ........................................ 13

1.3.4 Hydrophobic interactions ............................................... 14

1.3.5 Intrinsic disorder and flexible proteins ......................... 17

1.4 Protein response to stress ................................................... 18

1.5 Project goals ..................................................................... 24

2 Computer Simulations of Proteins ......................................... 25

2.1 Overview ........................................................................... 25
2.2 Model system ................................................................. 26
2.3 Computer simulation in protein studies ............................. 28
  2.3.1 Early techniques ................................................. 28
  2.3.2 Ensemble dynamics ............................................. 29
  2.3.3 Restrained dynamics ........................................... 31
  2.3.4 Simplified model ............................................... 32
  2.3.5 Quantum mechanical techniques ............................... 33
2.4 Molecular dynamics .................................................. 34
  2.4.1 Equations of motion ............................................ 34
  2.4.2 Potential energy function ..................................... 37
  2.4.3 Periodic boundary conditions and minimum image convention ... 40
2.5 Methodologies .......................................................... 42
2.6 Application of stresses ............................................... 46
2.7 Analysis techniques .................................................. 48

3 Insulin Simulations: Ambient Conditions and Thermal Stress 58
  3.1 Overview ............................................................... 58
  3.2 Brief history .......................................................... 59
  3.3 Insulin structure .................................................... 61
  3.4 Methodologies ....................................................... 65
  3.5 Isolated insulin chain-B ........................................... 66
    3.5.1 Structural Stability .......................................... 66
    3.5.2 Comparison with NMR experimental results .................. 67
    3.5.3 Structural evolution ......................................... 68
    3.5.4 Conformational transitions .................................. 71
    3.5.5 Dipole moment ............................................... 78
    3.5.6 Radius of gyration ........................................... 79
    3.5.7 Solvent accessible surface area .............................. 79
    3.5.8 Biological implications ..................................... 80
3.6 Monomeric insulin chain-B ........................................ 82
  3.6.1 Stability ......................................................... 82
  3.6.2 PEPCAT analysis ............................................... 84
  3.6.3 Dipole moment ................................................. 87
  3.6.4 Radius of gyration ............................................ 87
  3.6.5 Solvent accessible surface area ............................... 88
3.7 Conclusions ........................................................ 88

4 Insulin Simulations: Static Electric Field Stress 90
  4.1 Overview .......................................................... 90
  4.2 Methodologies .................................................... 91
  4.3 Isolated insulin chain-B ......................................... 93
    4.3.1 Secondary structure ......................................... 93
    4.3.2 Dipole moment distribution ................................ 95
    4.3.3 PEPCAT analysis .............................................. 96
    4.3.4 Root mean square deviation ................................. 100
    4.3.5 Radius of gyration .......................................... 103
    4.3.6 Helical content .............................................. 104
    4.3.7 Solvent accessible surface area ............................ 108
  4.4 Chain-B in monomeric environment .............................. 109
    4.4.1 Secondary structure ......................................... 109
    4.4.2 Dipole moment distribution ................................ 111
    4.4.3 Root mean square deviation ................................ 112
    4.4.4 Radius of gyration .......................................... 115
    4.4.5 Helical content .............................................. 116
    4.4.6 Solvent accessible surface area ............................ 118
  4.5 Conclusion ........................................................ 119

5 Insulin Simulations: Oscillating Electric Field Stress 121
  5.1 Overview .......................................................... 121
5.2 Methodologies .................................................. 122
5.3 Isolated insulin chain-B ......................................... 123
  5.3.1 Secondary structure ...................................... 123
  5.3.2 Dipole moment distribution ............................... 125
  5.3.3 PEPCAT analysis ........................................... 127
  5.3.4 Root mean square deviation ............................. 131
  5.3.5 Radius of gyration ....................................... 133
  5.3.6 Helical content .......................................... 135
  5.3.7 Solvent accessible surface area ......................... 138
5.4 Chain-B in monomeric environment ............................ 139
  5.4.1 Secondary structure ...................................... 139
  5.4.2 Dipole moment distribution ............................... 141
  5.4.3 Root mean square displacement .......................... 142
  5.4.4 Radius of gyration ....................................... 144
  5.4.5 Helical content .......................................... 145
  5.4.6 Solvent accessible surface area ......................... 148
5.5 Conclusion .................................................... 149

6 Insulin Simulations: Frequency Dependent Electric Field Stress 151
  6.1 Overview ..................................................... 151
  6.2 Methodologies ............................................... 152
  6.3 Isolated insulin chain-B ..................................... 153
    6.3.1 Secondary structure .................................... 153
    6.3.2 Dipole moment distribution ............................. 156
    6.3.3 Root mean square deviation ............................. 157
    6.3.4 Radius of gyration ..................................... 160
    6.3.5 Helical content ........................................ 160
    6.3.6 Solvent accessible surface area ......................... 164
  6.4 Chain-B in monomeric environment ......................... 165
6.4.1 Secondary structure .............................................. 165
6.4.2 Dipole moment distribution ....................................... 167
6.4.3 Root mean square displacements .................................. 167
6.4.4 Radius of gyration .................................................. 170
6.4.5 Helical content ...................................................... 170
6.4.6 Solvent accessible surface area .................................... 173
6.5 Conclusion ............................................................. 174

7 Conclusions and Future Work ............................................. 176
  7.1 Conclusions ............................................................. 176
  7.2 Future work ............................................................. 180
List of Tables

3.1 Primary sequence of porcine insulin ........................................ 61
3.2 Primary sequence differences between different species ............... 62
3.3 Summary of simulation conditions used in this chapter. ............... 66
3.4 Retained number of all-helical and $\alpha$-helical residues for isolated insulin chain-B under ambient conditions and thermal stress, averaged over the data collection stage. ......................... 71
3.5 Radius of gyration of each simulation, averaged over data collection stage, after the first 1 ns. ........................................ 79
3.6 Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation. ......................... 79
3.7 Retained number of $\alpha$-helical residues for monomeric insulin chain-B under no field conditions, averaged over the data collection stage. ......................... 83
3.8 B9–B19 helical backbone RMSD of each monomeric chain-B simulation under no field conditions, averaged over the last 1 ns. .......................... 84
3.9 Radius of gyration of each monomeric chain-B system under no field conditions, averaged over data collection stage, after the first 1 ns. ......................... 88
3.10 Percentage of solvent accessible surface area of chain-B under monomeric condition throughout the data collection stage after the first 1 ns, relative to the NM300 simulation. .......................... 88
4.1 Summary of simulation conditions used in this chapter. ............... 92
4.2 Comparison of sampled states for isolated insulin chain-B under static electric field stress chosen for PEPCAT analysis with systems N300A and N400.

4.3 B9–B19 helix backbone RMSD of the isolated chain-B systems under static electric field, averaged over the last 1 ns.

4.4 Radius of gyration of each isolated chain-B system under static electric field, averaged over data collection stage, after the first 1 ns.

4.5 Retained number of all-helical and α-helical residues for isolated insulin chain-B under static field, averaged over the data collection stage.

4.6 Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation.

4.7 B9–B19 helical backbone RMSD of each monomeric chain-B simulation under static electric field, averaged over the last 1 ns.

4.8 Radius of gyration of each monomeric chain-B system under applied static electric field, averaged over data collection stage, after the first 1 ns.

4.9 Retained number of α-helical residues for monomeric insulin chain-B under static electric field, averaged over the data collection stage.

4.10 Percentage of solvent accessible surface area of chain-B under monomeric condition throughout the data collection stage after the first 1 ns, relative to the NM300 simulation.

5.1 Summary of simulation conditions used in this chapter (frequency of 2.45 GHz).

5.2 Comparison of sampled states for isolated insulin chain-B under oscillating electric field stress chosen for PEPCAT analysis with systems N300A and N400.

5.3 B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field, averaged over the last 1 ns, except systems O1 and O2 which were averaged between 2 ns and 3 ns.

5.4 Radius of gyration of each isolated chain-B system under oscillating electric field, averaged over data collection stage, after the first 1 ns.
5.5 Retained number of all-helical and α-helical residues for insulin chain-B under oscillating field, averaged over the data collection stage. ............... 135
5.6 Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation. ............... 139
5.7 B9–B19 helical backbone RMSD of each monomeric chain-B simulation under oscillating electric field, averaged over the last 1 ns, except for system OM1, where it was averaged between 1 ns and 2 ns. ................. 144
5.8 Radius of gyration of each monomeric chain-B system under applied oscillating electric field, averaged over data collection stage, after the first 1 ns. .......................................................... 144
5.9 Retained number of α-helical residues for monomeric insulin chain-B under oscillating electric field, averaged over the data collection stage, except system OM1 where averaging was done up to 2 ns. ................. 145
5.10 Percentage of solvent accessible surface area of chain-B under monomeric conditions throughout the data collection stage after the first 1 ns, relative to the NM300 simulation. .................. 148

6.1 Summary of simulation systems and conditions used in this chapter. ........ 153
6.2 B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field of different frequencies, averaged over the last 1 ns. 158
6.3 Radius of gyration of each simulation, averaged over data collection stage, after the first 1 ns. .......................................................... 160
6.4 Retained number of all-helical and α-helical residues for insulin chain-B under oscillating field of different frequencies, averaged over the whole data collection range. ................................. 161
6.5 Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation. ............... 164
6.6 B9–B19 helical backbone RMSD of each monomeric chain-B simulation under oscillating electric field of different frequencies, averaged over the last 1 ns. .......................................................... 168
6.7 Radius of gyration of monomeric chain-B systems under applied oscillating electric field of different frequencies, averaged over data collection stage, after the first 1 ns. ........................................... 170

6.8 Retained number of $\alpha$-helical residues for monomeric insulin chain-B under oscillating electric field of different frequencies, averaged over the data collection stage. ........................................... 171

6.9 Percentage of solvent accessible surface area of chain-B under monomeric conditions throughout the data collection stage after the first 1 ns, relative to the NM300 simulation. ........................................... 173
List of Figures

1.1 Schematic diagram of the two amino acid enantiomers. There are 20 known variations for the R side chain. ..................................................... 3
1.2 A schematic representation of the polypeptide nature of a protein. The protein backbone is highlighted in orange. The individual residues, peptide bonds, and the $\phi$, $\psi$, and $\omega$ angles are also shown. ................................. 3
1.3 The sulfur atoms (yellow) from two cysteine residues form a disulfide bond in an oxidation process. The reverse process is referred to as a reduction process. ................................................................. 4
1.4 Illustration of $\alpha$-helix, showing the hydrogen bonding (magenta) between the amino nitrogen and carbonyl oxygen. Carbon is shown in cyan, nitrogen is in blue, oxygen in red, and hydrogen in white. The protein backbone has been highlighted in orange. ...................................................... 5
1.5 Illustration of parallel and anti-parallel $\beta$-sheet secondary structures, with hydrogen bonding shown in red. ............................................................. 6
1.6 Ramachandran map, showing regions corresponding to common secondary structures. Darker colours signify increased probability of the secondary structure. ................................................................. 7
1.7 Ribbon representation of the tertiary structure of insulin, showing its three-dimensional backbone configuration in space. Chain-A is shown in red and chain-B in blue. The helix-loop-helix motif of chain-A is clearly visible. ... 8
1.8 Three insulin dimers form a hexamer in the presence of zinc ions and phenols. The view is along the symmetry axis formed by the zinc ions (not shown).

1.9 Damage probability to a protein at different stress power levels.

2.1 Summary of the accessible size and time scale for different types of simulation studies.

2.2 Different forces experienced by surface and bulk atoms.

2.3 Two dimensional illustration of periodic boundary conditions. The central box highlighted in orange is the main simulation box.

2.4 Two dimensional illustration of the minimum image convention. The dotted box represents all the particles that can be considered in the potential calculations on the yellow particle. The blue circle represents the use of potential cutoff. Only particles located within this circle experience interatomic interaction with the yellow particle.

2.5 The effect of potential truncation and switching function to a typical non-bonded interaction potential.

2.6 A typical simulation setup. The protein is shown in green, immersed in water inside a simulation box, which is then replicated in all three dimensions.

2.7 Summary of simulation stages used in this study.

2.8 Schematic diagram of the dipole moment that arises from helical regions of a protein, and the z-direction component that is analysed in this work.

2.9 Criteria used in the hydrogen bonding calculation. D is the donor atom, A is acceptor, and H is hydrogen. Hydrogen bond is highlighted in red.

2.10 Secondary structure evolution map of isolated insulin chain-B at 300 K. The vertical black line at around 8 ns denotes the point where the transition from R-state to T-state occurred.

2.11 Schematic diagram of solvent accessible surface area calculation. The solvent accessible surface area is the area traced by the centre of a probe on the van der Waals surface of a protein.
2.12 Solvent accessible surface area of insulin chain-B. The protein’s van der Waals surface is shown in yellow. A probe of radius 1.4 Å was used to generate the solvent accessible surface area, shown in red.

2.13 Illustration of PEPCAT classification process, reproduced from O’Donohue et al. (2000) with permission. The diagram shows the classification of a pyro-EEDCK trajectory in vacuo using the PEPCAT methodology. A trajectory consists of a time sequenced set of peptide coordinates (bottom right). Each frame is classified using a set of user specified descriptors (top right), and the set of values is looked up in a database of known states (left). If the set of descriptor values is not found in the database, a new entry is created. The conformational state identifier is returned (left) and stored with the frame. See text for more detail.

2.14 A typical PEPCAT conformation map. The circles denote conformational states, identified by each state ID, with size proportional to its relative population in the trajectory. The arrows connecting the circles denote transitions between the states, with size proportional to the amount of transitions between them. Refer to the text for full explanation of map features.

3.1 Tertiary structure of insulin. Chain-A is in blue, chain-B is in red, and disulfide bonds are in yellow.

3.2 Different conformations of insulin chain-B: R-state (yellow), T-state (green), R-state (pink), and O-state (orange). The conserved helical region comprising residues B9–B19 is highlighted in blue.

3.3 RMSD of backbone atoms of residues B9–B19 compared to the starting structure as a function of time for all isolated chain-B systems. The plot shows a moving average of 250 ps.
3.4 The NOE energy derived from the distance restraints of the starting structure compared to all systems. The shaded region represents the MD structures with NOE energy less than the starting structure. The plot shows a moving average of 250 ps.

3.5 Secondary structure plot for all systems. (A) N300A, (B) N300B, (C) N300C, (D) N300D, (E) N400.

3.6 Conformational transition map for all systems. (A) N300A, (B) N300B, (C) N300C, (D) N300D, (E) N400. The circles, labelled with the state identification number, represent conformational states visited during the simulation, and the size of the circle is proportional to the frequency. The lines between the circles indicate the conformational transitions and the width is proportional to the frequency. For classification of the states please refer to the accompanying text.

3.7 Typical structures from the major conformational states for selected systems. (A) N300A, (B) N300B, (C) N300D, (D) N400. The structures show the variation in the N-terminal region.

3.8 Conformational changes of insulin chain-B at 300 K. The initial change in structure of Gly B8 leads to two distinct conformations: Pathway A leads to T-state like conformations with low NOE energy in the N300A system; Pathway B leads to conformations trapped in local energy minima with high NOE energy in the N300D system. Both structures are stabilised by localised hydrophobic interactions.

3.9 Comparison of NOE energy and distance (C$_\alpha$ atoms) between N- and C- termini in selected systems (residues B6 to B26). (A) N300A. (B) N300D. The plot shows a moving average of 250 ps.

3.10 Time dependence plot of the distance between residues B5 and B13 (C$_\alpha$ atoms) for N300A, N300D and N400 systems. The plot shows a moving average of 250 ps.
3.11 B9–B19 helical dipole moment distribution in the \( z \)-direction of isolated insulin chain-B with no field applied. Both ambient and elevated temperature conditions show no directional preference.

3.12 Illustration of the role of individual residues. (A) Starting structure. (B), (C) low NOE energy structure (N300A system). A comparison between structures (A) and (B) show the stabilisation of the T-state conformation by a hydrogen bond interaction between the side chains of His B5 and Glu B13; and the movement of the side chains of Phe B24 and Phe B25 between the closed (A) and open (B) conformation. Structure (C) shows the two distinct binding sites of insulin to its receptor exposed with movement of the C-terminus.

3.13 Secondary structure evolution of the monomeric chain-B systems under static electric field.

3.14 100 ps running average of the B9–B19 helix backbone RMSD of the monomeric chain-B systems.

3.15 PEPCAT transition maps for systems considered for PEPCAT analysis. The yellow circles represent a unique state ID, with size proportional to the relative population of that state within the simulation timeframe. The grey arrows indicate transitions between states, with size of arrowheads representing the relative number of transitions to that state. Refer to Section 2.7 for full explanation of map features. For clarity, only states appearing more than 10% of the total simulation timeframe are presented, except for system NM300 where all states are presented.

3.16 Snapshots of the two most occupied states for the systems considered in this section.

3.17 Dipole moment distribution in the \( z \)-direction of insulin chain-B in the presence of chain-A.

4.1 Secondary structure evolution of the isolated chain-B systems under static electric field.
4.2 Illustration of the secondary structure disruption to system S1. System S1 at 0.5 ns (blue) is aligned with respect to the structure at the start of the simulation (red) on the remaining intact helical residues. The B9–B19 helix is highlighted in yellow. The secondary structure disruption is evident towards the N-terminal end of the helix.

4.3 B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under static electric field. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.

4.4 Snapshot of the last frame of the simulations of isolated chain-B systems, showing the alignment of the helix with respect to the electric field applied in the z-direction. N-terminus is coloured blue, C-terminus is coloured red, and the B9–B19 helix is highlighted in black.

4.5 PEPCAT transition maps for systems considered for PEPCAT analysis. The yellow circles represent a unique state ID, with size proportional to the relative population of that state within the simulation timeframe. The grey arrows indicate transitions between states, with size of arrowheads representing the relative number of transitions to that state. Refer to Section 2.7 for full explanation of map features. For clarity, only states appearing more than 10% of the total simulation timeframe are presented.

4.6 Snapshots of the two most occupied states for the isolated insulin chain-B under static electric field stress.

4.7 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under static electric field.

4.8 100 ps running average of the B9–B19 helical content retained as a function of time for the isolated chain-B systems under static electric field. Red lines indicate the all-helical content, and the green lines indicate the \( \alpha \)-helical content.
4.9 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for systems S1, S2, and S3. .......................... 108
4.10 Secondary structure evolution of the monomeric chain-B systems under static electric field. .......................................................... 110
4.11 B9–B19 helical dipole moment distribution in the $z$-direction of insulin chain-B in the presence of chain-A. Each plot has been displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field. .......................................................... 111
4.12 Illustration of the stronger coupling between the electric field and the chain-B helical dipole moment. For each electric field strength, the dipole moment in the $z$-direction for monomeric insulin chain-B (red) is stronger than for isolated chain-B (green). Each of the plots has been displaced by 0.2 for clarity. .......................................................... 112
4.13 Snapshot of the last frame of the simulations of monomeric chain-B (left of each subfigure) systems and isolated chain-B (right of each subfigure), showing the stronger alignment of the monomeric chain-B helix with respect to the electric field applied in the $z$-direction, compared to the isolated chain-B helix. N-terminus is coloured blue, C-terminus is coloured red, and the B9–B19 helix is highlighted in black. Chain-A is present throughout the simulations of monomeric insulin chain-B, but is omitted from the diagrams for clarity. .......................................................... 113
4.14 100 ps running average of the B9–B19 helix backbone RMSD of the monomeric chain-B systems under static electric field. ......................... 114
4.15 100 ps running average of the helical content retained as a function of time for the monomeric chain-B systems under static electric field. Red lines indicate the all-helical content, and the green lines indicate the $\alpha$-helical content. .......................................................... 117
4.16 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for systems SM1 and SM2. ......................... 118
5.1 The oscillating electric field was approximated by changing the electric field strength every 20 fs. 123
5.2 Comparison of static and oscillating electric fields applied to the system. One period of the oscillation is highlighted in black. 123
5.3 Secondary structure evolution of the systems under oscillating electric field. 125
5.4 Comparison of systems S2 and O2, taken at the end of the simulation. System S2 is shown in red and system O2 is shown in green. The B9–B19 residues are shown in yellow. 126
5.5 Illustration of the T-like state formed by systems N300A (red) and O4 (green). The starting, Rf-state (blue) is shown as a comparison. Residues B9–B19 are highlighted in black. 127
5.6 B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under oscillating electric field. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field. 128
5.7 PEPCAT transition maps for systems considered for PEPCAT analysis. The yellow circles represent a unique state ID, with size proportional to the relative population of that state within the simulation timeframe. The grey arrows indicate transitions between states, with size of arrowheads representing the relative number of transitions to that state. Refer to Section 2.7 for full explanation of map features. For clarity, only states appearing more than 10% of the total simulation timeframe are presented. 129
5.8 Snapshots of the two most occupied states for the isolated insulin chain-B under oscillating electric field stress. 130
5.9 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field. 132
5.10 Illustration of the secondary structure disruption to systems O1 and O2. Both structures were obtained at 1.5 ns. The B9–B19 helix is highlighted in orange. 134
5.11 100 ps running average of the B9–B19 helical content retained as a function of time for the isolated chain-B systems under oscillating electric field. Red lines indicate the all-helical content, and the green lines indicate the α-helical content. .................................................. 136

5.12 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for systems O1 and O2. ......................... 138

5.13 Secondary structure evolution of the monomeric chain-B systems under oscillating electric field. ................................................................. 140

5.14 B9–B19 helical dipole moment distribution in the z-direction of insulin chain-B in the presence of chain-A. Each plot has been displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field. .................................................. 141

5.15 Comparison of the B9–B19 helical dipole moment distribution in the z-direction of the monomeric insulin chain-B (red) and the isolated chain-B (green). Each of the plots have been displaced by 0.2 for clarity. ........... 142

5.16 100 ps running average of the B9–B19 helix backbone RMSD of the monomeric chain-B systems under oscillating electric field. ......................... 143

5.17 100 ps running average of the helical content retained as a function of time for the monomeric chain-B systems under oscillating electric field. Red lines indicate the all-helical content, and the green lines indicate the α-helical content. .................................................. 146

5.18 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for systems OM1 and OM2. ................. 147

6.1 Comparison of oscillating electric field frequencies of 2.45 GHz, 1.225 GHz, and 4.9 GHz. ................................................................. 152

6.2 Secondary structure evolution of the isolated chain-B systems under oscillating electric field of different frequencies. ............................... 154

xxxii
6.3 Snapshots of isolated insulin chain-B conformations at 6 ns for systems O2H, O2D, O3H, and O3D. Disruption to the secondary structure is observed in all systems except O3H. Residues B9–B19 are highlighted in orange.

6.4 Comparison of the oscillating fields of different frequencies shows that within the same time period, the low frequency oscillating field does not change as much as the high frequency oscillating field. The maximum of each electric field frequency has been aligned with respect to the maximum of the electric field of frequency 4.9 GHz.

6.5 B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under oscillating electric field of different frequencies. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.

6.6 Comparison of the B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under static electric field and oscillating electric field of frequencies 2.45 GHz, 1.225 GHz, and 4.9 GHz. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.

6.7 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field of different frequencies.

6.8 100 ps running average of the B9–B19 helical content retained as a function of time for the isolated chain-B systems under oscillating electric field of different frequencies. Red lines indicate the all-helical content, and the green lines indicate the α-helical content.

6.9 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for isolated chain-B systems under electric field of different frequencies with strength equal to or above $10^8$ V/m.

6.10 Secondary structure evolution of the monomeric chain-B systems under oscillating electric field of different frequencies.
6.11 B9–B19 helical dipole moment distribution in the z-direction of insulin chain-B in the presence of chain-A, under the application of oscillating fields of different frequencies. Each plot has been displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.

6.12 100 ps running average of the B9–B19 helix backbone RMSD of the monomeric chain-B systems under oscillating electric field of different frequencies.

6.13 100 ps running average of the helical content retained as a function of time for the monomeric chain-B systems under oscillating electric field of different frequencies. Red lines indicate the all-helical content, and the green lines indicate the α-helical content.

6.14 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for systems OM2H and OM2D.
Chapter 1

Introduction

“What is the secret of life?” I asked.

“I forget,” said Sandra. “Protein,” the bartender replied.

Cat’s Cradle (1963) by Kurt Vonnegut, Jr. (1922–)

1.1 Overview

In this chapter, a brief introduction to protein structure is presented in Section 1.2. A great emphasis is placed on the discussion of secondary structure, as it is later used to quantify the structural and conformational changes on insulin chain-B in subsequent chapters. In Section 1.3, a detailed review and history of the current emphasis of research efforts in the field of protein studies is presented. This section leads to the topic with which this thesis is concerned, which is the protein response to stress. A detailed review of the current knowledge on protein response to various kinds of stresses is presented in Section 1.4. Finally, the aims of this project are given in Section 1.5.
1.2 Protein structure

A protein is a biological macromolecule composed of carbon, nitrogen, oxygen, sulfur, and hydrogen atoms. It is an essential component of life with very diverse roles, such as regulating metabolic pathways, carrying out the transport and storage of small molecules, making up a large part of the structural framework of cells and tissues, and controlling the growth and repair of living tissues [1]. Some of the important kinds of protein include hormones, antibodies, and enzymes.

Due to their structural complexity, protein structure is classified by several hierarchical categories. In order of increased complexity, the structural categories are: primary, secondary, tertiary, and quaternary. The more basic structural features determine the nature of the more complex ones.

1.2.1 Primary structure

The primary structure of a protein refers to a sequence of amino acid residues, which are the building blocks that uniquely define a protein. The sequence of residues that forms a protein is encoded in the DNA (deoxyribonucleic acid). The residues are composed of an amine group followed by an α-carbon (C$_{\alpha}$) and a carboxyl group. A variable R-group called a side chain is attached to the α-carbon, forming the individual amino acid residues and determining their properties. Even though there are only 20 commonly occurring amino acids, they can combine linearly in an arbitrary way, forming an enormous collection of proteins sized from ten to ten thousand of residues. The chirality of the C$_{\alpha}$ atom gives rise to two enantiomers, illustrated in Figure 1.1.

The amino acids combine with each other to form proteins (short proteins are known as peptides) by forming a peptide bond, illustrated in Figure 1.2. The resulting –N–C$_{\alpha}$–C– sequence is called the protein backbone. There are 3 essential torsion angles that determine the conformation of the protein backbone: φ, ψ, and ω. φ angle is the angle subtended by the two planes formed by the C–N–C$_{\alpha}$–C atoms (i.e. rotation about the N–C$_{\alpha}$ line). ψ angle is the angle subtended by the two planes formed by the N–C$_{\alpha}$–C–N atoms (i.e.
CHAPTER 1. INTRODUCTION

Figure 1.1: Schematic diagram of the two amino acid enantiomers. There are 20 known variations for the R side chain.

rotation about the Cα–C line). ω angle is the angle subtended by the planes formed by the Cα–C–N–Cα atoms (i.e. rotation about the C–N line). However, ω angle always has a value close to 180° (except for cis-prolines, which have ω-angle close to 0°) and forms a plane called the peptide plane. Therefore, it is the φ and ψ angles that determines the major structural features of a protein, which is reflected in its secondary structure. In the literature, the dihedral angles are usually cited as (φ,ψ) in units of degrees.

Figure 1.2: A schematic representation of the polypeptide nature of a protein. The protein backbone is highlighted in orange. The individual residues, peptide bonds, and the φ, ψ, and ω angles are also shown.

Cysteine residues have a special property that allows them to form a covalent bond with another cysteine residue in an arrangement called a disulfide bond (shown in Figure 1.3). Being a covalent bond, the disulfide bond is extremely strong compared to ionic and
hydrophobic bonds, and is essential to the proper folding, structure, and functioning of a protein.

Figure 1.3: The sulfur atoms (yellow) from two cysteine residues form a disulfide bond in an oxidation process. The reverse process is referred to as a reduction process.

1.2.2 Secondary structure

The secondary structure refers to ordered local structural features within a protein. It can be classified into several categories: helices, sheets, turns, and loops.

1. Helices

Helices are formed when sequences of amino acid residues form a winding motif, like a spiral staircase. The motif is stabilised by the formation of parallel hydrogen bonds between the backbone atoms along the amino acid sequence. The dipole
moments that arise from the slight electronegativity of the carboxyl oxygen of each residue add up constructively to form the helix dipole moment. As such, helices are inherently quite strong and usually resistant to external stresses. The helix dipole moment is known to contribute to the binding of charged groups at the helix termini, the long range attraction of charged molecules, and the enhancement of reaction rates [2]. There are 3 kinds of commonly occurring helices: \( \alpha \)-helices, \( 3_{10} \)-helices, and \( \pi \)-helices—varying according to the hydrogen bond coordination of their backbone atoms. \( \alpha \)-helices have bonds between residues \( i \) and \( (i+4) \); \( 3_{10} \)-helices have bonds between residues \( i \) and \( (i+3) \); and \( \pi \)-helices have bonds between residues \( i \) and \( (i+5) \). Ideal \( \alpha \)-helices have dihedral angles of \((-57.8, -47.0)\); ideal \( \pi \)-helices have dihedral angles of \((-57.1, -69.7)\); and ideal \( 3_{10} \)-helices have dihedral angles of \((-74.0, -4.0)\). Figure 1.4 illustrates an \( \alpha \)-helix, stabilised by the hydrogen bonds between the backbone atoms.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{alpha_helix.png}
\caption{Illustration of \( \alpha \)-helix, showing the hydrogen bonding (magenta) between the amino nitrogen and carbonyl oxygen. Carbon is shown in cyan, nitrogen is in blue, oxygen in red, and hydrogen in white. The protein backbone has been highlighted in orange.}
\end{figure}

2. \( \beta \)-sheets

\( \beta \)-sheets are composed of elongated strands of the peptide chain created by the formation of hydrogen bonds between backbone atoms, perpendicular to the direction of the strand. The hydrogen bonding is weaker than in helices, but \( \beta \)-sheets are very important in processes like aggregation once a hydrophobic contact has been made [3, 4]. \( \beta \)-sheets have average dihedral angles of \((-120, 120)\) and come in two varieties: parallel and anti-parallel. An illustration of \( \beta \)-sheets is given in Figure 1.5.
3. Turns and loops

Turns and loops are essential in enabling a protein to fold and form compact tertiary structures that minimise the protein’s solvent accessible surface area (SASA). They are also essential in enabling flexibility of regions in a protein structure, often related to its activity.

The secondary structures are formed when consecutive residues have very similar $\phi$ and $\psi$ angle values that lie within known regions on the Ramachandran map [5]. The Ramachandran map provides a convenient way of categorising a protein’s structure based on the dihedral angles. Broadly speaking, there are three different regions in the Ramachandran plot, which correspond to the secondary structures with well-defined dihedral angles, illustrated in Figure 1.6: the right-handed $\alpha$-helix, the left-handed $\alpha$-helix, and the $\beta$-sheet. There is a preference towards right-handed $\alpha$-helix due to the prevalence of L-amino acids.

Only selected regions of the Ramachandran plot are accessible to amino acids because of steric restrictions of the side chains. The only exception to this rule is the glycine (Gly) residue due to its lack of side chain. This allows Gly a greater degree of freedom compared to other amino acid residues and thus it is often associated with turns in a protein’s secondary structure. This feature can be important for protein activity, and the presence of Gly is often conserved across species and across a protein family as the
CHAPTER 1. INTRODUCTION

1.2.3 Tertiary and quaternary structure

The tertiary structure refers to a protein’s three-dimensional configuration in space. There are some well known motifs of a tertiary structure, for example the well studied helix-loop-helix motif, appearing in insulin chain-A. A typical schematic of tertiary structure is shown in Figure 1.7. The formation of tertiary structure is driven by hydrophobic interactions between the secondary structures which minimise the hydrophobic surface area accessible to water [8]. The role of hydrophobic residues in directing the formation of ter-

terminal residue of a helical section of the secondary structure. Gly residue can also occur in the middle of a helix. In this case, the greater conformational flexibility of Gly enables it to function as a helix breaker which allows the protein to adopt biologically active conformations, as part of the normal dynamics behaviour [6]. An example of such behaviour can be found in the insulin molecule where the Gly residue allows insulin chain-B to adopt its biologically active T-state conformation [7].

Figure 1.6: Ramachandran map, showing regions corresponding to common secondary structures. Darker colours signify increased probability of the secondary structure.
tiary structure has also been confirmed by studies of hydrophobic dipole moment (dipole moment arising from the difference in hydrophobicity) [9] and buried surface areas [10]. Aside from aiding in the formation of tertiary structure, hydrophobic residues also play an important role in the stabilisation of tertiary structure [11].

Figure 1.7: Ribbon representation of the tertiary structure of insulin, showing its three-dimensional backbone configuration in space. Chain-A is shown in red and chain-B in blue. The helix-loop-helix motif of chain-A is clearly visible.

Figure 1.8: Three insulin dimers form a hexamer in the presence of zinc ions and phenols. The view is along the symmetry axis formed by the zinc ions (not shown).

The quaternary structure is formed when a group of proteins behave in a collective manner. While usually the proteins are of the same species, different proteins can also
form a quaternary structure. An example of quaternary structure is the formation of the insulin dimer, consisting of two monomers stabilised by hydrogen bonds. Three insulin dimers can cluster around metal ions such as zinc and form hexamers [12], a form of insulin stored inside the body for future use. This arrangement is illustrated in Figure 1.8.

1.3 Protein folding problem

1.3.1 Energy landscape

The protein folding problem refers to the process where a protein in a random coil state rearranges to adopt a predetermined and functional three-dimensional shape. For years this problem has eluded the best minds in biochemistry, since the information that allows a protein to adopt the three-dimensional structure seems to be encoded within its primary structure, without any external influence [13, 14]. In other words, the formation of the native state is achieved by sampling the conformational space available to that protein until it reaches its native state. A related problem that eluded biochemists is the mechanism of native state formation by creation of disulfide bonds involving cysteine residues. Specifically, the formation of disulfide bonds seems to be random at first, with subsequent thermodynamically-driven rearrangement taking place to form the native state [13]. This is referred to as the thermodynamic control of protein folding, since the goal of folding is believed to be to find the conformation with the lowest energy.

The problem with this approach is that a protein would take an astronomically large time scale to do a blind search in the available conformational space to reach its native state. A conservative estimate puts the number of conformations sampled by a typically sized protein to be $10^8$ states before the native fold can be found, and this would take many years to achieve. However, it is clear that in reality, most proteins fold in a matter of seconds. This seemingly conflicting behaviour is known as the Levinthal’s paradox [15]. Levinthal suggested the existence of a folding pathway allowing the protein to fold in a realistic timeframe, thereby providing a kinetic control to the folding process.
CHAPTER 1. INTRODUCTION

While the pathway model satisfies the requirement for a protein to fold within biological time scale, it implies the existence of only one route for a protein to fold/unfold. The existence of intermediate states (I) in relation to the folded (N) and unfolded (U) states, therefore, have been classified as on-pathway [16]:

\[ U \rightleftharpoons I \rightleftharpoons N \]

and off-pathway [17]:

\[ U \leftarrow I \rightarrow N \]

with off-pathway intermediates believed to take no part in the actual folding process. It soon became obvious that rather than a single pathway, a protein can fold via many different pathways, each with its own set of transition states [18, 19]. All the different pathways lie on a funnel-like energy landscape, which is referred to as the folding funnel model [20, 21]. In this model, a protein starts from a region of high energy and gradually goes down the folding funnel until it reaches the region of minimum energy. While there is statistical preference towards a specific pathway [22], the “off-pathway” intermediates may also appear in pathways that lead to the native state. This model satisfies Anfinsen’s thermodynamic control and Levinthal’s kinetic control. The existence of a protein folding pathway was confirmed as early as 1975 by Levitt and Warshel in their computer simulation study of bovine pancreatic trypsin inhibitor (BPTI) [23].

The real protein folding energy landscape is very rugged with local minima and energy barriers [24, 25]. The ruggedness means there are many pathways to protein folding, with different pathways leading to different local minima, where the protein can be trapped for some time before it continues on the folding funnel. In some cases, different pathways may lead to the same local minimum, which forms a rate-limiting process in the folding event [26]. In some proteins, the potential energy landscape can be smooth on one side and rough on the other, giving rise to fast and slow folding, respectively. This has been
observed in hen lysozyme [27, 28]. The local minima in the energy landscape give rise to intermediate states. The folding funnel model has been confirmed experimentally using NMR, which revealed the presence of a transition state ensemble, which is dominated by compact conformations [29].

The first suggestion that a protein folds through some intermediates was made by Paul Flory in 1968 [30]. This idea was then extended by Ptitsyn [31], who suggested that there are two types of intermediate states, one with slightly disordered secondary structure but with similar tertiary structure to the native conformation, and another with more defined secondary and tertiary structures that resemble the native conformation. Ptitsyn’s insight proved to be correct when an intermediate state of α-lactalbumin was identified which had native-like secondary structure, but fluctuating tertiary structure [32]. This was subsequently referred to as the molten globule conformation, due to the “liquid-like” behaviour it exhibited. The molten globule conformation has since been identified in many other proteins [33–36]. This conformation has now been widely accepted as a third distinct thermodynamic state of a protein, alongside the folded (native) and unfolded states [37].

1.3.2 Reversibility

Many proteins can undergo complete and reversible folding/unfolding, and have been observed to do so from as early as the 1930s [38]. Some of the proteins exhibit reversibility under the influence of different types of stresses, such as the mouse prion protein under the influence of guanidinium chloride [39] and sepiolite under thermal stress [40]. Helices are known to undergo reversible folding/unfolding and have been used as a model system to predict the properties and timeframe of protein folding [41, 42]. Engineered proteins can also exhibit this behaviour and have been used to study processes such as β-sheet formation [43].

The principle of microscopic reversibility states that folding and unfolding processes are the same under the same conditions [44]. Protein engineering has provided evidence that microscopic reversibility is applicable to protein folding/unfolding processes [45, 46],
with transition states of folding shown to be identical to the transition states in unfolding [45]. The first known reversible folding/unfolding study using computer simulations was performed by Daura et al. in 1998. A $\beta$-heptapeptide in methanol solution was studied using 50 ns molecular dynamics simulations at different temperatures. They concluded that folding can occur within the simulation timeframe as long as the solvent molecules are present to facilitate the refolding [47]. Methanol was used in this study because it is a strongly hydrogen bonding solvent with lower density than water, thereby less expensive to use in simulations. However, a more accurate description of conditions inside real biological systems calls for the use of water as a solvent. This presented a problem, as the longest simulation of protein folding in water thus far—at the time of writing—lasted for 1 $\mu$s without actually observing the folding process [48]. There are a number of ways to actually study the folding process. Some of the most popular methods include starting with an initial guess of the pathway between limiting structures [49], directing the simulation based on known initial and target structures [50], using specialised potential to drive protein to its native state [51], and using activated dynamics [52, 53] with higher temperature [54]. The latter is especially useful in the study of the unfolding event, as the reversibility theorem implies that the early stages of unfolding are identical to the later stages of folding [55, 56].

The use of high temperature to induce the unfolding event has been questioned by several authors because of concerns that the high temperature unfolding does not necessarily follow the folding process at physiological conditions [57–59]. The theory of microscopic reversibility strictly holds true only if the conditions for forward and reverse processes are the same. However, there have been several studies conducted on thermal unfolding and refolding of proteins which showed that the transition states sampled during the unfolding and refolding events agree very well [59–61]. Dinner and Karpplus noted that although the pathways for folding and unfolding are very similar, high temperature simulations do not sample intermediate states, and are best applied to fast folding events [59]. Finally, a study by Day et al. showed that the overall pathway of unfolding is independent of temperature and the process is merely accelerated by increasing temperature. They proved this by
performing an unfolding study of chymotrypsin inhibitor 2 (CI2) at a lower, experimentally accessible temperature, and found very good agreement between their simulations and experiments [62].

1.3.3 Helices and folding mechanisms

A cooperative process in protein folding is a process where the formation of a hydrogen bond lowers the energy requirement for subsequent hydrogen bonds. This view is the current accepted model used to explain the folding process [63] and α-helix formation [64]. Dintzis has established that the folding process is not a random process, but rather a sequential addition of amino acid residues that progresses from the N-terminus towards the C-terminus [65]. The addition of hydrogen bonds has also been shown to proceed from the N-terminus subject to the degree that solvent can penetrate to and interact with the potential hydrogen bonding groups [66]. This cooperative process is referred to as the “zipper” model [63, 67]. It was first introduced to describe the mechanics of helix–coil transition [41, 42]. The zipper model was observed to occur in the ribonucleic acid (RNA) helix, with the formation of a few, correct adjacent base pairs, which then “zippered up” rapidly [68]. Baldwin et al. then hypothesised that protein folding also follows the same nucleation mechanism [18]. The intermediate state of ribonuclease A was subsequently discovered using the T-jump experiment, which confirmed the nucleation or zipper mechanism in the protein folding process [16]. The theory was later extended by Dill et al. to cover early stages of protein folding [67] and verified by Bakk et al. in a study of heat capacity of myoglobin [69]. Bakk et al. also extended the zipper model by including the possibility of multiple sites for folding/unfolding and introduction of water interactions [70]. This model successfully explained phenomena previously unaccounted for in the original zipper model, such as cold and warm denaturation [71–74].

Using computer simulations, the time scale of secondary structure formation (e.g. helix) was suggested to be about 0.1–1 ns [75], with complete formation within 20–70 ns [76]. This time scale is also supported by spectroscopic measurements [77]. The formation of tertiary structure, however, is much slower. With the advent of faster and more powerful
computers, numerous protein folding studies have been conducted. However, the simulation timeframe for these studies has been limited to tens of nanoseconds to a microsecond at best [78], which is the lower limit of the observed unfolding and folding timeframe for real proteins [29, 30, 79–81]. Indeed, even with 1 µs of continuous simulation, Duan et al. could only see the early stages of protein folding rather than the folding process itself [78].

The study of helix–coil transformation provides a useful insight into the early stages of folding. Unfolding studies conducted in water suggested that the helix unfolding mechanism involves $3_{10}$-helices as intermediates in the transformation from $\alpha$-helix to random coil conformation, with strong preference for unfolding to be initiated from the C-terminus [54, 82–84]. An alternate mechanism was observed in an unfolding study of a 36-residue $\alpha$-helical protein conducted in a weakly polar medium. In this study, the unfolding proceeds through the formation of local regions at the centre of the helix, stabilised by $\pi$-helical hydrogen bonds (hydrogen bonds between residues $i$ and $(i + 5)$), which propagate to metastable structures with partially solvated backbone intermediates, forming an $\alpha$-helix–random coil transition [85].

1.3.4 Hydrophobic interactions

Hydrophobic bonding is defined by Kauzmann as the gain in free energy upon the transfer of nonpolar residues from an aqueous environment to the protein interior [86]. In order for the protein atoms to interact with the solvent molecules, they must be able to form van der Waals contacts, the process described quantitatively by a concept of accessible surface area by Lee and Richards [87]. The hydrophobic interaction has since been accepted as one of the major forces involved in protein folding, confirmed experimentally by Chothia who calculated that every $\AA^2$ of accessible surface area removed from contact with water gives the free energy gain of 25 cal/mol [88]. Chothia also noted that secondary structures are very hydrophobic, with two third of non-polar and more than half of the polar groups forming hydrogen bonds [8]. The formation of a protein’s secondary structure leads to the burial of main chain polar atoms, leaving more hydrophobic surface groups accessible to water. To increase the solubility of the final structure, the hydrophobic surface must
be removed from contact with water by burying it between the fragments of secondary structure [8].

Due to its important role in protein structure and function, there have been many attempts to describe hydrophobicity by quantifying the free energy change by the transfer of amino acid chains from water to an apolar environment [88–91]. However, the hydrophobicity scales depend on the exact details of the method and conditions used. In 1982, Eisenberg et al. combined hydrophobicity scales taken from 5 methods and formed the Eisenberg-Weiss hydrophobicity scale, considered to be the most reliable [91]. They also developed a method to calculate the “hydrophobic moments” by analogy to the electric moments. The first moment (hydrophobic dipole moment) is analogous to the electric dipole moment, and is a measure of the amphiphilicity (asymmetry of hydrophobicity) of the structure [9, 91]. A large hydrophobic dipole moment indicates the structure is mostly hydrophobic on one side and hydrophilic on the other. In globular proteins, the hydrophobic dipole moments in different regions point towards a common centre. This means the hydrophobic surfaces form a cluster around a common centre, and is indicative of the role of such surfaces in directing the folding of proteins. This finding is confirmed by Rose et al. who determined that hydrophobicity correlates well with the average buried surface area in globular proteins [10].

The hierarchical folding model states that the folding process is initiated locally and the formation of the native local structures driven by hydrophobic interactions (“hydrophobic folding units”) precedes the formation of tertiary interactions [92, 93]. The folding units then group together to form domains, which then lead to the formation of the final tertiary structure. Hydrophobic interaction also plays a major role in the formation of \( \beta \)-hairpins. A study conducted on \( \beta \)-hairpin formation indicated they are stabilised by both hydrogen bonding and hydrophobic interactions, and exhibit two-state behaviour and a funnel-like, partially rugged energy landscape [94]. The works by Pande and Rokshar [60] and Muñoz et al. [95] established that the positioning of the side chains is important for the formation of a hydrophobic cluster which in turn is essential for the folding of the hairpin structure. However, the details of the two works are slightly different. Muñoz et al. argued for the
zipper model where folding is initiated from the $\beta$-turn, which then zips up the remaining hydrogen bonds. In contrast, Pande and Rokshar suggested the formation of hydrophobic clusters which then collapse, with non-native hydrogen bonds being repaired as the process progresses.

The two groups present two differing views in protein folding theory. The model put forward by Muñoz et al. is very similar to the diffusion-collision model, where the first event in protein folding is the formation of secondary structure elements, which then diffuse and coalesce to form the stable protein [96–99]. This model was challenged in 1975 by Levitt and Warshel who observed the forming and breaking of secondary structure, which is inconsistent with the theory that stable secondary structures form first in the folding process [23]. However, it has been successfully applied to describe the folding of the N-terminal domain of $\lambda$-phage repressor protein [100]. The model suggested by Pande and Rokshar, on the other hand, is based on the hydrophobic collapse model, where the rapid formation of a hydrophobic core precedes the formation of secondary and tertiary structures [60, 101]. This model has been successfully used to explain the folding of an $\alpha/\beta$ protein and confirmed by experiment [102]. Both of these folding models concentrate on the specific pathway towards the folded conformation, with the intermediate states providing crucial information towards understanding the folding process.

Hydrophobic residues also play an integral role in the stabilisation of a folded protein. For example, Lasalle et al. recently concluded that hydrophobic interactions are essential to stabilise the molten globule structure of $\alpha$-lactalbumin based on a study on pressure-induced unfolding of the protein [11]. Hagen et al. suggested a biological significance of the mechanism of protein collapse: the rapid collapse of hydrophobic residues is essential to prevent aggregation in unchaperoned proteins. Once the initial collapse has taken place, the formation of native structure may proceed more slowly [79]. Another study of the hydrophobic domain cores showed higher resistance to denaturation compared to a hydrogen bonding network, thus demonstrating the robustness of the hydrophobic domain [103]. The hydrophobic interactions were able to maintain the peptide fold despite the repulsive force between $\beta$-strands [104]. This implies that the disruption
of the hydrophobic core is one of the major steps in the folding/unfolding process for most proteins [105]. For example, the intermediate states in an unfolding study of barnase have compromised hydrophobic cores and lost some of their native hydrogen bonds [106].

Another model put forward is the nucleation-propagation model, where the early stages in folding may involve the independent formation of native domain contacts in different regions of the protein ("nucleation") which then grow rapidly ("propagation") by adding peptide chain segments that are close to the nucleus in the amino acid sequence to form the native structure in the protein [107].

1.3.5 **Intrinsic disorder and flexible proteins**

Traditionally, it was thought that there is a connection between protein structure and function, with an active protein having a certain conformation. This view was fostered by the notion of the "lock-and-key" mechanism proposed by Fischer [108, 109]. It was further confirmed by an independent review compiled by Mirsky and Pauling [38] which showed a correlation between the loss of pepsin activity and the degree of protein denaturation. The first specific reference relating protein function to protein structure occurs in a paper by Wu [110], as noted by Edsall [111]. White and Anfinsen also investigated the structure-function relationship, along with identifying the residues that are essential for ribonuclease's activity [112]. In addition, they also deduced that disulfide bonds are essential for activity—reduction caused loss of activity, whereas upon oxidation, the enzyme regained its activity.

Recent discoveries of intrinsically disordered proteins suggest, however, that the structure-function relationship paradigm is incomplete. Many proteins, found to show activity, are not restricted to ordered conformation. Intrinsic disorder can give functional advantages to a protein, including the ability to bind to several different targets and more precise control over the thermodynamics of the binding process. Interestingly, some of these proteins can become structured upon binding to the target [113–118]. Spolar and Record noted that the folding induced upon binding to the target can increase the specificity of molecular recognition. This is because of the free energy requirement needed to fold
the protein, which means the protein will only bind to targets whose complementarity is maximal [119]. These discoveries led Dunker and coworkers to coin the “Protein Trinity” model, in which the ordered, molten globule, and random coil conformations are equally important [120]. Protein function can arise from any of the conformations and transitions between them.

Conformational flexibility has a very important role in the functional properties. For example, proteins that are composed of multiple domains may rely on the flexibility offered by the hinge region in order to adopt an active conformation [121]. Another example of the inherent conformational flexibility can be found in insulin [122], which has two major forms, R-state and T-state [123]. The R-state has residues B1–B8 in an α-helical conformation, while the active, T-state, has residues B1–B8 in an extended conformation, which exposes the hydrophobic core region of insulin chain-B, long suspected to be important for receptor binding [124].

1.4 Protein response to stress

“Stress” in physiology is defined by the Random House Unabridged Dictionary as “a specific response by the body to a stimulus, such as fear or pain, that disturbs or interferes with the normal physiological equilibrium of an organism”. While the stimulus itself can take various forms (e.g. thermal, chemical, pressure, electromagnetic, mechanical), the end result in proteins is usually the same: a change in conformation which inevitably leads to the disruption of the proper functioning of a protein. For example, changes in protein conformation can induce the formation of amyloid fibrils [125], a basis for diseases such as Alzheimer’s disease, cystic fibrosis, and variant Creutzfeldt-Jakob disease (vCJD).

Some of the common forms of stress include:

Thermal stress

Thermal stress is normally associated with an increase in heat (thermal energy) which forces a protein to adopt conformational states that are not normally accessible to it.
Changes to protein conformation resulting from the application of prolonged thermal stress are generally irreversible, possibly because of the increase in hydrophobic interactions which tend to destabilise the protein [126]. Thermal stress simulations have been used to encourage a protein to overcome the potential energy barriers and accelerate the folding/unfolding process [60, 103, 126–128].

Experiments show that the nature of damage due to heating depends on the exposure time scale. For example, at milliseconds to submilliseconds time scale, the damage is thermal in nature [129], whereas at nanoseconds to microseconds time scale, the damage is mechanical in nature [130], with explosive evaporation and spallation being the primary cause of damage [131].

**Chemical stress**

Chemical stress occurs when a protein is subjected to a change of pH, or other chemical conditions that damage certain covalent bonds between the atoms in the protein, such as the application of Cleland Reagent (DTT, dithiotreitol) which damages the disulfide bonds in a protein [132]. The pathways of protein denaturation under thermal and chemical stresses have been found to be quite different, since early acid denaturation includes not only changes in secondary structure, but also tertiary structure as well. In addition, the position of the transition state on the reaction pathway depends on the concentration of the denaturant [133]. Gu et al. observed that the secondary structure elements of the protein, along with the loop regions become more parallel to each other as the simulation progresses. This process is accompanied by stabilisation due to non-native side chain contacts, and provided a framework for $\beta$-sheet extension, which eventually will lead to protein aggregation [128]. In contrast, no such effect was observed upon application of a thermal stress. Another example of a chemical stress is the aggregation observed in a study of insulin upon the introduction of chlorpromazine by Bhattacharyya and Das [134].

A mechanism of damage by denaturants such as urea was suggested as preferential solvation of the protein surface, reducing the hydrophobic interaction and providing enhanced competition for native hydrogen bond interactions in the protein [135, 136]. Such
a mechanism implies that the protein can undergo refolding upon removal of the denaturants from the system.

**Pressure**

It is well known that the use of high temperature results in simultaneous changes in volume and thermal energy that are difficult to separate [137]. In addition, the conformational changes due to high temperature are also generally irreversible. However, chemical stress and pressure cause reversible conformational changes in most proteins [126]. In some situations, it is preferable not to have unnecessary introduction of chemical agents. In those cases, high pressure provides a good alternative and may also aid in the interpretation of high temperature experiments [138]. High pressure has also been suggested as a useful, adjustable parameter in the experimental study of protein aggregation and misfolding [126].

High pressure denatures a protein by forcing the protein-solvent system to occupy a smaller volume. It also causes oligomeric proteins to dissociate due to a decrease in the net volume [138]. The denaturation is especially evident if the folded structure or oligomeric form of a protein has cavities, which are forced to collapse by the pressure [137]. Single chain proteins undergo denaturation at around 500 MPa (5000 atm), while oligomeric proteins undergo dissociation at around 300 MPa (3000 atm) [139]. While there have been a number of studies that utilised high pressure to induce denaturation [11, 36, 140], the application of high pressure on insulin was shown to promote circularly shaped amyloid [141]. This effect may be caused by anisotropic distribution of “void volumes” in the normal insulin fibre, which then collapses asymmetrically under high pressure. The study also raised some concern about undesirable biological activity of protein aggregates induced by high-pressure food processing [141].

**Electromagnetic stress**

The proliferation of radio-frequency electromagnetic devices has raised concerns regarding the effects of electric and magnetic radiation on human health. Electromagnetic field
exposure is ubiquitous in our everyday life—including occupational, medical, and residential exposure. Some familiar concerns, particularly with respect to children, include close proximity to power transmitters and the rapid increase in the use of mobile phones. It is generally assumed that the power levels delivered by these devices are small and have no ill effect [142]. A study conducted on rats exposed to microwaves with similar frequency to that of a mobile phone yielded no effect [143] or even decreased susceptibility to cancer growth [144]. However, there are also reported cases where the presence of weak microwave fields was enough to induce physiological changes in organisms [145, 146]. Recent studies also showed evidence of non-thermal microwave damage to brain tissues in exposed rats [147] and increased susceptibility to cancer growth in human [148] and cancer-prone rats [149].

The non-thermal effect has been suggested in recent studies where microwave radiation was shown to alter protein conformation without bulk heating [125, 150]. One of the possible explanations for this “non-thermal” effect was that the application of pulsed microwave radiation causes a rapid temperature rise that returns to the baseline temperature faster than can be detected by normal thermometry [151]. The fast temperature rise can alter protein conformation, which in turn can change the protein’s activity, causing detrimental effects.

There is an ongoing debate concerning the true nature of electromagnetic radiation effects on humans. The studies reported in the literature are difficult to interpret and are inconsistent in their conclusions (see Ahlborn, Green et al. 2004 [152] and references therein). The exact nature and extent of electromagnetic field exposure is extremely difficult to measure. Moreover, the lag-time period of studies thus far is relatively short (regular mobile phone use has only become prevalent in mid 1990s), and may not be sufficiently long to cover the period between exposure and disease manifestation. For example, one study reported at least 3-fold increase in acoustic neuroma among users of analogue mobile phone (a common source of electromagnetic radiation), regardless of the induction period [148]. In a more recent study, it was concluded that short-term exposure to mobile phones does not increase the risk of acoustic neuroma, although they indicated a
possible increase in risk for long-term exposure (at least 10 years in duration) to analogue phones [153].

The specific effects of electric and magnetic field exposure on biological systems have been the subject of a number of investigations showing conflicting results. A study by Capri et al., investigating the effect of 50-Hz sinusoidally oscillating magnetic field exposure on human peripheral blood mononuclear cells reported that magnetic fields of strengths up to 2.5 mT do not affect the activity and proliferation of the cells [154]. In contrast, in a study conducted by Selmaoui and Touitou, it was reported that the exposure of rats to 50-Hz oscillating magnetic fields of strength up to 0.1 mT yields depression in pineal serotonin-N-acetyltransferase [155]. Exposure to magnetic field was also reported to increase the risk of breast cancer in women living close to power lines by 60% [156].

Research using 60-Hz oscillating electric fields of strengths up to 130 kV/m showed severe alteration to the pineal melatonin content and serotonin-N-acetyltransferase [157] and interference in the normal expression of nocturnal melatonin levels in rats [158]. This positive result is offset by a negative result obtained by Grota et al., which did not detect any change to the activities of pineal melatonin and serotonin-N-acetyltransferase in rats after exposure to 60-Hz oscillating electric fields of strength 65 kV/m for up to 30 days [159]. However, Grota et al. found a reduction in serum melatonin level and suggested that exposure to the electric field probably stimulated degradation of tissue uptake of melatonin. Also of some concern is the relatively small number of studies investigating the effects of static magnetic fields, which occur in industrial and medical processes [160]. A review of the available literature covering the biological effects of exposure to magnetic resonance imaging did not find evidence of any health hazard, although it was suggested that there are physical mechanisms of interaction between tissues and the magnetic field [161].

**Protein response**

When subjected to electromagnetic or other stress, a protein may begin to unfold. When this happens, there are several mechanisms to prevent permanent damage to the system.
This is illustrated in Figure 1.9. If the damage is not extensive, the protein may simply self-refold [47]. This event is intrinsic to every protein and directly relates to the condition set out by the reversibility theorem. If the damage is extensive, a class of proteins called the small Heat Shock Proteins (sHSPs) will be called into action to prevent permanent damage by assisting to refold the protein [151]. The sHSPs are molecular chaperones that either assist in the refolding of the protein, or alternatively, in the degradation of the protein. The over-expression of sHSPs has been associated with increased oncogenesis and metastasis, as well as increased resistance to anticancer drugs (see French et al. 2000 [162] and the references within).

![Figure 1.9](image.png)

**Figure 1.9:** Damage probability to a protein at different stress power levels.

In terms of response to microwave irradiation, studies have shown an increase in expression of sHSPs, even with no discernible temperature rise [146, 163]. Recent work suggests the expression of sHSPs occurs in a limited range of power levels when tissues are exposed to microwaves [151]. At low power levels, sHSPs are not activated resulting in partial unfolding and disruption to protein function. At higher power levels, the sHSPs are expressed and protect the protein from conformational changes that may disrupt their function. At very high power levels, the sHSPs cannot prevent the protein unfolding. This leads to aggregation and precipitation, and therefore cell death. If the damage is irreparable, then the protein is destroyed [164].
1.5 Project goals

It is clear from the outline above that the current understanding of protein response to external stresses is incomplete. In particular, electromagnetic stress has been a subject of some controversy. This study focuses on the electric field component of the electromagnetic stress. The aim of this study is to use computer modelling to investigate the mechanistic response of a well-studied prototype protein, the insulin hormone, to the applied electric field of various strengths at a time scale not readily accessible by conventional experiments. The theory of very fast temperature excursion not detectable by normal thermometry is also explored by imposing thermal stress conditions onto the same protein. The effects of the simulations under thermal stress are then compared with the results from simulations under electric field stress. The effects of static and oscillating electric fields of different strengths and frequencies are also compared in order to understand the nature of interaction of each type of field with the protein.
Chapter 2

Computer Simulations of Proteins

The important thing in science is not so much to obtain new facts as to discover new ways of thinking about them.

Sir William Henry Bragg (1862–1942)

2.1 Overview

This chapter presents a review of history and methodologies involved in computer simulations of proteins. Section 2.2 presents an account of the development of model systems, from early conception as a physical model to the development of virtual models on a computer. A comprehensive review of the use of computer simulations in protein studies is given in Section 2.3. This review covers the early development with crude models, which are gradually refined and extended to include quantum mechanical effects. Molecular dynamics—the method used in this project—is described in great detail in Section 2.4. In Section 2.5, the simulation details used in this work are described. The method of the application of stresses to the systems is described in Section 2.6. Finally, the analysis techniques used in this work are discussed in Section 2.7.
2.2 Model system

While traditional experiments are useful in determining many properties of materials, there are cases where their use is prohibitively difficult, if not impossible. These include short time scale, short length scale, and extreme environmental conditions.

Early efforts in modelling proteins began with the development of a scale model made from hard wood connected by steel rods and clamps (1 inch = 1 Å) and plastic connected by snap fasteners (1 inch = 2 Å) by Corey and Pauling [165]. This model was based on structural data obtained from x-ray diffraction (XRD) and used correct atom proportions based on their van der Waals radii. In 1960, Koltun improved upon the original Corey-Pauling model. The resulting model was named the Corey-Pauling-Koltun (CPK) model and has proven to be very useful in visualising and making accurate structural measurements of a protein.

Nowadays, computer simulations techniques have progressed so far that we routinely carry out simulations at femtosecond time step and ångström length unit. An advantage of computer simulation over traditional experimental methods is the ability to highlight individual evolutions of a system of interest under controlled conditions. Whereas in traditional experiments, we can only obtain ensemble-averaged information, computer simulations can show the atomistic details of the system of interest. It should be noted, however, that traditional experiment is now beginning to be able to probe short time scale. For example, a recent x-ray diffraction study succeeded in obtaining data at picosecond time scale with resolution up to 1.8 Å [166].

The computer simulation techniques can be classified into several broad categories, shown in Figure 2.1. The most accurate method is the \textit{ab-initio} method, which calculates properties of a system from first principles by solving Schrödinger equations numerically. In this method, all electrons are treated explicitly and the interactions between atoms are determined solely from their electronic configurations (the atomic number) and position of the atoms. While highly accurate, this method is very expensive and can only realistically be implemented for small sized systems and very short time scales. In the semi-empirical methods, various simplifications are used, e.g. only treating the valence electrons explicitly.
and fitting several parameters to experimental data, which allows them to probe larger system sizes and longer time scales compared to \textit{ab-initio} methods. Both of these methods are used to investigate processes that involve bond breaking/formation and electronic rearrangement.

The molecular mechanics approach is used for larger system sizes and time scales. The heart of this approach is the \textit{Born-Oppenheimer approximation}, which allows the expression for the electronic movement to be eliminated from the Hamiltonian of the system, leaving only the nuclear variables. In this approximation, the microscopic state of the system can be determined using only the positions and the velocities of the atoms. The molecular mechanics approach is typically used for simulations up to hundreds of thousands of particles, with time scale up to a microsecond. The interaction between atoms is modelled using interatomic potentials (forcefield) derived from either highly accurate \textit{ab-initio} calculations or fitted to experimental data. This approach is used to obtain thermodynamic and transport properties and to study a wide range of processes such as interactions of molecules with surfaces, phase transitions, and protein dynamics. The two methods mainly implemented in these system sizes and time scales are Monte Carlo (MC) and molecular dynamics (MD).
For even larger systems, the molecular mechanics approach is further simplified by introducing restrictions to the degree of freedom and treating groups of atoms as one contiguous unit. This method is referred to as mesoscale modelling, and is very efficient in accessing quite large system sizes and long time scales, at the expense of atomic-scale accuracy. At the largest size and time scale, continuum methods enable one to simulate a wide range of dynamical processes, which can be compared directly to experiments.

2.3 Computer simulation in protein studies

2.3.1 Early techniques

With the introduction of electronic computers in the 1940s, people have attempted to develop new ways to exploit the power of automated calculation of repetitive tasks. In 1953, Metropolis et al. successfully applied the well-known MC method for calculating the equation of state of matter using computers [167]. Another big step occurred in late 1957 when Alder and Wainright introduced the MD technique which was used to study the dynamical behaviour of hard spheres [168]. Both techniques are still being used today as the mainstay techniques of computer simulation of matter.

During the time when computer simulation was just beginning to be accepted as a valid complement to experimental approach in its own right, the protein folding problem still eluded computational biochemists, due to the long simulation time scale necessary to probe the event (microseconds). Some progress has only been made within the last several years with the advent of the reversibility theorem that states proteins can unfold via the same path as they fold, but in a greatly accelerated time scale (nanoseconds) [60, 106, 169–174].

The main technique used in protein studies is MD, because of its exquisite atomistic resolution. The first time an all-atom model of a protein was used for computer simulation was in the refinement of x-ray diffraction structures of lysozyme by performing minimisation of energy as a function of the atomic coordinates [175, 176]. In 1975, Levitt and Warshel demonstrated the first application of MD to the protein folding problem by
performing thermalisation/minimisation MD cycles to bovine pancreatic trypsin inhibitor (BPTI) [23]. Due to the limitation of computer power at that time, they used simplified protein representation. This consisted of replacing each side chain with a centroid (an average of the atomic position of the atoms in the side chain), restricting the dihedral angles so that only a torsion angle along the line of two C\textalpha{} atoms was considered, and simplifying the non-bonded interactions by an effective potential. Even with these restrictions, Levitt and Warshel proved the technique could be used to find stable conformations of a folded protein. In 1976, Levitt repeated the feat using the same technique and succeeded in performing a 3 ns MD simulation of BPTI folding with accuracy within 1 Å of the true native structure [177]. In addition, the minimum energy conformation found using this approach also yielded a root mean square deviation (RMSD) less than 3 Å of the true native structure.

With the increase of computational power, it has become possible to represent each atom explicitly, including the surrounding water molecules within a simulation box [44, 47, 51, 55, 122, 178–188]. However, the simulation timeframe still rarely exceeds tens of nanoseconds, which is far shorter than the time scale necessary to probe protein folding. Before Duan and Kollman performed their groundbreaking 1 µs continuous simulation of villin headpiece subdomain in explicit water [78], the longest simulation performed in explicit water was by Daura et al., who managed to run multiple 50 ns simulations of \beta{}-heptapeptide. Even with the long simulation time, Duan and Kollman only observed the precursor to folding, not the folding process itself. This highlights the problem of a single dynamics run to study the protein folding problem, as the folding process is a heterogenous reaction involving an ensemble of transition states [189].

### 2.3.2 Ensemble dynamics

The perceived weakness of a single long dynamics run lead to the introduction of the ensemble dynamics, where many MD simulations are performed on the same system. Even though each run only lasts for a few nanoseconds, the sheer number of parallel simulations performed means the combined simulation time can reach up to hundreds of microseconds.
The crux of the ensemble dynamics method is that complex phenomena such as protein folding have a complex potential energy surface with many local minima. To escape from these potential minima, the system must rely on receiving energy from thermal fluctuation. A system at high temperature therefore will be able to cross the potential barrier easily, whereas a system at lower temperature will have a much lower, non-zero probability of doing so. In ensemble dynamics, multiple independent simulations are performed, starting from one potential well. When one simulation crosses the potential barrier, all of the other simulations are transferred to the new potential well and restarted. This process is repeated as many times as needed. By running multiple coupled simulations to explore the potential energy surface, ensemble dynamics can achieve linear speedup scaling according to the number of processors [190, 191]. The power of this method is illustrated by the folding simulation of the BBA5 mutant in implicit solvent performed by Snow et al. [174]. In that work, the authors ran thousands of simulations totaling 700 µs, each one lasting only 5–20 ns. Even though observing a folding event in a single 20-ns run is improbable, given that a small two-state protein obeys exponential kinetics with a mean folding time of 10 µs, folded conformation should be observed in roughly 10 out of 10 000 simulations after 10 ns. As a comparison, if the ensemble dynamics were performed as a single run on a typical modern CPU, it would have taken decades to complete a 10 µs simulation. Comparison with experimental data based on circular dichroism and fluorescence spectroscopy indicated that the results of the ensemble dynamics simulations agreed with them very well. Using the same method, Zagovic et al. also obtained not one, but eight complete and independent folding pathways of a β-hairpin protein. The results are also consistent with the existing experimental data [191].

A related technique to the ensemble dynamics is the replica exchange molecular dynamics (REMD) [192, 193]. In this method, a number of concurrent, non-interacting simulations (“replicas”) are performed at different temperatures. Every so often, the atomic coordinates from these replicas are exchanged with each other. In effect, REMD enables significantly enhanced sampling of the conformational space by providing a greater chance for a system at low temperature to cross energy barriers by being exchanged with
CHAPTER 2. COMPUTER SIMULATIONS OF PROTEINS

a replica at a higher temperature. This method has been used extensively with very good agreement with experimental data, suggesting that REMD can be used for structure prediction as well as studying the protein folding and dynamics problem [192–194]. In addition, REMD also has the advantage of improved convergence compared to classical MD [195].

2.3.3 Restrained dynamics

Another method that is often employed in protein folding studies is based on restraining the protein dynamic pathway using geometrical data obtained from experimental methods, such as the modification of the potential energy function using interproton distance restraints obtained from nuclear magnetic resonance experiments [51]. This method was used to observe rapid folding (within 9 ps) of crambin to the correct tertiary structure. Even with the somewhat artificial nature of the pathway, it was able to provide insight into the protein folding process. Specifically, correctly folded structure is achieved only when the secondary structure elements are at least partly formed before the tertiary structure elements are introduced [51]. This may be analogous to the protein folding mechanism suggested in the diffusion-collision model [100].

A closely related simulation technique to the restrained molecular dynamics is the targeted molecular dynamics. In this method, the starting and target structures are known, and the simulation progresses by utilising geometrical constraints which essentially simplify the configurational space search to sample only pathways accessible at a given temperature [50, 196]. As such, the targeted molecular dynamics method has been used to study the conformational transition pathways between two forms of a protein. An example of this application can be found in the study of the transition between the T ↔ R states, which happens cooperatively in all subunits of insulin hexamer during storage, but rarely happens in a monomer [197, 198]. Targeted molecular dynamics was shown to yield several pathways of the transition due to the relative freedom of the system. In contrast to other constrained MD methods [199, 200], no information of the preceding pathway calculations [49] were included. Recently, the stability of targeted molecular dynamics has
been criticised due to the pathways’ dependence on simulation direction and its sensitivity to perturbation of the initial conditions [201].

2.3.4 Simplified model

Full atomistic modelling of proteins is not the only method of simulating the dynamics of proteins. Studies have also been performed with simplified representations of proteins. Although this type of simulation cannot give any dynamical information of proteins, it can give some insights into the folding mechanism and sequence-dependent stability of amino acid chains. Most of these studies involve the use of lattice Monte Carlo, where a protein is modelled by beads connected by fixed chains. The simplest of lattice models has the beads arranged in a cubic lattice [59, 202, 203]. The first known simulation of proteins on a lattice was performed by Taketomi et al., using a two-dimensional lattice system by the Metropolis Monte Carlo method, and suggested that non-local contacts are important in generating the native state [204]. Later, Anfinsen and Scheraga argued that the folded structure is principally determined by local interactions, with the non-local contacts merely providing non-specific stability to the compact state [205].

The lattice Monte Carlo technique has also been used successfully to study the process of folding at a tractable time scale. For example, if the folding/unfolding simulation that Dinner and Karplus conducted on a 125-residue protein [59] was performed using classical MD, it would have taken around 100 years to simulate with the then available computing power [179]. Even with the simplified model, the authors could still obtain valuable information about the nature of folding. It was found that structures contributing to the unfolding pathways resemble those in fast folding events, rather than slow folding events.

In the lattice models, the beads are arranged in a cubic lattice, which restricts the angles between bonds to either 90° or 180°. The off-lattice models eliminate this restriction, while maintaining the simplified bead model of the residues. This technique also yields good agreement with experiment, with RMSDs of many proteins including cytochrome c, α-lactalbumin, and barnase found to agree well with x-ray crystallographic experiments [206]. The stability of the α-helix was also reproduced by the same author, justi-
fying the use of simplified models in protein folding studies [206]. Furthermore, dynamic and equilibrium properties of an apomyoglobin model were obtained within computational times an order of magnitude shorter than classical MD, with a very good agreement with MD and experimental NMR results [207].

2.3.5 Quantum mechanical techniques

While molecular mechanics methods such as MC and MD have proven to be very useful to study protein dynamics and to probe protein folding/unfolding, they cannot be used to study processes that involve changes at the electronic level. In order to study these processes, a quantum mechanical approach must be used. With this method, properties such as energies, vibrational frequencies, magnetic shielding, and thermochemical properties can be calculated with near chemical accuracy (\(\sim 1–2\) kcal/mol). Despite these advantages, the use of a pure quantum mechanical approach is still limited due to the exponential complexity in modelling even the smallest protein. Indeed, most of the application of quantum mechanical protein simulation involves a hybrid quantum mechanics/molecular mechanics technique, where only the region of interest is simulated using the full quantum mechanical treatment, with the rest simulated using molecular mechanics (see review by Monard and Merz and the references within [208]). Even with this limitation, the studies conducted so far have shown a lot of potential. The technique was first introduced by Field et al. in 1990, who obtained satisfactory results in agreement with experimental data [209]. Another study utilising the quantum mechanics/molecular mechanics technique on a protein yielded better depiction of active site geometry and dynamics in zinc metalloprotein human carbonic anhydrase than molecular mechanics alone [210]. The approach has also been used to study the reaction pathway of intermediate formation in acetylcholinesterase enzyme [211] and the catalytic mechanism of aldose reductase [212] with considerable success. More recently, a pure quantum mechanical approach was used to obtain thermochemical properties from closed-shell and open-shell forms of peptides, such as those involved in abnormal aggregation and amyloid formation [213] and to provide insights into conformational preference and kinetics of protein folding [214].
is no doubt in the power of quantum mechanics in protein studies. Efforts are now under way to optimise the technique to approach linear scaling and refine the hybrid method by including a polarised atom model in the molecular mechanics component [208, 215].

2.4 Molecular dynamics

One of the most important and widely used molecular simulation techniques used today is MD. This method was created subsequent to MC in order to study the dynamical behaviour of many-particle systems. The first successful application was performed by Alder and Wainright [168, 216] on a system of hard spheres, as a model for atomic-scale systems. Ever since then, it has been applied to a variety of other systems, including polymers and biological materials [48, 178, 179].

While MC is used in lattice simulation studies of protein folding [48, 59], MD is preferred in the study of the dynamical behaviour of proteins. This is due to the long correlation time caused by the inertia force term, which gives MD a better sampling efficiency compared to MC [217]. In contrast, MC was shown to have almost no correlation in the motion of the particles due to the randomness in each step, and is thus incompatible with the collective nature of protein motion arising from the covalent bonding between the atoms.

2.4.1 Equations of motion

The heart of the molecular mechanics method is the separation of nuclear coordinates from electron coordinates which are subsequently averaged out, by way of the Born-Oppenheimer approximation [218]. The separation is possible because of the large difference in mass between the nuclei and the electrons. By using the approximation, it is possible to express the Hamiltonian of the system as a function of nuclear variables. The complete microscopic state of the system can then be described only by the positions and velocities of the atoms, as shown in Equation 2.1:
\[ H(q, p) = K(p) + V(q) \]  \hspace{1cm} (2.1)

where \( H \) is the Hamiltonian, \( K(p) \) is the kinetic energy as a function of momenta, and \( V(q) \) is the potential energy as a function of generalised coordinates. The kinetic energy term contains the momenta of each atom, while the potential energy term contains the information of the interatomic interaction.

Provided the potential energy term is well behaved (i.e. the function and its first derivative exist and are continuous), it is possible to construct an equation of motion from the Hamiltonian. In MD studies, the equation of motion is usually constructed in Newtonian form:

\[ \frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} \]  \hspace{1cm} (2.2)

where \( F_i \) is the total force experienced by each atom \( i \) which acts in the direction \( r \), \( m_i \) is the mass of atom \( i \), and \( r_i \) is the position of atom \( i \).

The basic idea of MD is that given the positions and velocities at time \( t \), we want to find the positions and velocities at time \( t + \delta t \), where \( \delta t \) is the time interval (time step) between two simulation snapshots. The time step is chosen such that it is smaller than the fastest vibration in the system, but large enough to avoid unnecessary and expensive computation. Typically for an organic molecule, this would be in the order of 1 fs, which corresponds to \( \frac{1}{10} \) of the C–H bond vibration period, the fastest vibration in the system. However, methods exist to increase the time step to up to 8 fs at the expense of reduced accuracy at long simulation time. These involve fixing certain bonds, such as the hydrogen bonds, thereby reducing the number of degrees of freedom in the system [219].

In order to find the positions and velocities at time \( t + \delta t \) we need to solve the equation of motion given in Equation 2.2 by an integration scheme. While there are many techniques available, the most commonly used one is the Verlet algorithm [220], which requires the knowledge of current position, \( r(t) \); acceleration, \( a(t) \); and position from the previous step, \( r(t - \delta t) \). The position of the next step can then be found using the formula:
The velocities of the current step can be calculated by using $\mathbf{r}(t - \delta t)$ and $\mathbf{r}(t + \delta t)$ to perform a Taylor series expansion about $\mathbf{r}(t)$ to give:

$$
\mathbf{v}(t) = \frac{\mathbf{r}(t + \delta t) - \mathbf{r}(t - \delta t)}{2\delta t} \tag{2.4}
$$

While the positions found using Equation 2.3 are correct within errors of the order of $\delta t^4$, the velocities found using Equation 2.4 are only correct within errors of the order of $\delta t^2$. To alleviate this problem, improvements upon the original Verlet algorithm were made over the years. Most notable is the velocity Verlet algorithm, which offers improved accuracy and energy conservation [221]. In this scheme, the positions and velocities are obtained from the formulae:

$$
\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t) + \frac{1}{2} \delta t^2 \mathbf{a}(t) \tag{2.5}
$$

$$
\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \frac{1}{2} \delta t [\mathbf{a}(t) + \mathbf{a}(t + \delta t)] \tag{2.6}
$$

with the velocity mid-step computed using:

$$
\mathbf{v}(t + \frac{1}{2} \delta t) = \mathbf{v}(t) + \frac{1}{2} \delta t \mathbf{a}(t) \tag{2.7}
$$

After calculating the forces and accelerations at time $t + \delta t$, the velocities at time $t + \delta t$ can be found using:

$$
\mathbf{v}(t + \delta t) = \mathbf{v}(t + \frac{1}{2} \delta t) + \frac{1}{2} \delta t \mathbf{a}(t + \delta t) \tag{2.8}
$$
2.4.2 Potential energy function

The forces present in a MD simulation can arise from internal and external sources. The internal sources are in the form of interaction between bonded and non-bonded atoms. The external sources are environmental conditions imposed on a system, e.g. an electric field.

To calculate the forces, the derivative of the potential energy function—often known as the forcefield—is obtained (Equation 2.9). This potential energy function approximates $n$-body interactions of the particles, with the contributions arising from the external field, pair interactions, and triplet interactions. However, calculating 3-body interaction is very complex. Therefore, most early forcefields ignored its contribution altogether, opting for effective pair potential instead, which gives a very good description of liquid properties [222, 223]:

$$\mathbf{F}_i = -\nabla E_{pot}$$

where $\mathbf{F}_i$ is the force acting on a particle $i$, and $E_{pot}$ is the potential energy function, given by the equation:

$$E_{pot} \approx \sum_i v_1(r_i) + \sum_i \sum_{j>i} v_2^{eff}(r_{ij})$$

where the $v_1(r_i)$ is the contribution from the external field, $v_2^{eff}(r_{ij})$ is the contribution from pair interaction, and the notation $\sum_i \sum_{j>i}$ denotes a summation over all distinct pairs $i$ and $j$ without counting any pair twice (i.e. $ij$ and $ji$).

The simplest potential energy function is the hard sphere potential, which is defined as:

$$E = \begin{cases} 
0 & r \leq a \\
\infty & r > a 
\end{cases}$$

where $a$ is the hard sphere radius. This potential has no attractive nor repulsive part. A particle will travel in a straight line until it hits another particle, whereupon it undergoes elastic scattering.
While the hard sphere potential was successfully used to obtain many thermodynamic properties of liquid systems, it is not realistic. It was then superseded by potentials such as the Lennard-Jones potential, which has an attractive and repulsive component [224, 225]. The most commonly used Lennard-Jones potential has an exponent of 12 for the repulsive part, and 6 for the attractive part:

\[
E = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right]
\]  

(2.12)

where \( \epsilon \) is the well depth and \( \sigma \) is the hard sphere radius.

In order to simulate a more complicated system—such as proteins—a more sophisticated equation is needed to account for different degrees of freedom associated with it. Class I forcefields are forcefields that only incorporate diagonal terms and non-bonded interactions. Some examples of these forcefields are CHARMM [226], AMBER [227], and GROMOS [228]. The potential energy function in a Class I forcefield takes the form of:

\[
E_{\text{diagonal}} = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{torsions}} K_\phi[1 + \cos(n\phi - \delta)]
\]

(2.13)

\[
E_{\text{non-bonded}} = \sum_{\text{electrostatic,ij}} \frac{q_i q_j}{4\pi D r_{ij}} + \sum_{\text{van der Waals,ij}} \epsilon_{ij} \left[ \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right]
\]

(2.14)

where \( b \) is the bond length, \( \theta \) is the bond angle, \( n \) is dihedral multiplicity, \( \phi \) is the dihedral angle, \( \delta \) is the dihedral phase, \( \psi \) is the improper (out-of-plane) angle, \( r_{UB} \) is the Urey-Bradley term (1,3 non-bonded interaction); subscript 0 denotes equilibrium value of a property; \( K_b, K_\theta, K_\phi, K_\psi, K_{UB} \) are the force constants for the bond length, bond angles, dihedrals, impropers, and Urey-Bradley terms respectively; \( q \) is the charge of a particle, \( D \) is the dielectric constant of a material, \( r_{ij} \) is the distance between the centres of particles.
i and j, \( \epsilon_{ij} \) is the minimum value of the van der Waals term, and \( R_{\text{min}} \) is the radius where the van der Waals term is at the minimum.

Class II forcefields are natural extensions to the Class I forcefields. Class II forcefields incorporate cross terms into the potential energy function shown in Equations 2.13 and 2.14. Examples of the Class II forcefields include CFF [229], MM3 [230], and MMFF94 [231]. The cross terms that appear in Class II forcefields include the bonds-bonds (\( E_{bb'} \)), angles-angles (\( E_{aa'} \)), bonds-angles (\( E_{ba} \)), bonds-dihedrals (\( E_{bd}, E_{b'd} \)), angles-dihedrals (\( E_{ad} \)), and angles-angles-dihedrals (\( E_{aad} \)). These terms account for the cooperative effects in the system, e.g. relaxation in one bond length upon relaxation in the opposite direction, and the distortion to bond angles upon stretching of the bonds. These cross terms are illustrated by the following equations:

\[
E_{bb'} = \sum_{\text{bonds}} \sum_{\text{bonds}'} K_{bb'} (b - b_0)(b' - b_0')
\]

\[
E_{aa'} = \sum_{\text{angles}} \sum_{\text{angles}'} K_{\theta\theta'} (\theta - \theta_0)(\theta' - \theta_0')
\]

\[
E_{ba} = \sum_{\text{bonds}} \sum_{\text{angles}} K_{b\theta} (b - b_0)(\theta - \theta_0)
\]

\[
E_{bd} = \sum_{\text{bonds}} \sum_{\text{dihedrals}} (b - b_0)[K_{\phi,b1} \cos \phi + K_{\phi,b2} \cos 2\phi + K_{\phi,b3} \cos 3\phi]
\]

\[
E_{b'd} = \sum_{\text{bonds}'} \sum_{\text{dihedrals}} (b' - b_0')[K_{\phi,b'1} \cos \phi + K_{\phi,b'2} \cos 2\phi + K_{\phi,b'3} \cos 3\phi]
\]

\[
E_{ad} = \sum_{\text{angles}} \sum_{\text{dihedrals}} (\theta - \theta_0)[K_{\phi,\theta1} \cos \phi + K_{\phi,\theta2} \cos 2\phi + K_{\phi,\theta3} \cos 3\phi]
\]

\[
E_{aad} = \sum_{\text{angles}} \sum_{\text{angles}'} \sum_{\text{dihedrals}} (\theta - \theta_0)(\theta' - \theta_0')[\cos \phi]
\]
2.4.3 Periodic boundary conditions and minimum image convention

Particles that lie on the surface experience different forces compared to particles that lie in the bulk (see Figure 2.2). The surface effect is especially evident in systems that contain a small number of atoms. For example, a simulation box containing 1000 particles arranged in a $10 \times 10 \times 10$ lattice has approximately 488 particles lying on the surface. The forces experienced by these particles influence the forces experienced by the rest of the system, yielding unwanted effects for simulations of bulk systems.

![Figure 2.2: Different forces experienced by surface and bulk atoms.](figure)

In order to properly simulate a system in bulk liquid (e.g. a protein in aqueous environment), it is necessary to have “infinite” extent in three dimensions. This can be achieved simply by directly increasing the system size. However, the increase in system size is not practical due to the computational cost involved and the fact that it does not completely eliminate the surface effects. *Periodic boundary conditions* (PBC) provide a way to completely eliminate the surface effects without increasing the system size [232].

In the PBC implementation, the central simulation box is replicated in three dimensions to form an infinite lattice. Figure 2.3 illustrates the PBC in action. As the blue particle leaves the central simulation cell from the lefthand wall, it is replaced by a replica entering the central box from the righthand wall. This process occurs within all of the simulation cells simultaneously and ensures that there is always a constant number of particles within any of the simulation boxes.
CHAPTER 2. COMPUTER SIMULATIONS OF PROTEINS

Figure 2.3: Two dimensional illustration of periodic boundary conditions. The central box highlighted in orange is the main simulation box.

Figure 2.4: Two dimensional illustration of the minimum image convention. The dotted box represents all the particles that can be considered in the potential calculations on the yellow particle. The blue circle represents the use of potential cutoff. Only particles located within this circle experience interatomic interaction with the yellow particle.
In order to calculate the correct forces acting on a particle, the contribution from all other particles and their replicas must be considered. In practice, this is impossible to do because of the infinite number of particles involved. There are two approximations that are used to solve this problem. The first approximation is the minimum image convention \[167\]. Consider a region with the same size and shape as the simulation box with the particle of interest at the centre. This region will often overlap between the central box and several replicas. The particle will interact with every other particle that lies within this region. The second approximation is the potential truncation. In this approximation, a spherical potential cutoff is employed around the particle of interest. This is because the greatest contribution to the potential and forces comes from neighbouring particles close to the particle of interest. This introduces a small perturbation to the potential and force calculations, and therefore a suitable potential cutoff must be chosen for a given simulation. The information lost due to the perturbation can be recovered by using long range corrections such as the Ewald summation \[233–235\]. Figure 2.4 illustrates the application of minimum image convention and the potential cutoff.

## 2.5 Methodologies

The common methodologies of the simulations used in this project are discussed in this section, with specific parameters discussed in their respective chapters. The major technique used in this project is classical MD \[236, 237\], as implemented in the program NAMD \[238\]. The use of MD allows us to explore a range of conformations that occur during the protein dynamics in full atomistic details. Periodic boundary condition in the form of a cubic cell was employed to provide the necessary bulk conditions to simulate a protein in solution \[236\].

The explicit water model, TIP3P \[239\] was used to fill the periodic box with solvent density of 1.0 g/cm\(^3\). The TIP3P model was chosen as it is well known and widely used in protein simulation studies \[183\]. The box size and water molecules extend to at least
15 Å in all directions from the protein, creating a cubic box sized at least 60 Å × 60 Å × 60 Å. This ensures the protein doesn’t experience self interaction.

The well-validated all-atom empirical forcefield, CHARMM27 [240] was used in this project. This forcefield represents all atoms in the system explicitly, and assigns the interaction potential to each atom based on the individual local atomic environment. A spherical residue-based potential truncation at 12 Å was used for the non-bonded interactions, with a linearly decaying switching function turned on between 10 Å and 12 Å, ensuring the potential remained continuous. Figure 2.5 illustrates the effect of truncation and switching function on the interaction potential. A neighbour list was maintained for atoms within 14 Å. This provides a good compromise between accuracy and speed of computation while still maintaining the long range interaction necessary for the folding/unfolding processes involving helices [58]. In order to account for the long range electrostatics, the Ewald summation was employed in the form of the Particle Mesh Ewald (PME) method [233, 235]. Without the use of the Ewald summation, artifacts such as increased water/vapour surface tensions and increased orientational polarisation decay parameters will arise [241].

![Figure 2.5: The effect of potential truncation and switching function to a typical non-bonded interaction potential.](image)
CHAPTER 2. COMPUTER SIMULATIONS OF PROTEINS

The SHAKE algorithm [242] was employed in the simulations to constrain the bond lengths to their energy minimised values. This allowed the vibrational modes of hydrogen bonds to be ignored, and a time step of 2 fs to be used. The bond constraint enables us to simulate up to 20 ns of protein dynamics to be performed within a reasonable computational time.

The starting structure coordinates were obtained from the Protein Data Bank (PDB, http://www.rcsb.org/pdb/) [243]. Most of the structures in the database were obtained using x-ray diffraction (XRD) and Nuclear Magnetic Resonance (NMR) techniques, ensuring the accuracy of the coordinates to at least 2.5 Å resolution. This work utilised the coordinates of porcine insulin (PDB accession code 1ZNI [244]) resolved using x-ray diffraction to 1.5 Å resolution.

All the simulations were performed on APAC (Australian Partnership for Advanced Computing) and VPAC (Victorian Partnership for Advanced Computing) machines.

Simulation stages

After obtaining the coordinates from PDB, the water molecules identified by experiment were removed, leaving just the protein molecule. The complete protein or chains of interest were isolated and put into a simulation box and immersed in water using the InsightII program [245]. A typical system setup is shown in Figure 2.6, and the outline of the simulation stages is given in Figure 2.7.

In the energy minimisation stage, the initial coordinates of the systems were optimised for 20 000 steps by way of conjugate gradient and line search algorithms with convergence criterion of $10^{-4}$ kcal/(mol·Å). This allows the system to relax and avoid steric clashes between the protein and water atoms. In this study, the protein and water were not minimised separately.

The minimisation stage was then followed by two equilibration stages, each lasting 600 ps. The first of these stages was performed in the constant volume (NVT) ensemble, followed by the second equilibration in the constant pressure (NPT) ensemble [246]. The target pressure was set to 1 bar, representing ambient atmospheric pressure. In order to
achieve this, the Nosé-Hoover algorithm was employed, in which fluctuations in the baro-
stat are controlled by Langevin dynamics [247], as implemented in NAMD. The temperature 
of the system was initially set by assigning random velocities according to a Maxwell dis-
tribution corresponding to the target temperature (300 K or 400 K for the thermal stress 
simulations) and controlled using velocity rescaling method. The equilibration stage was 
checked by monitoring the energy, temperature, and pressure fluctuations.

After the equilibration was reached, the data collection stage of up to 20 ns was 
performed in NPT ensemble using the same pressure control method as in the NPT 
equilibration stage. The temperature during the data collection stage was controlled by 
the Berendsen thermal bath coupling method [248].

Throughout the data collection stage, the atomic coordinates were saved every 2500 
steps (5 ps) for analysis. This procedure yielded up to 4000 frames for analysis.
## CHAPTER 2. COMPUTER SIMULATIONS OF PROTEINS

### 2.6 Application of stresses

During the data collection stage, various transient stresses were applied, i.e. thermal and electric field. The chemical stress was applied separately in the form of chain-B isolation from chain-A. The application modes of these stresses are described below.

<table>
<thead>
<tr>
<th>Stage 1</th>
<th>Minimisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjugate gradient and line search algorithms</td>
<td></td>
</tr>
<tr>
<td>20 000 steps</td>
<td></td>
</tr>
<tr>
<td>Convergence criterion of $10^4$ kcal/(mol·Å)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 2</th>
<th>Equilibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVT ensemble</td>
<td></td>
</tr>
<tr>
<td>600 ps</td>
<td></td>
</tr>
<tr>
<td>$T = 300$ K/400 K (velocity rescaling)</td>
<td></td>
</tr>
<tr>
<td>Coordinates saved every 5 ps</td>
<td></td>
</tr>
<tr>
<td>NPT ensemble</td>
<td></td>
</tr>
<tr>
<td>600 ps</td>
<td></td>
</tr>
<tr>
<td>$T = 300$ K/400 K (velocity rescaling)</td>
<td></td>
</tr>
<tr>
<td>$P = 1$ bar (Nosé-Hoover algorithm)</td>
<td></td>
</tr>
<tr>
<td>Coordinates saved every 5 ps</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 3</th>
<th>Data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambient conditions</td>
<td></td>
</tr>
<tr>
<td>10-20 ns</td>
<td></td>
</tr>
<tr>
<td>$T = 300$ K (heat bath coupling)</td>
<td></td>
</tr>
<tr>
<td>$P = 1$ bar (Nosé-Hoover algorithm)</td>
<td></td>
</tr>
<tr>
<td>Coordinates saved every 5 ps</td>
<td></td>
</tr>
<tr>
<td>Thermal stress</td>
<td></td>
</tr>
<tr>
<td>10 ns</td>
<td></td>
</tr>
<tr>
<td>$T = 400$ K (heat bath coupling)</td>
<td></td>
</tr>
<tr>
<td>$P = 1$ bar (Nosé-Hoover algorithm)</td>
<td></td>
</tr>
<tr>
<td>Coordinates saved every 5 ps</td>
<td></td>
</tr>
<tr>
<td>Electric field stress</td>
<td></td>
</tr>
<tr>
<td>0.6-10 ns</td>
<td></td>
</tr>
<tr>
<td>$T = 300$ K (heat bath coupling)</td>
<td></td>
</tr>
<tr>
<td>$P = 1$ bar (Nosé-Hoover algorithm)</td>
<td></td>
</tr>
<tr>
<td>$E = 10^{-10}$ V/m</td>
<td></td>
</tr>
<tr>
<td>Coordinates saved every 5 ps</td>
<td></td>
</tr>
<tr>
<td>$f = 0$ (static), 2.45 GHz, 1.225 GHz, 4.9 GHz</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 2.7:** Summary of simulation stages used in this study.
CHAPTER 2. COMPUTER SIMULATIONS OF PROTEINS

Chemical stress

The chemical stress has been modelled to simulate the suggested athermal effect of radiation [150, 151]. In this study, the chemical stress was simulated by removing the two disulfide bonds that hold insulin chains-A and B together (A7–B7 and A20–B19). This allowed the individual chains to move with less restriction compared to the bound monomer. In this study, the behaviour of chain-B in a monomeric insulin is compared with the behaviour of the chemically stressed, isolated chain-B.

Thermal stress

Thermal stress was applied to represent the short-lived thermal stress that might arise in response to microwave irradiation from localised, very fast heating cycles within nanosecond time scale, which is undetectable by normal thermometry as suggested by Laurence et al. [151]. More recently, these very slight temperature rises have been successfully recorded in experimental studies [249].

The thermal stress was simulated by running the MD simulations at 400 K instead of the ambient temperature of 300 K. This method proved to be effective in allowing the protein to adopt conformations not accessible by the simulation at 300 K, within the limitation of simulation timeframe of up to 20 ns.

Electric field stress

Electric field stress was applied in order to model the effects related to electromagnetic radiation. The electric field stress was simulated by placing on each atom in the system a force, $F$, proportional to its partial charge, $q$, according to the formula $F = E \times q$, where $E$ is the strength of the electric field. The partial charges were assigned from the CHARMM27 forcefield. Due to the limitation of NAMD, it was not possible to simulate more advanced effects such as charge shielding by neighbouring atoms, induced dipole moment, or polarisation due to the external field. In reality, an oscillating electric field induces a magnetic field. However, since organic materials present in protein simulations
are essentially non-magnetic (i.e. they have magnetic permeability close to that of free air [250]), this study concentrates on the electric field.

The electric field strengths in this work range from 1806 V/m to $10^9$ V/m. The value of 1806 V/m corresponds to the equivalent electric field strength of the maximum Specific Absorption Rate (SAR) value used in a study that detected some effects of 2.45 GHz microwave radiation on protein conformation [249]. The highest value corresponds to the electric field strength in a previous MD study where an observable change in the water structure was detected [251].

In this project, static and oscillating fields are studied. The frequencies used in the oscillating electric field studies were 2.45 GHz (corresponding to the frequency used in the previous study [249]), 1.225 GHz (half the experimental frequency), and 4.9 GHz (double the experimental frequency). The variations of frequency were tested in an attempt to quantify the possible frequency effects on the protein response to the electric field.

### 2.7 Analysis techniques

A variety of analysis techniques were used to quantify the structural and conformational changes experienced by the model protein under stress. The analysis procedures detailed below were performed using VMD [252], PEPCAT [253], X-PLOR [254], and InsightII [245]

**Radius of gyration**

Radius of gyration ($r_g$) is a quantity that measures the distribution of atoms relative to their centre of mass. It can be used as a measure of “size” and shape of a protein and can be related to the protein’s response to the applied field by comparing it to its normal behaviour. For an unweighted value, it is calculated using the formula:

$$r_g = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} |\mathbf{r}(i) - \mathbf{r}_{com}|^2}$$  \hspace{1cm} (2.15)
where $N$ is the number of atoms, $r(i)$ are the coordinates of atom $i$, and $r_{\text{com}}$ are the coordinates of the protein’s centre of mass.

**Root mean square deviation**

Root mean square deviation (RMSD, $\langle R \rangle$) of a protein gives a quantitative measure of structural variation as the simulation progresses and can be used as a useful measure of stability of the simulation. It is usually performed by comparing the structure of the protein at a particular simulation timeframe with a reference structure. For a well-ordered protein, this is typically around 0.5–1 Å relative to the crystal structure. The RMSD can be obtained using the following formula:

\[
\langle R \rangle = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |r_{\text{final}}(i) - r_{\text{initial}}(i)|^2}
\]

where $N$ is the number of atoms, $r_{\text{final}}(i)$ are the final coordinates of atom $i$, $r_{\text{initial}}(i)$ are the initial coordinates of atom $i$, and $t$ is time.

**Dipole moment**

It is well known that $\alpha$-helices possess dipole moments along their axes, pointing from the C-terminus towards the N-terminus [2, 255]. The helix dipole moment can contribute to the binding of charged substrates or coenzymes at helix termini, long range attraction of charged molecules, and the enhancement of reaction rates [2].

In our analysis, we have only considered the dipole moment component in one direction corresponding to the application of the electric field (parallel to the $z$-axis), illustrated in Figure 2.8. The plot of dipole moment in the $z$-direction over time, therefore, can be used to determine the degree of coupling between the direction of the electric field and the helical region of the protein. The dipole moments are calculated using the following formula:
\[ Q_0 = \sum_{i=1}^{N} q(i) \]  

(2.17)

\[ \rho(z) = \sum_{i=1}^{N} (q(i) - Q_0) \cdot r_z(i) \]  

(2.18)

where \( Q_0 \) is the total charge of the system, \( N \) is the number of atoms, \( q(i) \) is the partial charge of atom \( i \), \( r_z(i) \) is the \( z \)-coordinate of atom \( i \) with the coordinates of the protein’s centre of mass subtracted.

**Figure 2.8:** Schematic diagram of the dipole moment that arises from helical regions of a protein, and the \( z \)-direction component that is analysed in this work.

**Hydrogen bonding**

Hydrogen bonding is an important factor in stabilising the secondary structure of a protein. The loss of internal hydrogen bonding, either by direct competition with solvent or due to the applied stress, can cause elements of the secondary structure to destabilise and result in protein unfolding. Figure 2.9 illustrates the schematic of the hydrogen bonding criteria adopted in this study. In order to be classified as a hydrogen bond, the angle between the donor-hydrogen-acceptor atoms (\( \theta \)) must be within 150°–180°, and the donor-acceptor atoms separation (\( r \)) must be no more than 3.5 Å. The hydrogen bonding in the simulated systems was monitored as a function of time.
Secondary structure analysis

In order to simplify the analysis of dynamical tertiary conformational changes of a protein, it is useful to concentrate on the changes occurring to the secondary structure. This is achieved by assigning a secondary structure type to each residue based on certain criteria. The most commonly used technique is the STRIDE algorithm [256], implemented in programs such as VMD [252]. This algorithm uses the hydrogen bond energy and distances, in addition to the main chain dihedral angles.

The information obtained from observing how the secondary structures change over time is invaluable to investigate the key events in the protein’s conformation change. Important conformational transitions such as the helix breaking can easily be observed by tracking the changes in secondary structure over time.
An example of the secondary structure evolution is presented in Figure 2.10, as obtained from the STRIDE algorithm utilised by VMD. In conjunction with visual analysis, the map reveals the exact timeframe of the helix splitting event which transforms the R-state into an active, T-state [6, 7].

**Solvent accessible surface area**

The solvent accessible surface area (SASA) [87] is the area sampled by the centre of a probe tracing a protein’s van der Waals surface. The probe can have different radii, depending on the choice of solvent. In this work, the probe chosen has a radius of 1.4 Å, approximating the “radius” of a water molecule. The solvent accessible area indicates the area available for a direct contact between protein and solvent or similarly sized molecule. Under physiological conditions, the hydrophobic residues of a protein will arrange themselves so as to minimise their exposure to water. However, this is not necessarily the case when the protein is undergoing a conformational change or is under stress. A significant change in solvent accessible area therefore can be used a measure of the effect of stress on the protein.

![Solvent accessible surface area](image)

**Figure 2.11:** Schematic diagram of solvent accessible surface area calculation. The solvent accessible surface area is the area traced by the centre of a probe on the van der Waals surface of a protein.

Figure 2.11 shows a schematic diagram of SASA calculation, and Figure 2.12 illustrates the calculated SASA of insulin chain-B.
Figure 2.12: Solvent accessible surface area of insulin chain-B. The protein’s van der Waals surface is shown in yellow. A probe of radius 1.4 Å was used to generate the solvent accessible surface area, shown in red.

PEPCAT analysis

Throughout an MD simulation, accessed states of the system are stored in the form of simulation frames. Each of these frames contains the atomic coordinates of each particle and serves as a snapshot of the system at any particular time. At the end of the simulation, these frames are collated into a trajectory file that serves as a complete record of the states accessed during the simulation. From the trajectory file, it is possible to calculate structural and dynamical properties of a system as statistical ensemble averages.

Conformational analysis techniques rely heavily on utilising the structural properties in order to perform the classification of system state. One structural property often used for this purpose is the RMSD of the backbone atoms, which gives a good measure of structural similarity, and is insensitive to small changes in the atomic coordinates. However, in some cases a change in a single backbone dihedral angle can lead to a large overall RMSD value. For example, only one backbone dihedral angle change in the middle of a peptide with an extended β-strand conformation can lead to a large change in the RMSD due to the quite significant change in the overall shape of the peptide and the vast number of atoms that have experienced dramatic changes in their atomic coordinates. However, it gives no indication that it is a comparatively simple structural change (rotation of a single dihedral bond angle) that is responsible for the difference, and that the two substructures on either
side of the changed residue are still identical to the original peptide structure [253].

Other methods apart from RMSD have also been used to group peptide structures based on their geometric similarities. These methods include calculation of residue-residue contact map [257, 258], where two residues are considered in a contact if any of the atoms in the first residue are within a certain distance of any of the atoms in the second residue; and using sequences of dihedral angles from several residues as a “fingerprint” of a conformation [259]. Most of these conformational analysis techniques have predefined classification criteria which cannot be changed easily, and cannot be used in conjunction with each other. As a result, while one method may excel in defining the conformations in a ligand docking problem, it may perform poorly in defining conformations in a protein flexibility study.

These techniques do not provide information concerning the specific conformation and their frequency of occurrence throughout the simulation. The structures produced from the simulation need to be clustered and classified according to specific geometric descriptors. PEPCAT (PEPptide Conformational Analysis Tool) is a tool that was designed to address this issue by providing access to an array of conformational description criteria (descriptor types), which include atom-atom distance, residue-residue contact, and dihedral angles [253]. These three different geometric descriptor types can be used simultaneously with flexible cutoff criteria which can be defined to suit a particular problem. These criteria are then referred to as descriptors. The five-step process of a PEPCAT analysis is: conformation definition; classification; comparison; trajectory analysis; and visualisation of the results. Figure 2.13 (reproduced from O’Donohue et al. [253] with permission) illustrates the PEPCAT classification process.

In the classification stage, each descriptor gives a value (descriptor value) based on the criteria applied. For example, the dihedral descriptor type for Gly 2 in Figure 2.13 gives rise to four possible descriptor values. In the example, the application of three descriptors gives rise to three sets of descriptor values. Each unique set of descriptor values defines a conformational state, which is referred to by a number called a state ID. Each state ID is associated with a distinct conformational state, and is stored in a database. As
CHAPTER 2. COMPUTER SIMULATIONS OF PROTEINS

Figure 2.13: Illustration of PEPCAT classification process, reproduced from O’Donohue et al. (2000) with permission. The diagram shows the classification of a pyro-EEDCK trajectory in vacuo using the PEPCAT methodology. A trajectory consists of a time sequenced set of peptide coordinates (bottom right). Each frame is classified using a set of user specified descriptors (top right), and the set of values is looked up in a database of known states (left). If the set of descriptor values is not found in the database, a new entry is created. The conformational state identifier is returned (left) and stored with the frame. See text for more detail.

PEPCAT analyses a trajectory file, it classifies each frame of the simulation and compares its state ID to the known database. As the analysis progresses, new conformational states are identified and added to the database of known conformations. The state IDs of each frame in the trajectory are also recorded and form the basis for the visualisation stage.

At the end of the trajectory analysis, the conformational transition map and conformational distribution graph can be generated and visualised. The conformational distribution...
graph shows the population of each state in the trajectory and their relative similarity, and is useful in order to determine the stable conformational states present in the simulation. The conformational transition map shows the conformational changes that occur during the course of the trajectory by tracing the temporal progression of state IDs recorded in the trajectory analysis stage.

![Figure 2.14: A typical PEPCAT conformation map.](image)

A typical PEPCAT conformational transition map is shown in Figure 2.14. Each state ID is associated with a circle with size proportional to its relative population in the trajectory. The lines between the conformational states show the conformational transitions that occurred in that trajectory. The arrows identify the direction of the transitions, with width proportional to the number of transitions that occurred.

The PEPCAT descriptors applied in this work in order to classify the insulin chain-B dynamical behaviour are described below:
1. $\phi$- and $\psi$-dihedral angles of residues B8, B18, B19, B20, and B21, used to detect changes in both ends of the B9–B19 central helix. The division of the Ramachandran map is also shown, with the descriptor states labelled.

2. Contact distance between residues B16 and B24, used to examine changes at the C-terminal hydrophobic area. The division of the distance descriptor is shown on the right hand side.

3. Distance between the C$_\alpha$ atoms of residues B5 and B13, used to detect the packing direction of the N-terminus. In the figure below, the C$_\alpha$ atoms are highlighted in red. The division of the distance descriptor is shown on the right hand side.
Chapter 3

Insulin Simulations:
Ambient Conditions and Thermal Stress

The determination of the structure of insulin clearly opens up the way to similar studies on other proteins and already such studies are going on in a number of laboratories. These studies are aimed at determining the exact chemical structure of the many proteins that go to make up living matter and hence at understanding how these proteins perform their specific functions on which the processes of Life depend. One may also hope that studies on proteins may reveal changes that take place in disease and that our efforts may be of more practical use to humanity.

Frederick Sanger (1918–), Nobel Lecture, 11 December 1958

3.1 Overview

In this project, insulin has been used as the prototype protein for studying stress response. This chapter describes the reasons behind this choice. The brief history of insulin discovery and its importance is described in Section 3.2. In Section 3.3, the structural features of insulin are described in great detail. The methodologies involved in the simulation of insulin under ambient and thermal stress conditions are described in Section 3.4, along
with the details for each system. Section 3.5 describes the behaviour of isolated insulin chain-B under ambient and thermal stress conditions, which are then compared to the data obtained from NMR experiments. The behaviour of monomeric insulin chain-B (chain-B in the presence of chain-A with all disulfide bonds present) under ambient and thermal stress conditions is investigated in Section 3.6 and compared to the behaviour of the isolated insulin chain-B obtained in Section 3.5. Finally, the conclusion is presented in Section 3.7.

3.2 Brief history

First studies

Diabetes mellitus is one of the top ten diseases in the modern world. Due to its critical role in the treatment of diabetes, it is not surprising that insulin has been a subject of numerous experimental and computer simulation studies [12, 122, 260–270]. Aside from its therapeutical role, insulin also functions as a metabolite and weak growth hormone.

Insulin began its long and important role in science by a chance discovery of islet formation in pancreatic tissues by a German medical student named Paul Langerhans in Berlin in 1869. The islets were named Islets of Langerhans after him. At that time, the function of the islet was not recognised, although it was conjectured that it might produce a secretion that is useful to the digestive process because of its location in the pancreatic tissues.

It was not until 1889 that the link between pancreas and diabetes mellitus was shown when Oskar Minkowski removed the pancreas from a dog and noticed a swarm of flies feeding on the dog’s urine. It was shown upon testing that the urine contained a high concentration of sugar, which would otherwise not be present if the pancreas were still intact. In 1901, a major step forward took place when Eugene Opie showed that diabetes mellitus occurred only with partial or total destruction of the islets of Langerhans. This finding triggered a race to isolate the secretion of the islets as a treatment for diabetes mellitus.
For 20 years, there was no progress in isolating the secretion, until Frederick Banting realised that the islet secretion was being broken down by the digestive secretions. This led him to approach J.J.R. Macleod at the University of Toronto in 1921. Together with Charles Best, they successfully isolated the secretion, which they called isletin [271].

The first method they used was blocking the pancreatic ducts of a dog, which killed the pancreatic digestive cells, while preserving the islets. The second method they used was extracting the isletin from foetal calf pancreas, which had not developed pancreatic digestive cells, thereby simplifying the procedure immensely.

James Collip was recruited to help with purifying the isletin to use for a human trial. On 11 January 1922, Leonard Thompson became the first human recipient of an insulin injection. Unfortunately, he suffered a severe allergic reaction due to the impurities present in the insulin. On 23 January, a second dose was injected, this time the insulin was pure enough, and the trial was considered a complete success.

A few months after that, Eli Lilly was recruited to replace Collip in developing a way to mass produce insulin with sufficiently high purity. They succeeded in November 1922, and insulin began to be commercialised. For their work in isolating insulin, Macleod and Banting were awarded the 1923 Nobel Prize in Physiology or Medicine.

**Determination of primary and tertiary structures**

By the 1940s, it was known that the use of strong acids breaks down the peptide bonds of a protein, leaving individual amino acid residues. This, in fact, is what is happening when food is digested by gastric acids. However, if the acid is weak enough, or the enzyme has amino acid selectivity in breaking the peptide bonds, then it is possible to obtain fragments of polypeptides containing only few residues at a time.

Frederick Sanger used exactly this method, combined with a method to mark the terminal residues of a polypeptide chain using 2-4-dinitrofluorobenzene dye reagent (“Sanger’s reagent”), to determine the primary sequence of insulin [272]. From the work he had done with the dye reagent, he deduced that there had to be two chains in an insulin molecule, one with 21 residues, and another with 30 residues. Sanger spent 15 years...
painstakingly sorting through the amino acid fragments using various chromatographic and electrophoretic techniques, until in 1955, he finally solved the primary sequence of insulin, the first protein to be analysed so. For his work, he was awarded the 1958 Nobel Prize in Chemistry.

In 1967, the three-dimensional structure of insulin was solved by Dorothy Hodgkin by means of x-ray diffraction studies. For her earlier work which made possible the solving of the insulin structure, she was awarded the 1964 Nobel Prize in Chemistry.

### 3.3 Insulin structure

**Primary structure**

The primary sequence of porcine insulin (PDB accession code 1ZNI), as used in this study, is shown in Table 3.1. Chain-A consists of 21 residues and chain-B consists of 30 residues.

<table>
<thead>
<tr>
<th>Chain A</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Gly</td>
<td>Ile</td>
<td>Val</td>
<td>Glu</td>
<td>Gln</td>
<td>Cys</td>
<td>Cys</td>
<td>Thr</td>
<td>Ser</td>
<td>Ile</td>
</tr>
<tr>
<td>10</td>
<td>Cys</td>
<td>Ser</td>
<td>Leu</td>
<td>Tyr</td>
<td>Gln</td>
<td>Leu</td>
<td>Glu</td>
<td>Asn</td>
<td>Tyr</td>
<td>Cys</td>
</tr>
<tr>
<td>20</td>
<td>Asn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chain B</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Phe</td>
<td>Val</td>
<td>Asn</td>
<td>Gln</td>
<td>His</td>
<td>Leu</td>
<td>Cys</td>
<td>Gly</td>
<td>Ser</td>
<td>His</td>
</tr>
<tr>
<td>10</td>
<td>Leu</td>
<td>Val</td>
<td>Glu</td>
<td>Ala</td>
<td>Leu</td>
<td>Tyr</td>
<td>Leu</td>
<td>Val</td>
<td>Cys</td>
<td>Gly</td>
</tr>
<tr>
<td>20</td>
<td>Glu</td>
<td>Arg</td>
<td>Gly</td>
<td>Phe</td>
<td>Phe</td>
<td>Tyr</td>
<td>Thr</td>
<td>Pro</td>
<td>Lys</td>
<td>Ala</td>
</tr>
</tbody>
</table>

**Table 3.1:** Primary sequence of porcine insulin

The majority of insulin’s primary sequence is conserved in different species, including key residues that might be important to receptor binding (e.g. residues A3, B24, and B25). However, minor differences between species do occur, as shown in Table 3.2. While these variations give rise to the difference in therapeutic potency, they do not affect the tertiary structure of insulin [273].
CHAPTER 3. INSULIN SIMULATIONS: AMBIENT CONDITIONS AND THERMAL STRESS

<table>
<thead>
<tr>
<th>Species</th>
<th>A8</th>
<th>A10</th>
<th>B30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porcine</td>
<td>Thr</td>
<td>Ile</td>
<td>Ala</td>
</tr>
<tr>
<td>Bovine</td>
<td>Thr</td>
<td>Ile</td>
<td>Thr</td>
</tr>
<tr>
<td>Human</td>
<td>Ala</td>
<td>Val</td>
<td>Ala</td>
</tr>
</tbody>
</table>

Table 3.2: Primary sequence differences between different species

Tertiary and quaternary structure

An insulin monomer (Figure 3.1) consists of two chains, A and B. Chain-A consists of two helices ranging from A1–A8 and A13–A20, linked by a loop region and stabilised by an intrachain disulfide bond between residues A6–A11. Chain-B consists of a variable N-terminal region, a central helix at B9–B19, a loop at B20–B23, and an extended C-terminus. The central helix B9–B19 is conserved between species and is flanked by two disulfide bonds between A7–B7 and A20–B19. These two disulfide bonds lock the two chains together and exert a stabilising influence on the central helix of chain-B.

Figure 3.1: Tertiary structure of insulin. Chain-A is in blue, chain-B is in red, and disulfide bonds are in yellow.

The formation of insulin itself is believed to be mediated by a third chain, chain-C, located between the two main chains, which is then cleaved in the production process, along with several extra residues from the N-terminus of chain-B [274].

Under normal physiological conditions, insulin forms dimers by forming hydrogen bonds between residues B24 and B26 [275, 276]. These hydrogen bonds link the two C-termini of the two monomers and form an anti-parallel β-sheet structure. In the pres-
ence of ions such as Zn$^{2+}$ and Cl$^{-}$, the insulin dimers can form tetramers and hexamers, with the ions acting as their crystal centre [12, 244, 262]. The hexameric form is particularly important, as it is the form adopted when insulin is stored in the pancreas [277]. The quaternary structure of hexameric insulin is shown in Figure 1.8.

Tertiary structure variations

X-ray diffraction studies revealed variations in insulin tertiary structure, depending on its crystallographic environment [278]. The variations occur primarily in the N-terminal region of chain-B [12, 244, 260, 279, 280]. The two main conformation states are the R- and T-states [123]. In the R-state, residues B1–B8 are in $\alpha$-helical conformation, so that the overall helical region extends from B1 to B19. In the T-state, residues B1–B8 are in an extended conformation. In addition to the R- and T-states, other conformations also exist, such as the R$^f$-state ("frayed"-state) [260, 262] and the O-state [280]. In the R$^f$-state, residues B1–B3 are in an extended conformation, while B4–B8 are in $\alpha$-helical conformation. The O-state ("open"-state) is almost identical to the T-state, with the difference being the extended B1–B8 residues are further away from the main body of the insulin molecule, and are thought to be involved in receptor binding. The different conformations of insulin are shown in Figure 3.2.

The hydrophobic core region of insulin chain-B has long been suspected to be important for receptor binding of insulin [124]. Studies on despentapeptide insulin (DPI)—insulin which lacks residues B26–B30—have shown that it retains the full activity of insulin, while lacking the dimer-forming region [281]. This indicates that the C-terminus of chain-B is not necessary for receptor binding. However, its positioning is critical by exposing the hydrophobic patch located primarily in the N-terminal region of chain-A [124, 282, 283]. From the DPI studies, it is generally accepted that the active form of insulin in solution is a monomer [281]. However, efforts in isolating its crystal structure have been hampered by insulin’s solubility and propensity to self-association [7]. There is also evidence that the observed conformation of the “native” insulin structure is an artefact of the crystallographic process, where crystal contacts due to the packing of the
molecules modify the active form of the structure [284, 285]. Thus, the structural knowledge of native monomeric insulin has been primarily inferred from NMR experiments on engineered monomeric insulin, e.g. DPI, with the active conformation identified to be analogous to the T-state [286, 287].

Experimental studies on physiologically active monomeric forms of insulin have observed fluctuations in the N-terminus of chain-B [7, 286, 288, 289] as well as flexible regions in the C-terminus of chain-B [286, 289–291]. The importance of flexibility is also demonstrated by the lack of activity shown by insulin when residues A1 and B29 are crosslinked [283, 292]. This indicates that the native insulin monomer is intrinsically flexible and that this property is integral to the interaction of insulin and its receptor [284, 286]. Further support for insulin’s flexibility comes from the observation that native insulin has analogous conformation to inactive insulin, suggesting that the native insulin undergoes conformation change upon binding to its receptor [283].
3.4 Methodologies

The general methodologies used in this chapter were outlined in Section 2.5. In this chapter, we are focusing our attention on the behaviour of insulin chain-B in both isolated form and in the presence of insulin chain-A, under ambient conditions ($T = 300 \text{ K}, \ P = 1 \text{ bar}$) and thermal stress ($T = 400 \text{ K}, \ P = 1 \text{ bar}$).

Emphasis has been put on the isolated chain-B system, as it has been shown to be primarily stable independently of chain-A, with little change to its tertiary structure apart from increased mobility in the C-terminal region [261, 293]. The importance of chain-B has also been suggested by several studies that highlighted it as a folding template for chain-A [293–295].

To this end, five independent simulations of isolated insulin chain-B were performed, totalling nearly 100 ns: four simulations each lasting 20 ns were performed at ambient conditions, and a simulation lasting 18 ns was performed at an elevated temperature of 400 K. The latter was performed to simulate the ultra-fast temperature increase believed to be one of the interaction mechanisms of the electromagnetic radiation with tissues [151].

For the monomeric insulin chain-B, two systems were simulated, representing ambient conditions ($T = 300 \text{ K}, \ P = 1 \text{ bar}$) and elevated temperature ($T = 400 \text{ K}, \ P = 1 \text{ bar}$). The behaviour of isolated chain-B was then compared to the behaviour of insulin chain-B in the presence of chain-A and all disulfide bonds intact. The statistics of insulin behaviour at ambient conditions and under thermal stress were later used for comparison with the protein’s behaviour under electric field stress.

A complete list of the conditions considered in this chapter is given in Table 3.3. The letter “N” denotes normal conditions (no external field applied); the letter “M” denotes monomeric insulin (i.e. chain-A is present, and the disulfide bonds between the two chains left intact); and the letters “A” to “D” represent the four simulations performed on isolated insulin chain-B. In the following sections, system N300 refers to the average data accumulated from systems N300A to N300D.
### Isolated chain-B

<table>
<thead>
<tr>
<th>System name</th>
<th>Simulation length (ns)</th>
<th>Temperature (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N300A</td>
<td>20</td>
<td>300</td>
</tr>
<tr>
<td>N300B</td>
<td>20</td>
<td>300</td>
</tr>
<tr>
<td>N300C</td>
<td>20</td>
<td>300</td>
</tr>
<tr>
<td>N300D</td>
<td>20</td>
<td>300</td>
</tr>
<tr>
<td>N400</td>
<td>18</td>
<td>400</td>
</tr>
</tbody>
</table>

### Monomeric chain-B

<table>
<thead>
<tr>
<th>System name</th>
<th>Simulation length (ns)</th>
<th>Temperature (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM300</td>
<td>10</td>
<td>300</td>
</tr>
<tr>
<td>NM400</td>
<td>10</td>
<td>400</td>
</tr>
</tbody>
</table>

Table 3.3: Summary of simulation conditions used in this chapter.

#### 3.5 Isolated insulin chain-B

##### 3.5.1 Structural Stability

The central helix residues (B9–B19) in each MD simulation were aligned and compared to the minimised crystal structure. The RMSDs of the backbone of the central helix for the five systems are shown in Figure 3.3. The plot shows a moving average of 250 ps to enable effective comparison of the overall trends between the systems. The RMSDs were generally stable, even at the elevated temperature, however some interesting variations occurred. In the first 300 K simulation (system N300A), the system showed overall stability, although several sharp increases in RMSD reflect distinct conformational transitions (at approximately 3.0 ns the RMSD increases to 2.50 Å and at 11.5 ns to 3.25 Å). As expected, the RMSD at 400 K shows persistent variation, although within the upper and lower limits observed in the 300 K systems. The RMSDs suggest that the molecule is quite mobile and flexible during the MD simulations. Therefore NMR observations can provide useful insights into its behaviour.
3.5.2 Comparison with NMR experimental results

The data derived from the NMR structure of bovine insulin chain-B in solution [261] provide a useful benchmark for evaluating the conformations observed during the MD calculation. Hawkins, Cross, et al. 1995 derived nuclear Overhauser effect (NOE) constraints from a 250 ms NOESY spectrum of the oxidised chain B at 500 MHz, 300 K and pH of 2.2 to 2.5. As coordinates for the structure were not available, NOE distance restraints obtained from the authors were used for comparison [296]. Using the program X-PLOR [254], the interproton distance restraints were used to calculate energy values for the MD conformations from the five trajectories. To evaluate our results, we compared these values with the NOE energy value of the minimised structure. The NOE energy plot of the structures generated from the five simulations is shown in Figure 3.4. We found that the majority of the modelled structures satisfied the distance restraints with lower energy values compared to that of the minimised crystal structure. It is interesting to note the correlation at \( \sim 11.5 \) ns between the decrease in NOE energy and the previously observed sharp increase in RMSD in system N300A (Figure 3.3). This indicates the changes in conformation which occurred at this time point had fewer NOE violations, i.e., the structures were in closer agreement with the experimental data. The exception to the
generally low energy values occurred in one of the 300 K simulations (system N300D). In this simulation there was a noticeable increase in energy at about 8 ns indicating changes in conformation, which in this case, showed less agreement with experimental data. This observation stimulated further interest in the forces driving this divergence in structure, which will be discussed in later sections.

![Figure 3.4: The NOE energy derived from the distance restraints of the starting structure compared to all systems. The shaded region represents the MD structures with NOE energy less than the starting structure. The plot shows a moving average of 250 ps.](image)

3.5.3 Structural evolution

The secondary structure dynamics reveal details of conformational changes over the simulation time course. The secondary structure content as a function of time for the 300 K (systems N300A–N300D) and 400 K simulations are displayed in Figure 3.5. The secondary structure was classified by VMD using the algorithm STRIDE which utilises hydrogen bond energy and main chain dihedral angles in addition to hydrogen bond distances [256].

The content of the secondary structure plot of each system shows variation, while they all share some common elements of secondary structure. All of the simulations showed an initial change in Cys B19 from the starting structure, most noticeable in
systems N300A (Figure 3.5a) and N300C (Figure 3.5c). This can be expected because of the loss of the disulfide bonds with chain-A. The helical content was maintained in each secondary structure plot although with some variation. The changes observed were most distinctive in the N-terminus of the central helix (residues B4–B9). The secondary structure assignments for the first of the 300 K simulations system (Figure 3.5a) showed a loosening of atom-atom contacts by 3 ns, resulting in the \( \alpha \)-helix structure developing into \( \pi \)-helix structure and finally losing the hydrogen bonds of the helix altogether. Transitions between \( \alpha \)-helix structure and \( \pi \)-helix have been previously reported in MD simulations. They are believed to be genuine physical transitions and an important stage in helix-to-coil transitions [297]. Conversely, a recent study has indicated that the occurrence of \( \pi \)-helix
structure in MD may be an artefact of specific forcefields [186]. However, this investigation of CHARMM, and other forcefields, was performed using implicit solvent, which cannot necessarily be compared to simulations in explicit solvent, such as ours. The π-helix conformation is difficult to measure quantitatively by experiment. The conformation is rarely identified in crystal structures, although this could be attributed to the definition algorithm of helical structure. Modified π-helix definitions have shown that the π-helix is more than 10 times more prevalent than previously believed [298]. Therefore, despite the suggestion that the existence of π-helix can be an artefact of the forcefield, there is sufficient evidence to suggest that the π-helix does occur in the protein’s dynamic behaviour and plays an important role in the function of a protein [297, 298].

Also of interest was the change in structure at 8 ns, where the α-helix region between residues B4 and B7 reformed while accompanied by a change in structure of Gly B8. These results are consistent with experimental studies showing that Gly residues are often associated with perturbations in structure and are known as “helix-breakers” [6]. In addition, evidence of helical content in residues B4 to B7 was observed previously in the NMR solution structure of an insulin chain B mutant [299]. In our simulation, this trend was not observed in the N300C system where the helical content was mostly maintained with a small amount of fraying at the N-terminus (Figure 3.5c). The trend did occur, although to a lesser degree, in systems N300B and N300C (Figures 3.5b and 3.5d, respectively). The N300D system, however, showed variation where the simulation followed the same trend initially and then showed some distinctive differences. Between 5 ns and 8 ns, the secondary structure content showed strong similarities to the N300A system with a change in geometry of Gly B8 and the appearance of helical content in residues B4–B7. However, at about 8 ns there was a noticeable loss of helical content in residues B8–B12. It was at this time period in the simulation when a significant increase in the NOE energy was previously noted for this system (Figure 3.4) suggesting the simulation may have become trapped in conformations corresponding to energetically higher local minima.

Another interesting secondary structure feature was the movement of the β-turn, again, most pronounced in the N300A system (Figure 3.5a). During the simulation at about 11.5
ns, the turn moved from its initial position localised in residues B20–B23 in the starting (crystal) structure to residues B17–B20. This was in agreement with the turn-like structure defined in the oxidised NMR structure of chain-B for residues B17–B21 [261]. This movement was previously observed as an increase in RMSD (Figure 3.3) accompanied by a sharp decrease in NOE energy (Figure 3.4) confirming close agreement with experimental data.

In the 400 K simulation more secondary structure fluctuation was observed, however, the central helix was in general maintained (Figure 3.5e). The helical content was reduced to residues B10–B15, although there was a brief return to helical structure in the C-terminal end of the central helix. Notably, more movement was observed in the \( \beta \)-turn in the 400 K simulation, shifting in position between residues B16 and B27, thus reinforcing the view that the position of the turn is flexible.

Table 3.4 presents the number of all-helical residues (consisting of \( \alpha \)-, \( \pi \)-, and \( 3_{10} \)-helices) and \( \alpha \)-helical residues of insulin chain-B retained over the whole data collection stage. The helical contents of the simulations at 300 K have been included in the average and presented as system N300.

<table>
<thead>
<tr>
<th>System name</th>
<th>Retained all-helical residues</th>
<th>Retained ( \alpha )-helical residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>N300</td>
<td>10.7 ± 2.7</td>
<td>8.3 ± 3.9</td>
</tr>
<tr>
<td>N400</td>
<td>8.8 ± 2.4</td>
<td>7.9 ± 2.7</td>
</tr>
</tbody>
</table>

Table 3.4: Retained number of all-helical and \( \alpha \)-helical residues for isolated insulin chain-B under ambient conditions and thermal stress, averaged over the data collection stage.

### 3.5.4 Conformational transitions

The conformations sampled during the MD of each system were classified according to the structure of the N-terminus and central helix regions to identify stable states and transitions between them. In order to compare the conformational states between the five simulations performed in this section, the PEPCAT descriptors chosen were the 17 dihedral angles between residues B3 and B19. These residues are located in the central helix.
and the N-terminus. The descriptors were chosen in order to observe structural changes that occur within the helical region B4–B19 of the R\textsuperscript{f}-state throughout the simulation. The PEPCAT conformational transition maps displaying the relative population of each conformation (state) and the transitions between the states for each system are shown in Figure 3.6. The analysis of the first simulation (system N300A) in Figure 3.6a revealed only a relatively small number of conformational states (10), where state 0 represents the initial R\textsuperscript{f} conformation. The most populated states are also shown. The sizeable number of transitions, which occurred between state 0 and state 7 at 5.5 to 8 ns, was the result of the change in conformation of Gly B8 noted previously in the secondary structure analysis. The conformations of typical structures representing states 0 and 7 (shown in Figure 3.7a) illustrate the transition between the two states. It can be seen that this transition involved a change from the R\textsuperscript{f}-state (state 0) to a conformation that was very similar to the T-state conformation (state 7), which is believed to be the monomeric solution structure important for the hormone’s activity. The state 7 structure, stabilised by a hydrogen bond between residues B5 and B13, will be discussed in more detail later.

The conformational analysis of the second 300 K simulation (system N300B) revealed an increased number of structures compared to that of the previous simulation (86). For clarity, only the major conformations are displayed in the transition map in Figure 3.6b representing nearly 71% of the sampled structures. Conformational state 7 was only visited briefly and new pathways are followed through states 15, 32, and 48. The additional conformational states are largely structures with variations centred on Gly B8. Figure 3.7b shows typical structures of states 15, 32 and 48. After a transition through state 15, states 32 and 48 are generally similar to the T-state conformations and are stabilised by hydrophobic interactions between the N-terminus and the central helix.

The stability observed in the secondary structure in the third 300 K simulation (system N300C) was reflected in the small number of conformational states (8) sampled. The transition map in Figure 3.6c illustrates how the conformational changes did not progress much beyond state 0 (R\textsuperscript{f}-state). State 7, a key conformation observed in previous 300 K simulations, was only briefly visited. The conformations generated from this simulation
Figure 3.6: Conformational transition map for all systems. (A) N300A, (B) N300B, (C) N300C, (D) N300D, (E) N400. The circles, labelled with the state identification number, represent conformational states visited during the simulation, and the size of the circle is proportional to the frequency. The lines between the circles indicate the conformational transitions and the width is proportional to the frequency. For classification of the states please refer to the accompanying text.

are low NOE energy structures (Figure 3.4) and represent a subset of the structures already observed.

The conformational analysis in the fourth 300 K simulation (system N300D) revealed a moderately large number of structures classified in the conformational analysis (42). The major conformational states representing 84% of the sampled structures are shown in the conformational transition map in Figure 3.6d. As indicated previously, there was evidence of an alternative conformational transition. Initially states 0 and 1 were well populated, however the conformations moved quickly along a pathway, only briefly visiting state 7, to a geometry which was largely centred on state 95. Figure 3.6c shows typical structures of states 12 and 95, the most frequently conformed structures. It is worth noting the close proximity
Figure 3.7: Typical structures from the major conformational states for selected systems. (A) N300A, (B) N300B, (C) N300D, (D) N400. The structures show the variation in the N-terminal region.

between the N- and C-termini in both states 12 and 95, indicating these structures are stabilised by hydrophobic interactions.

At 400 K there was a significant increase in the number of conformational states observed (310), as expected. The most populated states are shown in Figure 3.6e representing 61% of the sampled structures. The purpose of increased temperature was to explore a broader region of conformational space in order to identify additional new states. Altogether, 46% of the states observed in the 300 K simulations were observed in the 400 K system (accounting for 34.5% of the total number of structures). The remaining states were classified as new. The conformations included many of the most populated ones previously observed at 300 K, for example, states 0, 7, 15 and 32. Many additional conformations were observed through the transition states 8 and 32, and the conformations were again largely differentiated around the pivotal turning point of Gly 8 with increased
mobility of the extended region of the N-terminus. The variations in the conformations of the major states are displayed in Figure 3.7d.

Even though our structural analysis showed a diverse range of conformations in all systems at 300 K and 400 K, overall 86% of structures satisfied the experimental NMR distance restraints. A comparison of the structures and transitions shows similarities in the conformations generated in the time course of the simulations. Many of the experimentally known structural properties were reproduced in the simulations. The central $\alpha$-helix (residues B9–B19) was largely conserved in all the simulations apart from some fraying occurring at the ends. This would suggest that the central helix is the most stable region and possibly the folding nucleus of the molecule. Evidence of helical structure was also observed in residues B4–B7 although this was far more transient in frequency. We observed the appearance of stabilising interactions, such as between the residues B4 to B7 and the central helix. However, one simulation at 300 K showed some distinct differences where the conformational dynamics leads to structures with lower NOE energy. A comparison of the N300A and N300D systems revealed more detail of this divergence and is schematically illustrated in Figure 3.8. In both simulations there was a change in geometry of Gly B8 between 5 ns and 6 ns. In the N300A system, the change in Gly B8 conformation lead to a change in the central helix characterised by a movement of the $\beta$-turn. This lead to a downward swing of the N-terminus with a corresponding outward movement of the C-terminus breaking the hydrophobic interactions with the central helix (illustrated in Figures 3.7a and 3.8). This resulted in a T-state conformation. Conversely, in the N300D system (Figure 3.7c) the hydrophobic interaction between the N- and C-termini resulted in the second helix (residues B4–B7) being pulled up towards the C-terminus instead of down as in N300A. Both sets of conformations were effectively trapped in local energy minima and stabilised by hydrophobic interactions.

The correlation between the NOE energy structures and the distance between the N- and C-termini for the two systems is shown in Figure 3.9. In the N300A system there was a marked drop in NOE energy associated with an increase in the distance between the two termini regions and the more T-state like conformation. In the N300D system there
was a marked increase in NOE energy associated with a decrease in distance between the two regions. This reinforces the importance of the T-state conformation.

The T-state-like conformation observed in our simulations is characterised by interactions between the N-terminus region and the central helix. This association is illustrated by the time history plot of the distance between residues B5 and B13 shown in Figure 3.10. These two residues were important in the stabilisation of the conformations in the N300A system by a hydrogen bond. The plot shows the strong interaction between these regions occurring in the N300A system. This was not evident in the N300D system where the plot showed the distance between these two residues markedly increased, a divergence occurring at approximately the same time in the simulation. This is a reflection of the different conformational pathways observed in these systems. Figure 3.10 also shows the distance over time in the structures from the high temperature simulation (system N400). Although the distance between the residues was not as close as that observed in the N300A system, the high temperature system does appear to avoid the structural entrapment which occurred in the N300D system.
Figure 3.9: Comparison of NOE energy and distance (C\textsubscript{\alpha} atoms) between N- and C-termini in selected systems (residues B6 to B26). (A) N300A. (B) N300D. The plot shows a moving average of 250 ps.

Figure 3.10: Time dependence plot of the distance between residues B5 and B13 (C\textsubscript{\alpha} atoms) for N300A, N300D and N400 systems. The plot shows a moving average of 250 ps.
Whilst there are differences in the order and population of the conformational states observed in the 400 K system, the structures generally satisfied the NOE experimental constraints, and there was a significant amount of residual helical structure maintained. These detailed observations and analyses suggest high temperature simulations are a valid means of avoiding entrapment in local minima on the potential energy surface. High temperature exploration of a wider conformational space results in the production of viable conformations, and possibly an intimation of unfolding.

3.5.5 Dipole moment

Because of the helical nature of the central region, insulin chain-B possesses a strong dipole moment, which has a potential to interact with an external electric field. However, under no field conditions, the protein is free to rotate along all of its axes. Figure 3.11 shows the distribution of dipole moment values of the B9–B19 helix along the $z$-axis. The dipole moments for all the systems at 300 K have been averaged and presented as system N300. The plot confirms the lack of directional preference in insulin with no electric field applied, and will be used for comparison with the electrically stressed behaviour in later chapters.

![Figure 3.11: B9–B19 helical dipole moment distribution in the $z$-direction of isolated insulin chain-B with no field applied. Both ambient and elevated temperature conditions show no directional preference.](image-url)
3.5.6 Radius of gyration

The average $r_g$ of the isolated chain-B system was calculated and presented in Table 3.5. The individual $r_g$ of the systems at 300 K have been averaged and presented in the table as system N300. The $r_g$ of the system under elevated temperature was observed to be 11% larger than the $r_g$ of the systems under ambient conditions. In addition, the variance of the $r_g$ of the thermally stressed system was also observed to more than double the variance of the unstressed systems. This is indicative of the higher degree of flexibility of the thermally stressed chain-B, and agrees with the PEPCAT analysis which shows a greater degree of conformational freedom.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [$\text{Å}$]</th>
</tr>
</thead>
<tbody>
<tr>
<td>N300</td>
<td>$9.9 \pm 0.5$</td>
</tr>
<tr>
<td>N400</td>
<td>$11.0 \pm 1.3$</td>
</tr>
</tbody>
</table>

Table 3.5: Radius of gyration of each simulation, averaged over data collection stage, after the first 1 ns.

3.5.7 Solvent accessible surface area

The SASA of the protein was calculated using the program X-PLOR [254] for the whole data collection stage after the first nanosecond of the simulation. For clarity, the resulting SASA and variance are expressed as a percentage of the SASA of the N300A system.

<table>
<thead>
<tr>
<th>System name</th>
<th>Percentage SASA</th>
</tr>
</thead>
<tbody>
<tr>
<td>N300A</td>
<td>$100.0 \pm 5.8$</td>
</tr>
<tr>
<td>N400</td>
<td>$107.7 \pm 6.1$</td>
</tr>
</tbody>
</table>

Table 3.6: Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation.

The result qualitatively agrees with the obtained values of the $r_g$, where the thermally stressed system had increased $r_g$ due to the application of increased temperature.
3.5.8 Biological implications

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

The central helix region has considerable impact on the dynamics of the protein and also plays a key role in the protein’s activity. Most of this region has retained its integrity throughout the simulation, even at 400 K. Gly B8 is highly conserved in all species of insulin, with experimental studies showing a loss of activity if Gly B8 is replaced by other residues [265, 300]. Gly is able to accommodate a range of dihedrals in the Ramachandran plot. Our results indicate that because of its unique geometric properties, Gly B8 acts as a pivotal point between the conformations necessary for activity. Examination of our simulated structures revealed evidence of a hydrogen bond commonly occurring between the side chains of His B5 and Glu B13 thus having the effect of stabilising the central helix. Figure 3.12a shows the starting structure, and Figure 3.12b shows the interaction between His B5 and Glu B13 in a low NOE energy structure from the N300A system. It must be noted that experimental results show that Glu B13 mutants result in decreased binding affinity [300], therefore confirming our findings of the important role of this residue in stabilising the protein’s structure.

The movement or flexibility of the $\beta$-turn from residues B20–B23 (in the x-ray structure) to residues B17–B20 resulted in the C-terminus moving away from the central helix thus exposing hydrophobic residues important for activity. The importance of the destabilisation of the $\beta$-turn and C-terminus with respect to receptor binding has been suggested in experimental studies [286]. Furthermore, this may also partly explain why alanine scanning mutagenesis experiments show enhanced binding affinity for Gly B20 mutants [300]. Flexibility in the position of the $\beta$-turn implies that Gly B20, unlike Gly B8, is not relied on for its unique geometric properties.
Figure 3.12: Illustration of the role of individual residues. (A) Starting structure. (B), (C) low NOE energy structure (N300A system). A comparison between structures (A) and (B) show the stabilisation of the T-state conformation by a hydrogen bond interaction between the side chains of His B5 and Glu B13; and the movement of the side chains of Phe B24 and Phe B25 between the closed (A) and open (B) conformation. Structure (C) shows the two distinct binding sites of insulin to its receptor exposed with movement of the C-terminus.

Phe B24 in chain-B is of particular interest because this residue has been shown to play an integral role in hormone-receptor binding [124, 290, 292]. In the theory discussed by Weiss [301], the receptor-binding surface of insulin is defined in terms of open and closed conformations. This theory suggests the role of Phe B24 in chain B. In the closed conformation Phe B24 anchors the conformation via hydrophobic interactions with residues B14–B20. In the open conformation, Phe B24 moves away from the helix region, thus exposing hydrophobic residues in the central helix. The transition between the two conformations was observed during the simulations and is illustrated in Figure 3.12. The starting structure in Figure 3.12a represents the closed conformation and the simulated structure in Figure 3.12b shows the position of side chains of Phe B24 and Phe B25 in the open conformation.

Since insulin is cross-linked to the receptor dimer, it has been suggested that there are two distinct binding sites of insulin to its receptor [302]. Residues from both chain-A and chain-B are thought to be involved in the binding site [303, 304]. The residues identified in chain-B include residues B12, B15, B16 located in the central helix, and residues B23
to B25 in the C-terminus. The simulated movement of the C-terminus exposing the two distinct binding sites is displayed in Figure 3.12c.

3.6 Monomeric insulin chain-B

3.6.1 Stability

The secondary structure evolution of insulin chain-B under monomeric condition is presented in Figure 3.13. In contrast to the secondary structures of the isolated chain-B systems, the secondary structures of these systems were observed to be very stable, which is not surprising due to the presence of the disulfide bonds between residues A7–B7 and A20–B19. These bonds stabilise the central helix of insulin chain-B, which results in remarkable stability even under the application of thermal stress. The stability of these systems is even more apparent through the lack of $\alpha$- to $\pi$-helix transition observed in the isolated chain-B systems.

Table 3.7 presents the retained number of $\alpha$-helical residues averaged over the whole data collection range of the simulations. Only the number of $\alpha$-helical residues are presented, since no transition towards $\pi$-helix was observed. Similarly to the isolated chain-B systems, the monomeric chain-B under thermal stress was observed to lose approximately 16% of its all-helical content. The loss of the helical content in system NM400 occurs at the N-terminus as a consequence of the increased mobility and flexibility at that region. Due to the disulfide bonds bracketing residues B7 to B19, the core helical residues that span from residues B9 to B19 are fully conserved in both ambient and thermally stressed systems. Flexibility was also observed at the C-terminus region, as can be observed from the secondary structure plot (Figure 3.13b).

The stability of the simulations was monitored by observing the RMSDs of the B9–B19 helix backbone over time, calculated relative to the start of the data collection stage, shown in Figure 3.14. The plots have been averaged by 100 ps for clarity. They confirm the finding in the secondary structure analysis that both ambient and thermally stressed systems are stable with B9–B19 backbone RMSD values rarely exceeding 0.8 Å.
CHAPTER 3. INSULIN SIMULATIONS: AMBIENT CONDITIONS AND THERMAL STRESS

Figure 3.13: Secondary structure evolution of the monomeric chain-B systems under static electric field.

Table 3.7: Retained number of $\alpha$-helical residues for monomeric insulin chain-B under no field conditions, averaged over the data collection stage.

<table>
<thead>
<tr>
<th>System name</th>
<th>Retained $\alpha$-helical residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM300</td>
<td>16.0 ± 0.2</td>
</tr>
<tr>
<td>NM400</td>
<td>13.5 ± 1.6</td>
</tr>
</tbody>
</table>

The RMSD values for the last 1 ns of the simulations are presented in Table 3.8. The RMSDs are relatively small with similar variance. The NM400 system has a slightly larger RMSD compared to the NM300 system and reflects the increased flexibility which arises from the application of increased temperature.
CHAPTER 3. INSULIN SIMULATIONS: AMBIENT CONDITIONS AND THERMAL STRESS

Figure 3.14: 100 ps running average of the B9–B19 helix backbone RMSD of the monomeric chain-B systems

<table>
<thead>
<tr>
<th>System name</th>
<th>RMSD [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM300</td>
<td>0.56 ± 0.09</td>
</tr>
<tr>
<td>NM400</td>
<td>0.7 ± 0.1</td>
</tr>
</tbody>
</table>

Table 3.8: B9–B19 helical backbone RMSD of each monomeric chain-B simulation under no field conditions, averaged over the last 1 ns.

3.6.2 PEPCAT analysis

In order to investigate the differences between conformations sampled by the isolated and monomeric chain-B, PEPCAT analysis was performed on the monomeric chain-B systems and compared to the PEPCAT analysis of systems N300A and N400. System N300A was chosen as a representative system for the ambient conditions, as it is the only one which sampled the T-like structure which is associated with insulin bioactivity.

Whereas the PEPCAT descriptors used in Section 3.5.4 were chosen to pick up structural changes in the B4–B19 helical region, in this section the descriptors described in Section 2.7 were used instead. These descriptors were designed to observe changes to the conserved B9–B19 helical region; monitor changes to the hydrophobic area of the C-terminal region; and detect the packing direction of the N-terminal region.

Figure 3.15 shows the transition maps for systems NM300, NM400, N300A, and N400. Refer to Figure 2.14 in Section 2.7 for detailed explanation of map features. The results show that the monomeric chain-B systems sample far fewer states compared to the isolated
Figure 3.15: PEPCAT transition maps for systems considered for PEPCAT analysis. The yellow circles represent a unique state ID, with size proportional to the relative population of that state within the simulation timeframe. The grey arrows indicate transitions between states, with size of arrowheads representing the relative number of transitions to that state. Refer to Section 2.7 for full explanation of map features. For clarity, only states appearing more than 10% of the total simulation timeframe are presented, except for system NM300 where all states are presented.
chain-B systems, as a consequence of the presence of chain-A with the connecting disulfide bonds. The systems at thermal stress show a larger number of states with interconnecting pathways between them. This is indicative of the flexibility and mobility of the protein under thermal stress, and also the lack of preference for a specific pathway.

Figure 3.16: Snapshots of the two most occupied states for the systems considered in this section.

Figure 3.16 shows the two most occupied states for the systems considered for PEPCAT analysis. In the monomeric systems, chain-A is present throughout the simulations, but not shown in the figure. The monomeric systems were observed to have very stable secondary structure, even at high temperature. This illustrates the strength of the disulfide bonds which connect the two chains together.
3.6.3 Dipole moment

The monomeric insulin contains three helical regions: the helical residues of chain-B, the A1–A8 helix, and the A13–A20 helix. As will be demonstrated in the following chapters, these helices have the potential to be affected by the applied electric field, yielding directional preference. However, under no field conditions, no such preference should be observed. Indeed, the lack of directional preference is demonstrated in Figure 3.17, showing the distribution of dipole moments in the $z$-direction for the monomeric insulin chain-B.

![Figure 3.17: Dipole moment distribution in the $z$-direction of insulin chain-B in the presence of chain-A.](image)

3.6.4 Radius of gyration

Table 3.9 shows the $r_g$ of the monomeric chain-B systems, averaged over the data collection stage, after the first 1 ns. Only chain-B is considered in the calculation. The NM300 system has a larger $r_g$ than the N300 system. This is due to the stability of the NM300 system, where the presence of disulfide bonds between chain-A and chain-B prevented chain-B in monomeric conditions to adopt more compact conformations.

Whereas system N400 has a larger $r_g$ compared to system N300 due to the more flexible termini regions, system NM400 was observed to have a smaller $r_g$ compared to system NM300. The increased kinetic energy due to the temperature increase allowed the
protein to adopt a more compact conformation, compared to the NM300 system, while still being tethered to chain-A. The increase in flexibility is also evident in the higher variance exhibited in the NM400 system, compared to the NM300 system.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM300</td>
<td>11.2 ± 0.2</td>
</tr>
<tr>
<td>NM400</td>
<td>10.55 ± 0.35</td>
</tr>
</tbody>
</table>

Table 3.9: Radius of gyration of each monomeric chain-B system under no field conditions, averaged over data collection stage, after the first 1 ns.

### 3.6.5 Solvent accessible surface area

Table 3.10 shows the percentage change of SASA for monomeric insulin chain-B systems, relative to the NM300 system. The values are calculated for the whole data collection stage after the first 1 ns. Chain-A has been ignored in this calculation, so that only the SASA of chain-B was considered.

<table>
<thead>
<tr>
<th>System name</th>
<th>Percentage SASA (relative to NM300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM300</td>
<td>100.0 ± 1.8</td>
</tr>
<tr>
<td>NM400</td>
<td>94.1 ± 4.4</td>
</tr>
</tbody>
</table>

Table 3.10: Percentage of solvent accessible surface area of chain-B under monomeric condition throughout the data collection stage after the first 1 ns, relative to the NM300 simulation.

The SASA of the thermally stressed system was shown to decrease by 6% compared to the NM300 system, and qualitatively agrees well with the results obtained from $r_g$ calculations, where a decrease in overall size was observed.

### 3.7 Conclusions

In this chapter, we have explored the behaviour of insulin chain-B in both isolated form and in the presence of insulin chain-A with all disulfide bonds intact, under ambient conditions and under the application of thermal stress.
Due to its importance as a benchmark for all the subsequent simulations, multiple simulations of isolated chain-B were performed to collect rich statistics. These were compared with the thermally stressed system and both data sets used to get insights into possible biological implications of the inherent flexibility of insulin, including the conformation dependency of receptor binding. The results highlight two distinct pathways with a common residue, Gly B8, playing a key role in initiating the structural change over the course of the simulations. One of these pathways leads to the formation of the T-like state, which is believed to be the active state of insulin.

The comparisons also highlight the importance of packing interactions in determining the conformational behaviour of chain B of insulin, producing conformations stipulated by localised hydrophobic interactions. We were able to predict features observed in the NMR structure, including generally satisfying NMR experimental NOE distant restraints better than the crystal structure.

On the other hand, insulin chain-B under monomeric conditions was observed to be very stable, owing to the presence of the disulfide bonds which tether the peptide to chain-A. The presence of these disulfide bonds is important in order to preserve the secondary structure of the monomeric insulin and also prevents insulin chain-B to undergo the helix-breaking process within the simulation timeframe.

The application of thermal stress was shown to increase the conformational dynamics of both isolated and monomeric insulin chain-B by increasing the mobility and flexibility of the termini regions. This is evident from the overall analysis which showed conserved B9–B19 helical core, while at the same time showed secondary structure disruption at both termini, which changed the overall size of the whole protein.

Most importantly, the observed experimentally consistent behaviour of insulin under the simulation conditions used in this chapter provided significant confidence for using the MD simulations to obtain insights into the protein’s response to applied electric field.
Chapter 4

Insulin Simulations:
Static Electric Field Stress

New opinions are always suspected, and usually opposed, without any other reason but because they are not already common.

*John Locke, English philosopher (1632–1704)*

4.1 Overview

In this chapter, the effects of static electric field on protein behaviour are investigated. In Section 4.2, this thesis describes the implementation of the static electric field. The details of simulation conditions for each system under the static fields are also provided. In Section 4.3, we focus on the effects of static electric field of various strengths on isolated chain-B immersed in water, and compare the results to the reference systems. In Section 4.4, the effects of static electric fields of various strengths on chain-B in the monomeric condition (i.e. in the presence of chain-A with all disulfide bonds present) are investigated and compared to the results obtained from Section 4.3 and the reference systems. Finally, the conclusion is presented in Section 4.5.
4.2 Methodologies

The electric field stress was applied in order to investigate the nature of interaction between the field and insulin chain-B. Within 
\textbf{NAMD}, the electric field is applied directly to the atoms as an external force, as discussed in Section 2.4.2. This simplifies the force calculations on the atoms, as effects such as charge screening from the surrounding environment are not considered.

In this study, the static electric field strengths considered range between $1806 \text{ V/m}$ to $10^9 \text{ V/m}$. The highest value corresponds to the electric field strength where previous molecular dynamics simulations observed structural changes in the water structure [251]. The lowest value ($1806 \text{ V/m}$) was taken from the results of an experiment conducted to determine the effect of the application of microwave radiation to various proteins [249].

In this experiment, a 2.45 GHz, 800 W domestic microwave oven was modified to deliver pulses of radiation every 6 minutes, in accordance to the guidelines given by the International Commission for Non Ionising Radiation Protection (ICNIRP) and the Australian Radiation Protection and Nuclear Safety Agency (ARPANSA) radiation protection standard [305]. The radiation pulses cause a temperature increase in the protein, which can be converted into the Specific Absorption Rate (SAR), a measure of absorbed power per unit mass. The SAR is related to the strength of the electric field component of an electromagnetic wave by the formula:

$$\text{SAR} = \frac{\sigma E^2}{\rho}$$

where SAR is the specific absorption rate (in W/kg), $\sigma$ is the specific conductivity of media (in S/m), $\rho$ is the mass density of media (in kg/m$^3$), and $E$ is the electric field strength (in V/m).

Given the main component of the simulation box is water, we can use the following values: $\rho = 997.05 \text{ kg/m}^3$ (at $25 \degree \text{C}$) [306], and $\sigma = 0.0550 \text{ mS/cm}$ (at $25 \degree \text{C}$) [307]. Utilising Equation 4.1, the maximum value of SAR obtained from the experiment (18 W/kg) can be converted into the equivalent electric field strength of $1806 \text{ V/m}$. It should be noted
that this SAR value exceeds the maximum allowable SAR value in Australia, as defined in the Australian Standard AS/NZS 2772.1 and implemented by the Australian Communications Authority Standard (ACA RS 1999) [308], set to 1.6 W/kg in the frequency range of 800–2500 MHz and averaged over 1 g of tissue.

A complete list of the conditions considered in this chapter is given in Table 4.1. The letter “S” denotes static electric field, and the letter “M” denotes monomeric insulin (i.e. chain-A is present, and the disulfide bonds between the two chains left intact).

<table>
<thead>
<tr>
<th>Isolated insulin chain-B</th>
<th>Monomeric insulin</th>
<th>Electric field strength (V/m)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 1.7</td>
<td>SM1 10</td>
<td>10^7</td>
<td>Observable effects in previous studies [251]</td>
</tr>
<tr>
<td>S2 5.9</td>
<td>SM2 10</td>
<td>5 × 10^8</td>
<td></td>
</tr>
<tr>
<td>S3 7.2</td>
<td>SM3 10</td>
<td>10^8</td>
<td></td>
</tr>
<tr>
<td>S4 10</td>
<td>SM4 10</td>
<td>5 × 10^7</td>
<td></td>
</tr>
<tr>
<td>S5 10</td>
<td>SM5 10</td>
<td>10^7</td>
<td></td>
</tr>
<tr>
<td>S6 10</td>
<td>-</td>
<td>5 × 10^6</td>
<td>Experimental work on citrate synthase [249]</td>
</tr>
<tr>
<td>S7 10</td>
<td>-</td>
<td>10^6</td>
<td></td>
</tr>
<tr>
<td>S8 10</td>
<td>-</td>
<td>10^5</td>
<td></td>
</tr>
<tr>
<td>S9 6</td>
<td>-</td>
<td>1806</td>
<td>Experimental work indicated effects in some proteins [249]</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of simulation conditions used in this chapter.
CHAPTER 4. INSULIN SIMULATIONS: 
STATIC ELECTRIC FIELD STRESS

4.3 Isolated insulin chain-B

4.3.1 Secondary structure

Figure 4.1 shows the time evolution of the secondary structure of the systems described in Table 4.1, taken using the STRIDE algorithm as implemented in VMD.

Figure 4.1: Secondary structure evolution of the isolated chain-B systems under static electric field. (continued on next page)
CHAPTER 4. INSULIN SIMULATIONS:
STATIC ELECTRIC FIELD STRESS

(g) System S7

(h) System S8

(i) System S9

Figure 4.1: Secondary structure evolution of the isolated chain-B systems under static electric field. (continued)

The majority of the systems retained most of the helical residues, except for the system at 10^9 V/m (system S1) where disruption to the N-terminal end of the B9–B19 helix took place, as illustrated in Figure 4.2. Interestingly, helix B12–B19 was still conserved to the end of the simulation.

At field strengths of 5 \times 10^7 V/m or lower, there was a tendency for the systems to undergo transitions between \( \alpha \)-helix and \( \pi \)-helix. This behaviour is consistent with the observation made with the reference systems, discussed in Section 3.5.3, and appears to be a direct consequence of the reduction of the disulfide bond tethers that normally bind chain-B to chain-A at residues B7 and B19. These bonds immobilise the central B9–B19 helix and help stabilise the end termini. Without these disulfide bonds, the ends of the central helix can freely undergo transitions between \( \alpha \)-helix and \( \pi \)-helix. Interestingly, the transitions were not observed in system S1, S2, and S3; and were only lightly observed in
systems S4 and S5. This indicates a possible role of electric field on the conformational behavior of a protein.

The active state of insulin is believed to be the T-state, which is characterised by the extended B1–B8 residues [281]. In order to form this state, it is necessary for insulin in the R- or R₁-state to first break its central helix and then fully extend residues B1–B8. This transition is centred at Gly B8, which is in keeping with the known propensity of glycine residues as helix breakers [6]. It should be noted that this event is only observed in two systems studied here, systems S5 and S7. In addition, systems S2 and S3 also exhibit reduced conformational flexibility, evident from the secondary structure maps. This indicates that the application of strong static electric field interferes with the helix-breaking mechanism and restricts the conformational dynamics of isolated insulin chain-B.

4.3.2 Dipole moment distribution

Insulin chain-B possesses a strong dipole moment due to the helical nature of the central region, and therefore has the potential to be affected by the application of an electric field. As the electric field is applied, the peptide is induced to align itself with respect to the field. Figure 4.3 shows the distribution of dipole moment values of the B9–B19 helix in the applied field direction (z-axis) for all the systems over the data collection stage. The
dipole moment distributions have been normalised to allow direct comparison between different electric field strengths.

![Figure 4.3: B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under static electric field. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.](image)

The static field systems consistently produced preferential alignment of the helix in the direction of the applied field, as illustrated by the well-pronounced peaks of the corresponding dipole moment distributions. The peak heights and widths are correlated with the strength of the applied electric field. Stronger field strengths yielded larger and narrower peaks, whereas lower field strengths yielded smaller and broader peaks. The representative conformations for the systems under applied static electric fields are shown in Figure 4.4 to demonstrate the observed alignments. The reference systems are also shown for comparison.

### 4.3.3 PEPCAT analysis

The conformational analyses were performed using the program PEPCAT on systems S4, S5, along with the reference systems N300A and N400 in order to investigate the conformational dynamics of each system. System N300A was chosen as a representative of
CHAPTER 4. INSULIN SIMULATIONS:  
STATIC ELECTRIC FIELD STRESS

Figure 4.4: Snapshot of the last frame of the simulations of isolated chain-B systems, showing the alignment of the helix with respect to the electric field applied in the z-direction. N-terminus is coloured blue, C-terminus is coloured red, and the B9–B19 helix is highlighted in black.

the system at ambient conditions. The electrical field stressed systems were selected in order to examine the more subtle disruption to the secondary structure compared to the systems at higher field strengths, as evident from Figure 4.1. Systems S4 and S5 provide
a representative of the systems with electric field strengths just under those that caused
direct structural damage to the protein structure. The descriptors used in this section are
described in Section 2.7.

The systems were classified into a number of states, which are shown in Table 4.2. For
clarity, transient conformational states that occupy less than 1% of simulation time have
been removed.

<table>
<thead>
<tr>
<th>System name</th>
<th>Electric field strength (V/m)</th>
<th>No. of sampled states</th>
<th>Percentage of time spent within states occupied by system N300A</th>
<th>Percentage of time spent within states occupied by system N400</th>
<th>Percentage of time spent outside states occupied by systems N300A and N400</th>
</tr>
</thead>
<tbody>
<tr>
<td>N300A</td>
<td>0</td>
<td>9</td>
<td>100%</td>
<td>49.3%</td>
<td>0%</td>
</tr>
<tr>
<td>N400</td>
<td>0</td>
<td>30</td>
<td>6.0%</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>S4</td>
<td>$5 \times 10^7$</td>
<td>7</td>
<td>18.8%</td>
<td>74.6%</td>
<td>21.0%</td>
</tr>
<tr>
<td>S5</td>
<td>$10^7$</td>
<td>19</td>
<td>43.3%</td>
<td>76.7%</td>
<td>13.4%</td>
</tr>
</tbody>
</table>

**Table 4.2:** Comparison of sampled states for isolated insulin chain-B under static electric field stress chosen for PEPCAT analysis with systems N300A and N400.

Here we compare the results with the N300A and the N400 systems, notably the
“time” spent in conformational states, and also the relative number of conformational states. All the static field systems were shown to spend more time in conformational states sampled by system N400 rather than by system N300A. This indicates that the time evolution of the system under static electric field stress has more in common with a system under thermal stress than under ambient conditions. However, these systems also spent a large amount of time (up to 21%) in conformational states not observed in either of the reference systems. This behaviour is more prominent in the systems with higher electric field strength, suggesting that the mechanism of interaction between the electric field and a protein differs from the interaction resulting from the thermal stress alone. It is interesting to note that there is a relatively low number of states that exhibit deviation from normal conformations accessible to the reference systems. For example, in system S4, there is only one identified state outside the range of conformational states sampled by the reference systems, yet it accounts for more than 21% of the simulation time.
Figure 4.5: PEPCAT transition maps for systems considered for PEPCAT analysis. The yellow circles represent a unique state ID, with size proportional to the relative population of that state within the simulation timeframe. The grey arrows indicate transitions between states, with size of arrowheads representing the relative number of transitions to that state. Refer to Section 2.7 for full explanation of map features. For clarity, only states appearing more than 10% of the total simulation timeframe are presented.

Figure 4.5 shows the PEPCAT transition maps for the systems considered in this section. Refer to Figure 2.14 in Section 2.7 for detailed explanation of map features. System N400 is shown to have more complex transition pathways, with no preference towards a specific state. This is characteristic of the effect of thermal stress, which enables the
protein to cross over energy barriers on the potential energy landscape. The restriction in conformational flexibility is observed in system S4, which yields a reduced number of conformational states compared to the reference systems. Preference to a specific conformational state (state 203) is observed in this system. System S5 showed an increased number of states and transition complexity, albeit not as extreme as system N400.

![Figure 4.6: Snapshots of the two most occupied states for the isolated insulin chain-B under static electric field stress.](image)

Figure 4.6 show the two most occupied states for the isolated chain-B systems considered for PEPCAT analysis. The figures show that they preserve most of their α-helical region and exhibit less flexibility on both termini regions compared to the N300A system. This α-helix preservation effect is exaggerated in the systems with stronger static electric field (refer to Figure 4.1b and 4.1c) and may result from the coupling of the helical dipole moment with the electric field, as illustrated in Figure 4.3.

### 4.3.4 Root mean square deviation

The structural stability of the B9–B19 helical region under the influence of the applied static fields was investigated by determining the RMSD of the backbone atoms compared to the start of the data collection stage. The plots of the RMSDs as a function of time for each system are shown in Figure 4.7. They show that all the systems considered in this...
chapter reached their equilibrium values relatively quickly. The average RMSD values for the last 1 ns of the simulations are presented in Table 4.3.

![Graphs of RMSD vs Time for Systems S1 to S6](image)

**Figure 4.7:** 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under static electric field. *(continued on next page)*

Similar to the system N300, systems S4, S6, S7, S8, and S9 all exhibited structural transitions towards \( \pi \)-helix (refer to Figure 4.1), resulting in their RMSD values increasing
Figure 4.7: 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under static electric field. (continued)

<table>
<thead>
<tr>
<th>System name</th>
<th>RMSD [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>2.9 ± 0.2</td>
</tr>
<tr>
<td>S2</td>
<td>1.5 ± 0.2</td>
</tr>
<tr>
<td>S3</td>
<td>1.54 ± 0.12</td>
</tr>
<tr>
<td>S4</td>
<td>2.26 ± 0.09</td>
</tr>
<tr>
<td>S5</td>
<td>1.57 ± 0.15</td>
</tr>
<tr>
<td>S6</td>
<td>2.18 ± 0.07</td>
</tr>
<tr>
<td>S7</td>
<td>2.17 ± 0.08</td>
</tr>
<tr>
<td>S8</td>
<td>2.17 ± 0.09</td>
</tr>
<tr>
<td>S9</td>
<td>2.16 ± 0.07</td>
</tr>
<tr>
<td>N300</td>
<td>2.25 ± 0.08</td>
</tr>
<tr>
<td>N400</td>
<td>1.5 ± 0.6</td>
</tr>
</tbody>
</table>

Table 4.3: B9–B19 helix backbone RMSD of the isolated chain-B systems under static electric field, averaged over the last 1 ns.
relative to the other static field systems. As the transitions between \( \alpha \)- and \( \pi \)-helix have been shown to be part of the normal dynamics of a protein, the transitions observed in these systems indicate that under static electric field stress of those strengths, insulin chain-B is still capable of reproducing the conformational dynamics of unstressed protein. In other words, the \( \alpha \)- to \( \pi \)-helix transition is not inhibited by the application of static electric field up to the strength of \( 5 \times 10^7 \) V/m.

In system S1, the secondary structure of the B9–B19 underwent extreme stretching of the N-terminal end of the helix (refer to Figure 4.2). This stretching unfolded this region and caused a much higher RMSD of the B9–B19 helix. It is remarkable, however, to note the resilience of helical residues B12–B19, which remained relatively undamaged, and maintained a relatively constant B9–B19 backbone RMSD of 2.9 Å.

The variances of the RMSDs were shown to be mostly of the same order as the unstressed system. The exceptions to this rule are the S1, S2, S3, and S5 systems, which showed increased variance up to 2.5 times the normal variance of the unstressed system. It is important to note, however, that these variances are less than the variance of the thermally stressed system. Because the variance of a property is directly related to how much individual data points differ from the average value, this indicates that the application of static electric field stress is responsible in making insulin chain-B adopt conformations that deviate from the average conformation more than in the unstressed system at the same temperature.

### 4.3.5 Radius of gyration

The average \( r_g \) of each isolated chain-B system under static electric field after the first 1 ns was calculated and the results are presented in Table 4.4. It can be seen that most of the systems under static electric field exhibit slightly larger \( r_g \) compared to system N300. However, there was no significant deviation from the values obtained from the reference systems until a field strength of \( 10^9 \) V/m was reached.

The variances of the static field systems were generally smaller than the reference systems, with decreasing trend as the field gets stronger. This agrees with the previous
analysis suggesting a stronger coupling between the helical dipole moment of the protein with the applied electric field at higher field strengths. This dipole moment coupling limits the conformations that can be accessed by the protein, as evident from Table 4.2, where system S4 adopts 7 conformational states compared to system S5 with 19 states sampled.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>15.8 ± 0.3</td>
</tr>
<tr>
<td>S2</td>
<td>11.3 ± 0.3</td>
</tr>
<tr>
<td>S3</td>
<td>10.4 ± 0.2</td>
</tr>
<tr>
<td>S4</td>
<td>10.1 ± 0.4</td>
</tr>
<tr>
<td>S5</td>
<td>10.5 ± 0.7</td>
</tr>
<tr>
<td>S6</td>
<td>9.6 ± 0.4</td>
</tr>
<tr>
<td>S7</td>
<td>9.5 ± 0.2</td>
</tr>
<tr>
<td>S8</td>
<td>11.7 ± 0.9</td>
</tr>
<tr>
<td>S9</td>
<td>10.1 ± 0.4</td>
</tr>
<tr>
<td>N300</td>
<td>9.9 ± 0.5</td>
</tr>
<tr>
<td>N400</td>
<td>11.0 ± 1.3</td>
</tr>
</tbody>
</table>

Table 4.4: Radius of gyration of each isolated chain-B system under static electric field, averaged over data collection stage, after the first 1 ns.

4.3.6 Helical content

The number of retained helical residues of each isolated insulin chain-B system under the application of static electric field was calculated using the STRIDE algorithm [256], as a measure of the stability of the helix over the data collection stage. The numbers of all-helical (consisting of $\alpha$-, $\pi$-, and $3_{10}$-helices) and $\alpha$-helical residues of insulin chain-B retained over the whole data collection stage are presented in Table 4.5.

The data showed that the number of retained all-helical and $\alpha$-helical residues for most systems are closer to the values for the N300 system than the N400 system. This is consistent with the previous findings that showed insulin chain-B retaining its normal conformational dynamics even at relatively high field strengths.
CHAPTER 4. INSULIN SIMULATIONS: STATIC ELECTRIC FIELD STRESS

Table 4.5: Retained number of all-helical and α-helical residues for isolated insulin chain-B under static field, averaged over the data collection stage.

<table>
<thead>
<tr>
<th>System name</th>
<th>Retained all-helical residues</th>
<th>Retained α-helical residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>5.8 ± 2.7</td>
<td>5.3 ± 3.1</td>
</tr>
<tr>
<td>S2</td>
<td>12.7 ± 1.8</td>
<td>12.5 ± 2.1</td>
</tr>
<tr>
<td>S3</td>
<td>12.7 ± 1.3</td>
<td>12.7 ± 1.3</td>
</tr>
<tr>
<td>S4</td>
<td>12.3 ± 1.3</td>
<td>9.3 ± 2.8</td>
</tr>
<tr>
<td>S5</td>
<td>9.9 ± 2.6</td>
<td>8.9 ± 3.1</td>
</tr>
<tr>
<td>S6</td>
<td>12.9 ± 1.9</td>
<td>6.3 ± 4.6</td>
</tr>
<tr>
<td>S7</td>
<td>11.4 ± 2.6</td>
<td>5.4 ± 4.2</td>
</tr>
<tr>
<td>S8</td>
<td>13.0 ± 2.1</td>
<td>4.5 ± 4.3</td>
</tr>
<tr>
<td>S9</td>
<td>12.4 ± 2.0</td>
<td>5.9 ± 4.5</td>
</tr>
<tr>
<td>N300</td>
<td>10.7 ± 2.7</td>
<td>8.3 ± 3.9</td>
</tr>
<tr>
<td>N400</td>
<td>8.8 ± 2.4</td>
<td>7.9 ± 2.7</td>
</tr>
</tbody>
</table>

Significant disruption of the helical content was observed at the highest field strength, and is consistent with the observation of the secondary structure (Figure 4.1a), $r_g$, and visual inspection of the protein (Figure 4.2).

Systems S2 and S3 suggest a degree of helix stabilisation present by the application of moderately high field strengths. As discussed previously, helix stabilisation is a consequence of the coupling between the dipole moment of the protein and the applied electric field. This coupling preserves the α-helical residues and disallows conformational transition towards π-helix as can be observed in the other systems. This conformational restriction is supported by the low variance exhibited in both of these systems, shown in Table 4.5.

In order to examine the changes to the central B9–B19 helix in terms of the number of helical residues, the plots of all-helical and α-helical residues of each insulin chain-B system under static electric field as a function of time are presented in Figure 4.8. The plots have been averaged over 100 ps for clarity.

Figure 4.8 indicates that while the all-helical residues remain roughly constant at 80% (8.8 residues), there is a gradual loss of α-helical content at lower field strength (below $5 \times 10^7$ V/m), with an average retained α-helical content of 20% (2.2 residues).

In the systems with moderately high field strengths (above $10^7$ V/m), the amount of α-
helix dominates the helical content. Systems S1, S2, and S3 do not exhibit the transitions between $\alpha$- to $\pi$-helix observed at lower field strengths. Systems S4 and S5 were also observed to have minimal $\alpha$- to $\pi$-helix transition. This indicates the restrictions in the
conformational dynamics of these systems. System S3, in particular, shows enhancement in the helical content, which further reinforces the stabilising effect of the static electric field to the protein. The rest of the systems exhibit conformational dynamics that closely resemble the dynamics of the reference systems with similar changes in the $\alpha$-helical region.

The percentage of retained backbone hydrogen bonds for systems S1, S2, and S3 are shown in Figure 4.9. The plots have been averaged over 100 ps for clarity. The loss of helical content in the B9–B19 helix of system S1 is correlated with the loss of hydrogen bonds between the backbone atoms. The retainment of helical content in system S3 is correlated with the enhancement in the backbone hydrogen bonds.
4.3.7 Solvent accessible surface area

The SASA was calculated using the program X-PLOR for the whole data collection stage after the first 1 ns. For clarity, the resulting SASA and variance are expressed as a percentage of the SASA of the N300A system, and are presented in Table 4.6.

The results showed that the SASA for static field systems did not change noticeably until a field strength of $10^9$ V/m was reached, the variance dropped significantly from field strength of $10^8$ V/m. This confirms the severe restriction of the mobility of the protein under applied static electric field.
| System name | Percentage SASA  
relative to N300A |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>123.6 ± 1.3</td>
</tr>
<tr>
<td>S2</td>
<td>106.4 ± 2.1</td>
</tr>
<tr>
<td>S3</td>
<td>101.2 ± 1.7</td>
</tr>
<tr>
<td>S4</td>
<td>102.9 ± 5.8</td>
</tr>
<tr>
<td>S5</td>
<td>106.7 ± 4.2</td>
</tr>
<tr>
<td>S6</td>
<td>99.0 ± 4.4</td>
</tr>
<tr>
<td>S7</td>
<td>97.4 ± 3.3</td>
</tr>
<tr>
<td>S8</td>
<td>108.6 ± 2.9</td>
</tr>
<tr>
<td>S9</td>
<td>104.1 ± 4.2</td>
</tr>
<tr>
<td>N300A</td>
<td>100.0 ± 5.8</td>
</tr>
<tr>
<td>N400</td>
<td>107.7 ± 6.1</td>
</tr>
</tbody>
</table>

Table 4.6: Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation.

4.4 Chain-B in monomeric environment

4.4.1 Secondary structure

The secondary structure evolution of insulin chain-B under monomeric condition is presented in Figure 4.10. In contrast to the isolated chain-B, the secondary structure of the monomeric chain-B is very stable, with no significant disruption observed until a field strength of $5 \times 10^8$ V/m was reached. None of the systems was found to undergo $\alpha$- to $\pi$-helix transitions. This is most likely due to the presence of the disulfide bonds which tether the end termini of the helical region, thus restricting its flexibility. Interestingly, systems SM1 and SM2 experience helix breaking at the valine residue B12, similar to system S1 for the isolated chain-B.
CHAPTER 4. INSULIN SIMULATIONS: STATIC ELECTRIC FIELD STRESS

Figure 4.10: Secondary structure evolution of the monomeric chain-B systems under static electric field.
4.4.2 Dipole moment distribution

The monomeric insulin contains three helical regions—in addition to the helical residues of chain-B, helices A1–A8 and A13–A20 are also present. As the electric field is applied, the dipole moments of these three helices interact with each other. This interaction may result in the alignment of the helices in the same direction. Under this circumstance, the combined dipole moment is stronger than the dipole moment of each individual helix.

The dipole moments in the $z$-direction for the monomeric insulin chain-B under static electric field stress are plotted in Figure 4.11. The plot shows that the distribution of the dipole moment for monomeric insulin chain-B is sharper and less spread, especially in the strongest field strength of $10^9$ V/m. This is a direct consequence of the stronger overall dipole moment possessed by the monomeric insulin compared to isolated chain-B. In the higher field strength systems, all three helices of the monomeric insulin point in the same direction, resulting in a stronger coupling between the protein and the applied electric field. Because of the stronger coupling, monomeric insulin chain-B is able to align itself with respect to the field at a lower strength than when isolated from chain-A.

\[\text{Figure 4.11: B9–B19 helical dipole moment distribution in the $z$-direction of insulin chain-B in the presence of chain-A. Each plot has been displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.}\]
CHAPTER 4. INSULIN SIMULATIONS: STATIC ELECTRIC FIELD STRESS

Figure 4.12 shows a comparison between the distribution of the dipole moment in the z-direction between isolated and monomeric insulin chain-B at different electric field strengths.

![Figure 4.12: Illustration of the stronger coupling between the electric field and the chain-B helical dipole moment. For each electric field strength, the dipole moment in the z-direction for monomeric insulin chain-B (red) is stronger than for isolated chain-B (green). Each of the plots has been displaced by 0.2 for clarity.](image)

The strong correlation between the helical dipole moment and the electric field is also evident from visual inspection of the preferred orientation adopted by the protein under stress, as shown in Figure 4.13. In all cases, the monomeric insulin chain-B shows stronger alignment with the electric field, applied in the z-direction, compared to the isolated chain-B.

4.4.3 Root mean square deviation

Figure 4.14 shows the RMSD of the B9–B19 helix backbone atoms from each simulation under static electric field, calculated relative to the start of the data collection stage. The plots have been averaged over 100 ps for clarity. They show that all the systems considered in this study reach their equilibrium values relatively quickly. Once they reach their equilibrium values, the RMSDs stay relatively constant.
Figure 4.13: Snapshot of the last frame of the simulations of monomeric chain-B (left of each subfigure) systems and isolated chain-B (right of each subfigure), showing the stronger alignment of the monomeric chain-B helix with respect to the electric field applied in the $z$-direction, compared to the isolated chain-B helix. N-terminus is coloured blue, C-terminus is coloured red, and the B9–B19 helix is highlighted in black. Chain-A is present throughout the simulations of monomeric insulin chain-B, but is omitted from the diagrams for clarity.
Figure 4.14: 100 ps running average of the B9–B19 helix backbone RMSD of the monomeric chain-B systems under static electric field.

The average RMSD values for the last 1 ns of the simulations are presented in Table 4.7. The RMSDs of the systems SM3, SM4, and SM5 are almost identical to the RMSDs of the reference systems. This indicates that the static electric field of moderate strength has very little effect on the conformation of the helical region of monomeric insulin chain-B.
CHAPTER 4. INSULIN SIMULATIONS: STATIC ELECTRIC FIELD STRESS

**Table 4.7:** B9–B19 helical backbone RMSD of each monomeric chain-B simulation under static electric field, averaged over the last 1 ns.

<table>
<thead>
<tr>
<th>System name</th>
<th>RMSD [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM1</td>
<td>3.5 ± 0.1</td>
</tr>
<tr>
<td>SM2</td>
<td>2.5 ± 0.2</td>
</tr>
<tr>
<td>SM3</td>
<td>0.62 ± 0.09</td>
</tr>
<tr>
<td>SM4</td>
<td>0.65 ± 0.07</td>
</tr>
<tr>
<td>SM5</td>
<td>0.68 ± 0.08</td>
</tr>
<tr>
<td>NM300</td>
<td>0.56 ± 0.09</td>
</tr>
<tr>
<td>NM400</td>
<td>0.7 ± 0.1</td>
</tr>
</tbody>
</table>

The electric field only has an effect at very high field strength. It is notable, however, that the application of electric field of strength $5 \times 10^8$ V/m disrupted the conformation of monomeric insulin chain-B to a larger extent than it did to the isolated chain-B. The reason for this is the presence of two helical regions of chain-A, which enhances the effects of the electric field due to the coupling of all helix dipole moments.

### 4.4.4 Radius of gyration

The $r_g$ of the monomeric chain-B systems are presented in Table 4.8. Chain-A has been ignored in this calculation, so that only the $r_g$ of chain-B was considered. Similar to the $r_g$ of the isolated chain-B systems under static electric field, monomeric chain-B systems also exhibit larger $r_g$ compared to the reference systems. This indicates the added stress exerted by the electric field. For the two highest electric field strengths (systems SM1 and SM2), the $r_g$ became very large with significant disruption to the secondary structure.

**Table 4.8:** Radius of gyration of each monomeric chain-B system under applied static electric field, averaged over data collection stage, after the first 1 ns.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM1</td>
<td>19.4 ± 0.4</td>
</tr>
<tr>
<td>SM2</td>
<td>14.9 ± 0.3</td>
</tr>
<tr>
<td>SM3</td>
<td>11.8 ± 0.7</td>
</tr>
<tr>
<td>SM4</td>
<td>12.2 ± 0.5</td>
</tr>
<tr>
<td>SM5</td>
<td>10.9 ± 0.2</td>
</tr>
<tr>
<td>NM300</td>
<td>11.2 ± 0.2</td>
</tr>
<tr>
<td>NM400</td>
<td>10.55 ± 0.35</td>
</tr>
</tbody>
</table>
4.4.5 Helical content

Table 4.9 presents the retained number of α-helical residues averaged over the whole data collection range of the simulations. Only the number of α-helical residues are presented, since virtually no transition towards π-helix was observed.

System name | Retained α-helical residues
---|---
SM1 | 1.7 ± 2.9
SM2 | 10.0 ± 4.4
SM3 | 14.4 ± 1.6
SM4 | 15.3 ± 1.1
SM5 | 16.0 ± 0.3
NM300 | 16.0 ± 0.2
NM400 | 13.5 ± 1.6

**Table 4.9:** Retained number of α-helical residues for monomeric insulin chain-B under static electric field, averaged over the data collection stage.

The result confirms the secondary structure analysis, which shows progressive loss of α-helical region as the field strength is increased. The most severe disruption to the helical region occurs with the application of electric field above $10^8$ V/m. Interestingly, the number of residues lost due to increased temperature is comparable to the number of residues lost at field strength slightly larger than $10^8$ V/m.

Figure 4.15 shows the percentage of retained helical content present in residues B9–B19 of chain-B, relative to the start of data collection stage. The plots have been averaged over 100 ps for clarity. In contrast to the results for isolated chain-B, there is virtually no π-helix component observed in monomeric chain-B. The helical content is conserved to the end of the simulations for systems SM3 to SM5. System SM1 exhibits total loss of the helical content within 2 ns, and system SM2 exhibits 40% loss of helical content within 2.5 ns. The disruptions experienced by these two systems are most likely caused by the presence of the two additional helices from the accompanying chain-A, which contributes to the enhancement of dipole moment of the whole protein. As a consequence, the SM2 system experienced more stretching by the electric field compared to the S2 system. In
fact, the loss of the helical content is similar to the one experienced by system S1 (compare Figures 4.15b and 4.8a).

![Graphs showing helical content retention over time for different systems](image)

**Figure 4.15:** 100 ps running average of the helical content retained as a function of time for the monomeric chain-B systems under static electric field. Red lines indicate the all-helical content, and the green lines indicate the α-helical content.
The loss of helical content in systems SM1 and SM2 is correlated to the loss of retained backbone hydrogen bonds, as shown in Figure 4.16. The plots have been averaged over 100 ps for clarity.

![Graphs showing hydrogen bond retention over time for systems SM1 and SM2.](Image)

**Figure 4.16:** 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for systems SM1 and SM2.

### 4.4.6 Solvent accessible surface area

Table 4.10 shows the percentage change of SASA for monomeric insulin chain-B systems under static electric field stress, relative to the NM300 system. The values are calculated for the whole data collection stage after the first 1 ns. Chain-A has been ignored in this calculation, so that only the SASA of chain-B was considered.

<table>
<thead>
<tr>
<th>System name</th>
<th>Percentage SASA (relative to NM300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SM1</td>
<td>121.5 ± 1.3</td>
</tr>
<tr>
<td>SM2</td>
<td>110.4 ± 1.5</td>
</tr>
<tr>
<td>SM3</td>
<td>98.9 ± 2.8</td>
</tr>
<tr>
<td>SM4</td>
<td>100.9 ± 2.4</td>
</tr>
<tr>
<td>SM5</td>
<td>95.0 ± 1.9</td>
</tr>
<tr>
<td>NM300</td>
<td>100.0 ± 1.8</td>
</tr>
<tr>
<td>NM400</td>
<td>94.1 ± 4.4</td>
</tr>
</tbody>
</table>

**Table 4.10:** Percentage of solvent accessible surface area of chain-B under monomeric condition throughout the data collection stage after the first 1 ns, relative to the NM300 simulation.
CHAPTER 4. INSULIN SIMULATIONS: STATIC ELECTRIC FIELD STRESS

The results show that the SASA stayed relatively constant for systems SM3 to SM5. Systems SM1 and SM2 showed much larger SASA compared to the reference systems, as a consequence of the stronger dipole moment coupling with the electric field due to the presence of chain-A. The stronger coupling led to the stretching of the helical region which caused the overall surface accessible area to increase. This is in agreement with previous analysis on the retained number of helical residues, which showed loss of helical regions in systems S1 and S2 due to the stretching of those regions.

The variance of the SASA for the static field systems was shown to be of the same order as the NM300 system. There is a tendency for the variance to decrease as the electric field strength is increased. This is indicative of the restrictive effect of the helical dipole moment, as discussed previously.

4.5 Conclusion

In this chapter, we have explored and compared the effect of static electric field of various strengths on the behaviour of both isolated and monomeric insulin chain-B.

The application of static electric field was shown to affect the behaviour of the protein as it aligns itself with respect to the electric field direction. This alignment arises because of the coupling between the electric field and the protein’s helical dipole moment. For simple proteins, with few long helical regions, this alignment may also induce the strengthening of the α-helix, thus hindering flexibility and can possibly prevent important conformational changes such as the transition to an active state or binding of other molecules. This transition was inhibited by the application of electric fields over $5 \times 10^7$ V/m.

When helical proteins are in the vicinity of other helical proteins, their dipole moments can affect one another, resulting in variations in the field strength necessary to align the proteins with respect to the field. This finding will be exploited in our future study of protein aggregation by inducing several proteins to align themselves with respect to the applied static electric field, due to the combined effects of their helical dipole moments.
The application of very strong fields has been shown to lead to extreme disruption of the secondary structure, due to the forceful stretching of the termini as a result of alignment with respect to the field, which causes the breaking of the hydrogen bonds holding the helical structure. Nevertheless, the helices are shown to be very resilient and in our study, only electric fields of at least $5 \times 10^8$ V/m were able to cause damage to the monomeric insulin chain-B. In the case of isolated chain-B, the field strength necessary is even stronger, at $10^9$ V/m.
Chapter 5

Insulin Simulations:
Oscillating Electric Field Stress

See first, think later, then test. But always see first. Otherwise you will only see what you were expecting. Most scientists forget that.

Douglas Adams (1952–2001)

5.1 Overview

In this chapter, the effects of oscillating electric field on protein behaviour are investigated. In Section 5.2, the implementation of the 2.45 GHz oscillating electric field is explained in detail, along with the details of simulation conditions for each system under oscillating fields. In Section 5.3, the effects of oscillating electric field of various strengths to isolated chain-B immersed in water are studied and compared to the results obtained from Section 4.3 and the reference systems. In Section 5.4, the effects of oscillating electric field of various strengths to chain-B in the monomeric condition (i.e. in the presence of chain-A with all disulfide bonds present) are investigated and compared to the results from the Sections 4.4 and 5.3, along with the reference systems. Finally, the conclusion is presented in Section 5.5.
5.2 Methodologies

Having tested the effects of static electric fields on the protein’s behaviour, the logical extension is to apply time varying electric fields. The maximum field strength of each simulation with oscillating fields ($E_{\text{max}}$) was chosen such that the field’s root mean square values ($E_{\text{RMS}}$) correspond to the maximum field strength used in the corresponding simulation with static electric field. This ensures equivalent power is delivered to both system.

The formula used to calculate $E_{\text{RMS}}$ is as follows:

$$E_{\text{RMS}} = \frac{E_{\text{max}}}{\sqrt{2}} \quad (5.1)$$

Several electric field frequencies were chosen for this work. The primary frequency is 2.45 GHz, which corresponds to the frequency of conventional microwave ovens, wireless, and Bluetooth devices. This frequency has been used in previous experimental work investigating the effect of microwave radiation on various proteins [249]. As a comparison, most digital mobile phones operate at frequencies ranging between 1.8–1.9 GHz with the third generation systems using a frequency of 2.2 GHz.

The oscillating field was implemented in NAMD by changing the field strength every 10 simulation steps (corresponding to 20 fs), illustrated in Figure 5.1. This value is short enough so that over one period of the electric field oscillation (408.16 ps), the change of electric field strength can be considered as continuous. Figure 5.2 illustrates the comparison between a static field and 2.45 GHz oscillating field.

The full list of the simulations performed is shown in Table 5.1. The letter “O” denotes oscillating electric field, and the letter “M” denotes monomeric conditions (i.e. chain-A is present, and the disulfide bonds between the two chains left intact).

Simulations using frequencies of 1.225 GHz and 4.9 GHz (half and twice the original frequency of 2.45 GHz, respectively) were also performed in order to investigate the effect of different frequencies on the protein’s conformation, and are discussed in Chapter 6.
Figure 5.1: The oscillating electric field was approximated by changing the electric field strength every 20 fs.

Figure 5.2: Comparison of static and oscillating electric fields applied to the system. One period of the oscillation is highlighted in black.

5.3 Isolated insulin chain-B

5.3.1 Secondary structure

Figure 5.3 shows the secondary structure evolution of the systems described in Table 5.1. In contrast to the results for the isolated chain-B systems under static field shown in Figure 4.1, the application of oscillating fields with strengths equal to or above $5 \times 10^8$ V/m (systems O1 and O2) severely disrupts the helical region of the protein. In system O1, all helical content was lost within 1.2 ns (2.8 electric field oscillations), while in
CHAPTER 5. INSULIN SIMULATIONS: OSCILLATING ELECTRIC FIELD STRESS

<table>
<thead>
<tr>
<th>Isolated insulin chain-B</th>
<th>Monomeric insulin</th>
<th>Electric field strength, $E_{RMS}$ (V/m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>System name</td>
<td>Simulation length (ns)</td>
<td>System name</td>
</tr>
<tr>
<td>O1</td>
<td>6</td>
<td>OM1</td>
</tr>
<tr>
<td>O2</td>
<td>6</td>
<td>OM2</td>
</tr>
<tr>
<td>O3</td>
<td>10</td>
<td>OM3</td>
</tr>
<tr>
<td>O4</td>
<td>10</td>
<td>OM4</td>
</tr>
<tr>
<td>O5</td>
<td>10</td>
<td>OM5</td>
</tr>
<tr>
<td>O6</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>O7</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>O8</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>O9</td>
<td>6</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5.1: Summary of simulation conditions used in this chapter (frequency of 2.45 GHz).

system O2, all helical content was lost within 2.2 ns (5.3 electric field oscillations). In contrast, systems S1 and S2 retained 23% and 65% of their helical content by the end of the simulations, respectively.

While there were some transitions between \( \alpha \)- and \( \pi \)-helix present, they were not as prevalent as they were in the static field systems. The only exception is system O3, which exhibited constant transitions between the two types of helices, which became more pronounced as the simulation progressed. System O3 is the oscillating field system with the highest electric field strength that exhibited this behaviour. In the static field, the highest strength where this transition was observed is in the S4 system. This indicates that features of the normal dynamic behaviour of the protein are observed at a slightly higher strength under oscillating field, compared to the static field stress.

Compared to the static field system, the oscillating field disrupted the structures of isolated insulin chain-B to a larger extent. This is illustrated by the comparison of typical sampled structures for systems S2 and O2, shown in Figure 5.4.

The helix breaking event at glycine B8 was observed in systems O4, O5, and O6. This illustrates the ability of systems under oscillating electric field stress of up to $5 \times 10^7$ V/m to undergo a process which normally occurs at ambient conditions. In particular, system O4 demonstrated the ability of a stressed system to form a T-like state, where the B4–B7 residues are pointing away from the main body of the protein, and is very similar to the conformation adopted by the N300A system (Figure 5.5).
5.3.2 Dipole moment distribution

Figure 5.6 shows the distribution of dipole moment values of the B9–B19 helix in the applied field direction (z-axis) for all the systems over the data collection stage. The dipole moment distributions have been normalised to allow direct comparison between different electric field strengths. The figure shows that the oscillating fields mostly produce a double
CHAPTER 5. INSULIN SIMULATIONS: OССILLATING ELECTRIC FIELD STRESS

Figure 5.3: Secondary structure evolution of the systems under oscillating electric field. (continued)

Figure 5.4: Comparison of systems S2 and O2, taken at the end of the simulation. System S2 is shown in red and system O2 is shown in green. The B9–B19 residues are shown in yellow.

peak, which is characteristic of the continuous realignment of the helix with respect to the field. In some cases (systems O3, O5, O6, and O7), the distribution is skewed towards one side of the plot. This type of distribution is caused by the large lag time before the protein can respond and align with the applied electric field.
Figure 5.5: Illustration of the T-like state formed by systems N300A (red) and O4 (green). The starting, $R^f$-state (blue) is shown as a comparison. Residues B9–B19 are highlighted in black.

It must be noted, however, that this does not depend on the initial configuration of the protein. For example, the initial dipole moment in the $z$-direction for all the systems studied is 11.95 eÅ. If this initial condition were influencing the dipole moment distribution, then the N300 and N400 systems would not have dipole moment distribution centred around zero.

5.3.3 PEPCAT analysis

PEPCAT analysis was performed on the oscillating field systems that correspond to the static field systems studied in Section 4.3.3 in order to compare the effects of the two types of moderate strength electric fields on protein conformation. The same PEPCAT descriptors were used for the analysis, allowing direct comparison with the results obtained in Section 4.3.3. The classification of the states found using PEPCAT is presented in Table 5.2. For clarity, transient conformational states that occupy less than 1% of simulation time have been removed.

Similar to the systems under static electric field stress, the isolated insulin chain-B systems under oscillating electric field stress were also observed to spend more time
CHAPTER 5. INSULIN SIMULATIONS: 
OSCILLATING ELECTRIC FIELD STRESS

Figure 5.6: B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under oscillating electric field. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.

System name | Electric field strength, $E_{RMS}$ (V/m) | No. of sampled states | Percentage of time spent within states occupied by system N300A | Percentage of time spent within states occupied by system N400 | Percentage of time spent outside states occupied by systems N300A and N400 |
--- | --- | --- | --- | --- | --- |
N300A | 0 | 9 | 100% | 49.3% | 0% |
N400 | 0 | 30 | 6.0% | 100% | 0% |
O4 | $5 \times 10^7$ | 18 | 32.4% | 72.4% | 14.6% |
O5 | $10^7$ | 12 | 70.6% | 83.1% | 11.3% |

Table 5.2: Comparison of sampled states for isolated insulin chain-B under oscillating electric field stress chosen for PEPCAT analysis with systems N300A and N400.

in conformational states sampled by system N400 rather than by system N300A. This further indicates that the time evolution of the system under electric field stress has more in common with a system under thermal stress than under ambient conditions.

Compared to the static electric field of the same strength, the oscillating field systems were observed to spend more time in conformational states sampled by the N300A system. For example, system O4 spent 32.4% of its simulation time in the conformational states sampled by the N300A system, while system S4 spent only 18.8% of its simulation time. In system O5, 70.6% of its simulation time is spent in the conformational states sampled by the N300A system, compared to system S5 with only 43.3%. This may be due to the
brief relaxation periods experienced by the systems under applied oscillating field as the electric field reverses its direction. In addition, stronger electric fields in both systems were shown to decrease the time spent in conformational states sampled by the N300A system, although the time spent in conformational states sampled by the N400 system is relatively unaffected. As there is an overlap between the states sampled by both reference systems, this can be explained by the electric field systems moving out from the local potential energy minima sampled by the N300 system.

Figure 5.7: PEPCAT transition maps for systems considered for PEPCAT analysis. The yellow circles represent a unique state ID, with size proportional to the relative population of that state within the simulation timeframe. The grey arrows indicate transitions between states, with size of arrowheads representing the relative number of transitions to that state. Refer to Section 2.7 for full explanation of map features. For clarity, only states appearing more than 10% of the total simulation timeframe are presented.
Figure 5.7 shows the **PEPCAT** transition maps for the systems considered in this section. Refer to Figure 2.14 in Section 2.7 for detailed explanation of map features. The transition maps of the reference systems are reproduced again for convenience. Both systems O4 and O5 exhibit increased number of states and transition complexity compared to system N300A, although not as extreme as system N400. The similarity between the oscillating systems and the N300 system in terms of the conformational pathways and the occupied states are displayed in the maps. The pathways of the three systems follow a similar trend with conformational state 2 well-represented, and also time spent in the common states 5, 11, 8, 21, and 22. The greater variability of the oscillating systems is also reflected in the occupation of some N400 states (for example, states 44 and 69).

![Figure 5.7: PEPCAT transition maps](image)

**Figure 5.7**: PEPCAT transition maps for the systems considered in this section.

Figure 5.8 shows the two most occupied states in the isolated chain-B systems considered for **PEPCAT** analysis in this section. In contrast to the systems under applied static field, the systems under oscillating field stress showed more flexibility in both termini regions compared to any other system except N400. A probable explanation is the added stress to the protein caused by the constant realignment with respect to the direction of the oscillating electric field.

![Figure 5.8: Snapshots of the two most occupied states](image)

**Figure 5.8**: Snapshots of the two most occupied states for the isolated insulin chain-B under oscillating electric field stress.
5.3.4 Root mean square deviation

The structural stability of the B9–B19 helical region under the influence of the applied oscillating fields was investigated by determining the RMSD of the backbone atoms compared to the start of the data collection stage. The plots of the RMSDs as a function of time for each system are shown in Figure 5.9. The RMSDs exhibited more variability compared to the RMSDs for static field systems. This is as expected due to the oscillatory nature of the electric field. Nevertheless, all the systems reached stable RMSD values by the end of the simulation timeframe. The average RMSD values for the last 1 ns of the simulations are presented in Table 5.3.

<table>
<thead>
<tr>
<th>System name</th>
<th>RMSD [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>4.72 ± 0.55</td>
</tr>
<tr>
<td>O2</td>
<td>4.33 ± 0.35</td>
</tr>
<tr>
<td>O3</td>
<td>2.2 ± 0.1</td>
</tr>
<tr>
<td>O4</td>
<td>1.41 ± 0.15</td>
</tr>
<tr>
<td>O5</td>
<td>1.9 ± 0.4</td>
</tr>
<tr>
<td>O6</td>
<td>1.5 ± 0.1</td>
</tr>
<tr>
<td>O7</td>
<td>2.20 ± 0.08</td>
</tr>
<tr>
<td>O8</td>
<td>2.1 ± 0.1</td>
</tr>
<tr>
<td>O9</td>
<td>1.19 ± 0.25</td>
</tr>
<tr>
<td>N300</td>
<td>2.25 ± 0.08</td>
</tr>
<tr>
<td>N400</td>
<td>1.5 ± 0.6</td>
</tr>
</tbody>
</table>

Table 5.3: B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field, averaged over the last 1 ns, except systems O1 and O2 which were averaged between 2 ns and 3 ns.

Systems O1 and O2 showed complete loss of the helical region, resulting in much larger RMSDs compared to the rest of the systems under oscillating electric field stress. An illustration of the structures of these two systems is shown in Figure 5.10. The helical region of both systems was lost at a much earlier time in the simulation compared to the insulin chain-B systems under static electric field stress.

The α- to π-helix transitions observed in some of the static field systems were also observed in systems O3, O7, and O8. In all of these systems, the RMSDs were shown to be comparable to the N300 system which also exhibited the same structural transitions. This
Figure 5.9: 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field. (continued on next page)

confirms the earlier observation made in the secondary structure discussion (Section 5.3.1).

The variances of the oscillating field systems were shown to be generally larger than their static counterparts. This reflects the constant conformational change experienced by insulin chain-B as it realigns with respect to the electric field direction. However, it
Figure 5.9: 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field. (continued)

is interesting to note that the variances are still lower than the variance of the thermally stressed system.

5.3.5 Radius of gyration

Table 5.4 presents the average $r_g$ of each isolated chain-B system under oscillating electric field after the first 1 ns. Similarly to the static field systems, the application of oscillating electric field was observed to exhibit larger $r_g$ compared to system N300. In contrast with the static electric field systems, the oscillating field disrupted the protein conformation significantly at the lower effective field strength of $5 \times 10^8$ V/m. The systems under applied oscillating field showed an increasing trend of the variance as the field strength increases.
CHAPTER 5. INSULIN SIMULATIONS: OSCILLATING ELECTRIC FIELD STRESS

Figure 5.10: Illustration of the secondary structure disruption to systems O1 and O2. Both structures were obtained at 1.5 ns. The B9–B19 helix is highlighted in orange.

Again, this is due to the protein constantly realigning with respect to the changing field direction.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>12.7 ± 2.3</td>
</tr>
<tr>
<td>O2</td>
<td>13.7 ± 1.3</td>
</tr>
<tr>
<td>O3</td>
<td>10.7 ± 0.8</td>
</tr>
<tr>
<td>O4</td>
<td>9.9 ± 0.5</td>
</tr>
<tr>
<td>O5</td>
<td>10.3 ± 1.0</td>
</tr>
<tr>
<td>O6</td>
<td>9.9 ± 0.4</td>
</tr>
<tr>
<td>O7</td>
<td>9.8 ± 0.5</td>
</tr>
<tr>
<td>O8</td>
<td>9.9 ± 0.2</td>
</tr>
<tr>
<td>O9</td>
<td>10.9 ± 0.5</td>
</tr>
<tr>
<td>N300</td>
<td>9.9 ± 0.5</td>
</tr>
<tr>
<td>N400</td>
<td>11.0 ± 1.3</td>
</tr>
</tbody>
</table>

Table 5.4: Radius of gyration of each isolated chain-B system under oscillating electric field, averaged over data collection stage, after the first 1 ns.
5.3.6 Helical content

The number of retained helical residues of each isolated insulin chain-B system under the application of oscillating electric field was calculated using the STRIDE algorithm [256], as a measure of the stability of the helix over the data collection stage. The retained numbers of all-helical (consisting of α-, π-, and 3_{10}-helices) and α-helical residues of insulin chain-B over the whole data collection stage are presented in Table 5.5.

<table>
<thead>
<tr>
<th>System name</th>
<th>Retained all-helical residues</th>
<th>Retained α-helical residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>1.6 ± 2.9</td>
<td>1.3 ± 2.8</td>
</tr>
<tr>
<td>O2</td>
<td>3.9 ± 4.3</td>
<td>3.5 ± 4.3</td>
</tr>
<tr>
<td>O3</td>
<td>11.8 ± 2.3</td>
<td>8.0 ± 4.3</td>
</tr>
<tr>
<td>O4</td>
<td>12.8 ± 2.3</td>
<td>12.7 ± 2.4</td>
</tr>
<tr>
<td>O5</td>
<td>9.5 ± 2.0</td>
<td>9.4 ± 1.8</td>
</tr>
<tr>
<td>O6</td>
<td>10.6 ± 1.9</td>
<td>10.4 ± 2.3</td>
</tr>
<tr>
<td>O7</td>
<td>12.7 ± 1.7</td>
<td>9.8 ± 3.2</td>
</tr>
<tr>
<td>O8</td>
<td>12.6 ± 1.4</td>
<td>9.4 ± 3.9</td>
</tr>
<tr>
<td>O9</td>
<td>11.4 ± 2.7</td>
<td>10.8 ± 2.9</td>
</tr>
<tr>
<td>N300</td>
<td>10.7 ± 2.7</td>
<td>8.3 ± 3.9</td>
</tr>
<tr>
<td>N400</td>
<td>8.8 ± 2.4</td>
<td>7.9 ± 2.7</td>
</tr>
</tbody>
</table>

Table 5.5: Retained number of all-helical and α-helical residues for insulin chain-B under oscillating field, averaged over the data collection stage.

Systems under oscillating electric field strength of less than or equal to 10^8 V/m showed a similar number of retained all-helical residues to the N300 system. However, these systems in general retained more α-helical residues compared to the N300 system.

Compared to the results for isolated chain-B under static electric field, these systems were observed to have larger variance. This larger variance reflects the constant stretching and relaxation experienced by insulin chain-B as it aligned itself with respect to the electric field direction.

Significant disruption to the helical content was observed at the two highest field strengths, and is consistent with the previous observation of the secondary structures (Figures 5.3a and 5.3b), r_g, and visual inspection (Figure 5.10).
System O4 showed a large retained number of $\alpha$-helices, albeit with a larger variance. Although this is similar to the helix stabilisation observed in the static field system, the large variance means the system does not experience the same restrictions as the static
In order to look at the changes to the central B9–B19 helix in terms of the number of helical residues, the plots of all-helical and α-helical residues of each insulin chain-B system under oscillating electric field as a function of time are presented in Figure 5.11. The plots have been averaged over 100 ps for clarity.

The loss of helical content in B9–B19 helical region of systems O1 and O2 is correlated with the loss of backbone hydrogen bonds in that region, as shown in Figure 5.12. The plots have been averaged over 100 ps for clarity.

As observed with the static field systems, insulin chain-B under oscillating field stress of strength less than or equal to $10^8$ V/m also showed all-helical residues to remain at 80%
at the end of the simulation. However, the $\alpha$-helical content showed greater preservation in most of the systems except O3, O7, and O8. This $\alpha$-helical retainment is due to the lack of $\alpha$- to $\pi$-helix transition observed in the N300 and most of the lower strength static field systems.

The simulation performed at $10^7$ V/m (system O5) was observed to exhibit a similar number of retained helical residues to the 400 K system. Along with the static system at $10^7$ V/m, the three systems show similarities in the secondary structure evolution, with the loss of B4–B8 helix within 5 ns of the simulations.

5.3.7 Solvent accessible surface area

The SASA was calculated using the program X-PLOR for the whole data collection stage after the first 1 ns. For clarity, the resulting SASA and variance are expressed as a percentage of the SASA of the N300A system, and are presented in Table 5.6.

In contrast to the static field systems, the application of oscillating fields severely disrupts the secondary structure at lower field strength of $5 \times 10^8$ V/m and exhibits large SASA variance. This finding is consistent with the other results presented above, and highlights the different mechanism by which the static and oscillating electric fields can cause a disruption to the normal protein behaviour.
CHAPTER 5. INSULIN SIMULATIONS: OSCILLATING ELECTRIC FIELD STRESS

<table>
<thead>
<tr>
<th>System name</th>
<th>Percentage SASA (relative to N300A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O1</td>
<td>118.8 ± 6.6</td>
</tr>
<tr>
<td>O2</td>
<td>120.8 ± 6.8</td>
</tr>
<tr>
<td>O3</td>
<td>107.7 ± 5.7</td>
</tr>
<tr>
<td>O4</td>
<td>100.7 ± 4.4</td>
</tr>
<tr>
<td>O5</td>
<td>105.1 ± 6.1</td>
</tr>
<tr>
<td>O6</td>
<td>98.9 ± 4.2</td>
</tr>
<tr>
<td>O7</td>
<td>98.3 ± 4.1</td>
</tr>
<tr>
<td>O8</td>
<td>100.1 ± 2.6</td>
</tr>
<tr>
<td>O9</td>
<td>106.1 ± 4.9</td>
</tr>
<tr>
<td>N300A</td>
<td>100.0 ± 5.8</td>
</tr>
<tr>
<td>N400</td>
<td>107.7 ± 6.1</td>
</tr>
</tbody>
</table>

Table 5.6: Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation.

5.4 Chain-B in monomeric environment

5.4.1 Secondary structure

Figure 5.13 shows the secondary structure evolution of insulin chain-B under monomeric conditions as obtained using the STRIDE algorithm. The plots reveal very little disruption to the secondary structure of monomeric insulin chain-B under applied oscillating field stress for field strength of $10^8$ V/m and lower. This is in agreement with the results obtained from the simulations of monomeric chain-B under static electric field stress.

In the system with the highest field strength (system OM1), a complete loss of the helical content was observed within 0.9 ns (2.2 electric field oscillations). This timeframe is slightly faster than the time it took to lose all helical content in the equivalent isolated chain-B system. System OM2, however, was observed to retain most of its helical content even after 10 ns of simulation time. This is in contrast with the isolated chain-B under the same oscillating electric field stress, which lost its helical content within 2.2 ns. The retainment of the helical content in the monomeric insulin chain-B is indicative of the stabilising effect offered by the presence of the disulfide bonds between chain-A and chain-
CHAPTER 5. INSULIN SIMULATIONS: 
OSCILLATING ELECTRIC FIELD STRESS

Figure 5.13: Secondary structure evolution of the monomeric chain-B systems under oscillating electric field.

B of insulin. The presence of the disulfide bonds is also responsible for the lack of transition from $\alpha$- towards $\pi$-helix, which was observed in the isolated chain-B systems.
5.4.2 Dipole moment distribution

The dipole moments in the $z$-direction for the B9–B19 helix of the monomeric insulin chain-B over the data collection stage are plotted in Figure 5.14. The dipole moments have been normalised to allow direct comparison between the individual plots. As previously observed in the isolated chain-B systems under oscillating electric field stress, the plots show two general conditions that can arise: a double peak, where the helix is continuously realigning with respect to the field direction; and a skewed distribution where the helix is not fast enough to align with the electric field.

![Dipole moment distribution](image_url)

**Figure 5.14:** B9–B19 helical dipole moment distribution in the $z$-direction of insulin chain-B in the presence of chain-A. Each plot has been displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.

In Section 4.4.2, it was found that insulin chain-B under monomeric conditions exhibited stronger coupling between the helical dipole moment and the static electric field, compared to isolated chain-B. Figure 5.15 shows the comparison between the helical dipole moment in the $z$-direction of insulin chain-B in monomeric conditions and the isolated chain-B. No difference in the coupling strength is observed, as evident by the same absolute values of the dipole moment in each electric field strength for both systems.
5.4.3 Root mean square displacement

The RMSDs of the B9–B19 helix backbone atoms from each simulation under oscillating electric field were calculated relative to the start of the data collection stage. These are plotted in Figure 5.16 and have been averaged over 100 ps for clarity. Similarly to the behaviour of monomeric chain-B under static field stress, the RMSDs of the monomeric chain-B under oscillating field stress stay relatively constant once the equilibrium values are reached.

The average RMSD values for the last 1 ns of the simulations are presented in Table 5.7. Systems OM3, OM4, and OM5 were found to be very stable with very little deviation from the normal RMSDs of the reference system. This is in agreement with the results from the monomeric chain-B under static field stress. The application of moderate strength oscillating electric field caused little effect on the conformational stability of insulin chain-B when tethered to chain-A. However, the application of oscillating electric field of strength $5 \times 10^8$ V/m and above was observed to disrupt the secondary structure to a larger extent compared to the application of static fields of similar strengths. The variances of these systems are also larger than the variances of the equivalent static field.
systems, and directly relate to the constant movement of the whole protein as it aligns with respect to the applied electric field. In this case, the application of the oscillating electric field destabilises the B9–B19 helix to a larger extent compared to the application of thermal stress. This is in contrast to the isolated chain-B systems where the application of

Figure 5.16: 100 ps running average of the B9–B19 helix backbone RMSD of the monomeric chain-B systems under oscillating electric field.
of thermal stress destabilises the B9–B19 helix to a larger extent than the application of oscillating electric field. This illustrates the stability offered by the presence of chain-A and the accompanying disulfide bonds.

<table>
<thead>
<tr>
<th>System name</th>
<th>RMSD [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM1</td>
<td>3.7 ± 0.4</td>
</tr>
<tr>
<td>OM2</td>
<td>2.6 ± 0.5</td>
</tr>
<tr>
<td>OM3</td>
<td>0.62 ± 0.09</td>
</tr>
<tr>
<td>OM4</td>
<td>0.48 ± 0.08</td>
</tr>
<tr>
<td>OM5</td>
<td>0.7 ± 0.1</td>
</tr>
<tr>
<td>NM300</td>
<td>0.56 ± 0.09</td>
</tr>
<tr>
<td>NM400</td>
<td>0.7 ± 0.1</td>
</tr>
</tbody>
</table>

Table 5.7: B9–B19 helical backbone RMSD of each monomeric chain-B simulation under oscillating electric field, averaged over the last 1 ns, except for system OM1, where it was averaged between 1 ns and 2 ns.

5.4.4 Radius of gyration

The average $r_g$ of monomeric insulin chain-B systems under oscillating electric field stress, after the first 1 ns are presented in Table 5.8. Chain-A has been ignored in this calculation, so that only the $r_g$ of chain-B was considered. Confirming previous observations in this section, the $r_g$ for systems under oscillating electric field up to and including $10^8$ V/m did not show significant deviations compared to the reference systems. Systems OM1 and OM2 showed large changes in $r_g$, which signifies a large amount of disruption to the secondary structure of the protein.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM1</td>
<td>13.3 ± 2.4</td>
</tr>
<tr>
<td>OM2</td>
<td>12.7 ± 1.5</td>
</tr>
<tr>
<td>OM3</td>
<td>10.6 ± 0.3</td>
</tr>
<tr>
<td>OM4</td>
<td>11.0 ± 0.2</td>
</tr>
<tr>
<td>OM5</td>
<td>11.05 ± 0.15</td>
</tr>
<tr>
<td>NM300</td>
<td>11.2 ± 0.2</td>
</tr>
<tr>
<td>NM400</td>
<td>10.55 ± 0.35</td>
</tr>
</tbody>
</table>

Table 5.8: Radius of gyration of each monomeric chain-B system under applied oscillating electric field, averaged over data collection stage, after the first 1 ns.
The variance of systems OM1 and OM2 also showed the different mechanism of damage caused by the oscillating electric field discussed in this chapter. The application of static electric field on monomeric insulin chain-B caused stretching of the chain which is accompanied by the restriction of the mobility and flexibility of the molecule. In the application of oscillating electric field, the increase of mobility and flexibility of the protein is reflected in a lower $r_g$ and larger variance.

5.4.5 Helical content

In Table 5.9, the retained number of $\alpha$-helical residues, averaged over the whole data collection range, was presented. Similarly to the monomeric insulin chain-B under static electric field stress, only the number of $\alpha$-helical residues are presented due to the lack of $\alpha$- to $\pi$-helix transitions.

<table>
<thead>
<tr>
<th>System name</th>
<th>Retained $\alpha$-helical residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM1</td>
<td>4.2 ± 4.9</td>
</tr>
<tr>
<td>OM2</td>
<td>9.0 ± 5.1</td>
</tr>
<tr>
<td>OM3</td>
<td>14.6 ± 1.5</td>
</tr>
<tr>
<td>OM4</td>
<td>16.0 ± 0.1</td>
</tr>
<tr>
<td>OM5</td>
<td>16.0 ± 0.2</td>
</tr>
<tr>
<td>NM300</td>
<td>16.0 ± 0.2</td>
</tr>
<tr>
<td>NM400</td>
<td>13.5 ± 1.6</td>
</tr>
</tbody>
</table>

Table 5.9: Retained number of $\alpha$-helical residues for monomeric insulin chain-B under oscillating electric field, averaged over the data collection stage, except system OM1 where averaging was done up to 2 ns.

The table confirms the loss of $\alpha$-helical region at the two highest electric field strengths. The variances of these two systems were shown to be larger than the variances of the equivalent static field systems, which confirms earlier observation of the damage mechanism of the applied oscillating electric field.

Compared to the equivalent static field systems which showed progressive loss of $\alpha$-helical region as the field strength was increased, the applied oscillating field did not show a significant effect until a field strength of $10^8$ V/m was reached. This is likely due to
the brief relaxation periods between the oscillations, where the instantaneous electric field strength is comparatively weak. These periods gave the protein a chance to refold the lost helical regions.

Figure 5.17: 100 ps running average of the helical content retained as a function of time for the monomeric chain-B systems under oscillating electric field. Red lines indicate the all-helical content, and the green lines indicate the $\alpha$-helical content.
Figure 5.17 shows the percentage of retained helical content present in residues B9–B19 of chain-B, relative to the start of the data collection stage. The plots have been averaged over 100 ps for clarity. It can clearly be seen that the helical content is dominated by $\alpha$-helices, with virtually no other helix types present. The helical content is conserved to the end of the simulations for systems OM3 to OM5. In system OM1, the helical content is lost within 0.9 ns, which agrees with the visual observation of the secondary structure map in Figure 5.13a. In system OM2, the helical content was reduced to an average of 50% after 2.5 ns. The loss of helical content in these systems is strongly correlated with the loss of hydrogen bonds, as shown in Figure 5.18. The timeframes of the secondary structure loss in these two systems are similar to the timeframes of secondary structure loss in systems O1 and O2. This indicates that the application of oscillating electric field has similar effect on insulin chain-B regardless of its environment (i.e. isolated or with the presence of chain-A). This is in contrast to the application of static electric field on monomeric chain-B, which showed more damage compared to isolated chain-B. This can be explained by the nature of the electric field oscillations, which offers a slight relaxation period every half period of the oscillation. This relaxation period is sufficient to lessen the cooperative effect of the helical dipoles present in the two chains.

![Figure 5.18: 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for systems OM1 and OM2.](image-url)
### 5.4.6 Solvent accessible surface area

Table 5.10 shows the percentage change of SASA for monomeric insulin chain-B systems under oscillating electric field stress, relative to the NM300 system. The values are calculated for the whole data collection stage after the first 1 ns. Chain-A has been ignored in this calculation, so that only the SASA of chain-B was considered.

<table>
<thead>
<tr>
<th>System name</th>
<th>Percentage SASA (relative to NM300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OM1</td>
<td>109.8 ± 7.9</td>
</tr>
<tr>
<td>OM2</td>
<td>101.8 ± 6.4</td>
</tr>
<tr>
<td>OM3</td>
<td>91.6 ± 4.0</td>
</tr>
<tr>
<td>OM4</td>
<td>95.7 ± 2.0</td>
</tr>
<tr>
<td>OM5</td>
<td>94.9 ± 1.9</td>
</tr>
<tr>
<td>NM300</td>
<td>100.0 ± 1.8</td>
</tr>
<tr>
<td>NM400</td>
<td>94.1 ± 4.4</td>
</tr>
</tbody>
</table>

**Table 5.10:** Percentage of solvent accessible surface area of chain-B under monomeric conditions throughout the data collection stage after the first 1 ns, relative to the NM300 simulation.

In contrast to the results for the equivalent static field systems, the systems under oscillating field stress show a decrease of overall SASA in systems with field strengths of $10^8$ V/m and lower. The average value of SASA for each of these systems is similar to the value obtained for the thermally stressed system. This indicates that the application of increased temperature or oscillating electric field of moderate strength causes insulin chain-B in a monomeric environment to adopt a more compact conformation. This suggests similarities in the mechanism of damage incurred by both types of stresses, where each one increases the mobility and flexibility of the protein.

Systems OM1 shows a large change in SASA, which is as expected due to the large strength of the applied electric field. System OM2, however, has similar value of SASA, compared to the NM300 system. These two systems are in contrast to their equivalent systems under application of static electric field, which showed much larger increase in SASA. As discussed in the previous section, this is due to the brief relaxation period
between oscillations, which limits the cooperative effects of the dipole moments present in the whole protein.

The variance of the systems under oscillating electric field stress shows increasing trend as the field strength is increased. This is in contrast to the variance of the monomeric insulin chain-B under static electric field stress, which shows a decreasing trend instead. The increasing variance is a direct consequence of the oscillating nature of the electric field, which causes the protein to constantly align itself with respect to the applied field.

5.5 Conclusion

In this chapter, we have explored and compared the effect of 2.45 GHz oscillating electric fields of various strengths on the behaviour of both isolated and monomeric insulin chain-B.

Compared to the application of static electric fields, the application of oscillating electric fields was observed to have a more destabilising effect. The oscillating field was shown to cause complete loss of secondary structure at a lower field strength compared to static electric field, suggesting the rapid change in electric field direction does more damage to the secondary structure than the application of a continuous electric field.

The mechanism of damage caused by oscillating electric fields is believed to stem from the rapid alignment process that the protein undergoes with the changing electric field strength and direction. As the field gets stronger, the protein undergoes a faster alignment process with the field with very minimal lag time every time the field changes direction. In order to do so, the protein has to move very rapidly. The process by which a protein moves rapidly is similar to the application of thermal stress. In addition to the rapid alignment process, the protein also experiences constant stretching and relaxation of the helical region as the electric field oscillation reaches its peaks and zeros. This process accelerates the unfolding process of the protein.
Under the application of oscillating electric fields, the presence of more than one helical region on a protein was observed to interfere with the cooperative effect of the dipole moments, as was observed upon the application of static electric fields. This results in less damage compared to the equivalent system under static electric field stress, and is more evident as the field strength is increased.
Chapter 6

Insulin Simulations: Frequency Dependent Electric Field Stress

The pure and simple truth is rarely pure and never simple.

_Oscar Wilde (1854–1900)_

### 6.1 Overview

As an expansion of Chapter 5, the effects of oscillating electric field of different frequencies on protein behaviour are investigated. The implementation of the 1.225 GHz and 4.9 GHz oscillating electric fields is described in Section 6.2, along with the details of simulation conditions for each system under those oscillating fields. In Section 6.3, the effects of the 1.225 GHz and 4.9 GHz oscillating electric field of various strengths on isolated chain-B immersed in water are studied and compared to the results obtained from previous chapters. In Section 6.4, the effects of the 1.225 GHz and 4.9 GHz oscillating electric fields of various strengths on chain-B in the monomeric condition (i.e. in the presence of chain-A with all disulfide bonds present) are investigated and compared to the results obtained in earlier chapters. Finally, the conclusion is presented in Section 6.5.
6.2 Methodologies

In the previous two chapters, we have determined the difference in the nature of the damage mechanism inflicted by the application of static and oscillating electric fields on insulin chain-B in isolated and monomeric conditions. We have also demonstrated the important role helical residues play in these interactions. In this section, we are exploring the effect of different oscillating electric field frequencies on the behaviour of insulin chain-B. Both the isolated and monomeric environment are studied and compared to the results from Chapters 4 and 5.

The method used in this section is identical to the method used in Chapter 5. However, instead of using a frequency of 2.45 GHz (or a period of 408.16 ps), electric field frequencies of 1.225 GHz and 4.5 GHz were used. These frequencies correspond to periods of 816.32 ps or 204.08 ps, respectively. Other simulation parameters are as outlined previously in Section 2.5. Figure 6.1 illustrates the difference between the three oscillating field frequencies.

![Figure 6.1: Comparison of oscillating electric field frequencies of 2.45 GHz, 1.225 GHz, and 4.9 GHz.](image)

The full details of the simulations performed are shown in Table 6.1. The letter “O” denotes oscillating electric field, the letter “M” denotes monomeric conditions (i.e. chain-A is present, and the disulfide bonds between the two chains left intact), the letter “H”
denotes half of the original electric field frequency \( f = 1.225 \text{ GHz} \), and the letter “D” denotes twice the original electric field frequency \( f = 4.9 \text{ GHz} \).

\[
\begin{array}{cccc}
\text{System name} & \text{Simulation length (ns)} & \text{System name} & \text{Simulation length (ns)} \\
\hline
O2H & 10 & O2D & 10 \\
O3H & 10 & O3D & 10 \\
O4H & 10 & O4D & 10 \\
O5H & 10 & O5D & 10 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{System name} & \text{Simulation length (ns)} & \text{System name} & \text{Simulation length (ns)} \\
\hline
OM2H & 10 & OM2D & 10 \\
OM3H & 10 & OM3D & 10 \\
OM5H & 10 & OM5D & 10 \\
\end{array}
\]

\[
\begin{array}{cccc}
\text{Electric field strength, } E_{\text{RMS}} \text{ (V/m)} \\
5 \times 10^8 & 5 \times 10^8 & 5 \times 10^7 & 10^7 \\
\end{array}
\]

Table 6.1: Summary of simulation systems and conditions used in this chapter.

### 6.3 Isolated insulin chain-B

#### 6.3.1 Secondary structure

Figure 6.2 shows the secondary structure evolution of the systems described in Table 6.1. All of the systems presented showed similarities to the secondary structure evolution of the systems under applied oscillating field of 2.45 GHz, shown in Figure 5.3.

Similarly to system O2 (see Figure 5.3b), both systems O2H and O2D exhibit disruption to the secondary structure. In the system with electric field frequency of 1.225 GHz, the helical region is lost within 5 ns (6.2 electric field oscillations). At \( 10^8 \text{ V/m} \), system O3D with electric field frequency of 4.9 GHz exhibited disruption to the secondary structure. This damage was not observed in system O3 (with a frequency of 2.45 GHz) as is the case with system O3H (with a lower frequency of 1.225 GHz). Instead, in system O3H, the \( \alpha \)-helix is conserved, similar to the stabilising effect of the application of a static field (refer to Section 4.3.1). The snapshots of insulin chain-B conformation of systems O2H, O2D, O3H, and O3D at 6 ns are shown in Figure 6.3.

The difference in behaviour of insulin chain-B under oscillating electric fields of different frequencies is intrinsically related to the nature of the oscillating electric field. Figure 6.4 illustrates that within the same timeframe of approximately 100 ps, the low
frequency electric field does not change as much as the high frequency oscillating field. As a result, compared to the application of high frequency electric field, the application of low frequency field within the same timeframe more closely resembles a static field.
CHAPTER 6. INSULIN SIMULATIONS:
FREQUENCY DEPENDENT ELECTRIC FIELD STRESS

Figure 6.2: Secondary structure evolution of the isolated chain-B systems under oscillating electric field of different frequencies. (continued)

Figure 6.3: Snapshots of isolated insulin chain-B conformations at 6 ns for systems O2H, O2D, O3H, and O3D. Disruption to the secondary structure is observed in all systems except O3H. Residues B9–B19 are highlighted in orange.

In terms of the simulation timeframe of 10 ns, this means the systems with low frequency electric field spend more time experiencing continuous strong instantaneous electric field compared to systems with high frequency electric field.

For the systems with field strength lower than $10^8$ V/m, no helix breaking event was observed, although the $\alpha$- to $\pi$-helix transitions occurred in all these systems except O5D.
CHAPTER 6. INSULIN SIMULATIONS: FREQUENCY DEPENDENT ELECTRIC FIELD STRESS

Figure 6.4: Comparison of the oscillating fields of different frequencies shows that within the same time period, the low frequency oscillating field does not change as much as the high frequency oscillating field. The maximum of each electric field frequency has been aligned with respect to the maximum of the electric field of frequency 4.9 GHz.

6.3.2 Dipole moment distribution

Figure 6.5 shows the distribution of dipole moment values of the B9–B19 helix in the applied field direction (z-axis) for all the systems over the data collection stage. The dipole moment distributions have been normalised to allow direct comparison between different electric field strengths.

Figure 6.5: B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under oscillating electric field of different frequencies. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.
The distributions of the dipole moment show the same trend as for the oscillating field systems at 2.45 GHz, shown in Figure 5.6. Double peaks were observed primarily in the systems with electric field frequency of 1.225 GHz at relatively high field strengths, indicating the continuous realignment of the helix with respect to the field. The remaining systems show a skewed distribution which is probably the result of the lag time of the protein response as it aligns with respect to the applied electric field.

Systems O2H and O3H exhibit stronger correlation to the applied field, compared to systems O2D and O3D. This is probably due to the longer time spent by chain-B under the presence of the field without the field changing its direction. The effect of the 1.225 GHz oscillating fields is more similar to the application of static fields compared to the 2.45 GHz and 4.9 GHz oscillating fields, as illustrated in Figure 6.6.

![Figure 6.6: Comparison of the B9–B19 helical dipole moment distribution in the z-direction of isolated insulin chain-B under static electric field and oscillating electric field of frequencies 2.45 GHz, 1.225 GHz, and 4.9 GHz. Each plot is displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.](image)

### 6.3.3 Root mean square deviation

The structural stability of the B9–B19 helical region under the application of oscillating fields of different frequencies was investigated by determining the RMSD of the backbone
atoms compared to the start of the data collection stage. The plots of the RMSDs as a function of time for each system are shown in Figure 6.7.

The RMSDs show comparable variability compared to the RMSDs under the oscillating field system of 2.45 GHz (refer to Section 5.3.4). This variability is expected due to the oscillatory nature of the electric field. The systems studied in this chapter all showed relative stability in the RMSD values by the end of the simulation timeframe, particularly in the systems with moderately low electric field strengths. In system O2H, a jump in RMSD value at 5 ns was observed, corresponding to the loss of the helical region of the secondary structure. A somewhat lesser jump was observed in system O3D, and is also related to changes in the secondary structure. Both of these conformations are shown in Figure 6.3.

<table>
<thead>
<tr>
<th>System name</th>
<th>RMSD [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>O2_</td>
<td>4.24 ± 0.35</td>
</tr>
<tr>
<td>O3_</td>
<td>0.9 ± 0.2</td>
</tr>
<tr>
<td>O4_</td>
<td>2.21 ± 0.09</td>
</tr>
<tr>
<td>O5_</td>
<td>2.2 ± 0.1</td>
</tr>
<tr>
<td>N300</td>
<td>2.25 ± 0.08</td>
</tr>
<tr>
<td>N400</td>
<td>1.5 ± 0.6</td>
</tr>
</tbody>
</table>

Table 6.2: B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field of different frequencies, averaged over the last 1 ns.

The average RMSD values for the last 1 ns of the simulations are presented in Table 6.2. The complete loss of the helical region in system O2H is reflected in the high RMSD and agrees very well with the value obtained for system O2. This high value of RMSD was caused by the constant realignment of insulin chain-B with the applied electric field. In contrast, system O2D retained some helical region, and the RMSD was not as high as for systems O2 and O2H. At 10⁸ V/m, the application of higher frequency electric field (system O3D) caused more disruption to the secondary structure compared to systems O3H and O3. In contrast, system O3H showed helix restriction, similar to the corresponding static field system (refer to Section 4.3.1), resulting in a lower RMSD compared to the other systems. The variances of the systems studied in this section were
CHAPTER 6. INSULIN SIMULATIONS: FREQUENCY DEPENDENT ELECTRIC FIELD STRESS

Figure 6.7: 100 ps running average of the B9–B19 helix backbone RMSD of the isolated chain-B systems under oscillating electric field of different frequencies.
observed to be comparable to the variances of the equivalent systems under field frequency of 2.45 GHz, which again, highlights the oscillatory nature of the electric field applied.

6.3.4 Radius of gyration

Table 6.3 presents the average $r_g$ of each isolated chain-B system under oscillating electric field after the first 1 ns. The $r_g$ of the systems under field frequency of 1.225 GHz and 4.9 GHz are almost identical and agree quite well with the $r_g$ for systems under oscillating field frequency of 2.45 GHz, presented in Section 5.3.5. They all showed larger $r_g$ compared to system N300, although within the limits set by the variance. The systems also showed lower threshold of conformational disruption compared to the static field systems. It is interesting to note that system O3D showed comparable value of $r_g$ to the N300 system, even though the RMSD value of the B9–B19 backbone indicated relatively large disruption. This can be explained by insulin chain-B adopting a compact conformation where the termini regions are folded towards the main helix, as illustrated in Figure 6.3d.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>O2_</td>
<td>12.2 ± 1.7</td>
</tr>
<tr>
<td>O3_</td>
<td>10.5 ± 0.4</td>
</tr>
<tr>
<td>O4_</td>
<td>10.1 ± 0.4</td>
</tr>
<tr>
<td>O5_</td>
<td>10.6 ± 0.7</td>
</tr>
<tr>
<td>N300</td>
<td>9.9 ± 0.5</td>
</tr>
<tr>
<td>N400</td>
<td>11.0 ± 1.3</td>
</tr>
</tbody>
</table>

Table 6.3: Radius of gyration of each simulation, averaged over data collection stage, after the first 1 ns.

6.3.5 Helical content

The STRIDE algorithm was again employed to obtain the number of retained helical residues of each isolated insulin chain-B system under the application of oscillating electric field of different frequencies, as a measure of the stability of the helix over the data collection stage. The retained numbers of all-helical (consisting of $\alpha$-, $\pi$-, and $3_{10}$-helices)
and α-helical residues of insulin chain-B over the whole data collection stage are presented in Table 6.4.

<table>
<thead>
<tr>
<th>System name</th>
<th>Retained all-helical residues</th>
<th>Retained α-helical residues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>_H</td>
<td>_D</td>
</tr>
<tr>
<td>O2_</td>
<td>4.0 ± 4.5</td>
<td>5.9 ± 3.7</td>
</tr>
<tr>
<td>O3_</td>
<td>11.8 ± 2.0</td>
<td>6.7 ± 2.6</td>
</tr>
<tr>
<td>O4_</td>
<td>12.6 ± 1.9</td>
<td>12.7 ± 2.1</td>
</tr>
<tr>
<td>O5_</td>
<td>12.4 ± 2.0</td>
<td>12.8 ± 1.3</td>
</tr>
<tr>
<td>N300</td>
<td>10.7 ± 2.7</td>
<td></td>
</tr>
<tr>
<td>N400</td>
<td>8.8 ± 2.4</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.4: Retained number of all-helical and α-helical residues for insulin chain-B under oscillating field of different frequencies, averaged over the whole data collection range.

The retained number of all-helical residues exhibit similar values and trend as those obtained from the simulations with oscillating field frequency of 2.45 GHz, presented in Section 5.3.6. However, there is more variation in the retained number of α-helical residues in the systems under oscillating field frequencies of 1.225 GHz and 4.9 GHz, compared to the systems under field frequency of 2.45 GHz. This is a reflection of the high degree of α- to π-helix transition observed in the systems presented in this section.

At field strength of $10^8$ V/m, significant variations were observed for oscillating fields of different frequencies. At a low frequency of 1.225 GHz (system O3H), stable helix was observed which preserved the α-helical region to the end of the simulation. This helix preservation is nearly as strong as in the case of a static field, as demonstrated by the O3H system preserving an average of 11.6 α-helical residues compared to 12.7 residues preserved by the application of static electric field of the same effective strength. At a field frequency of 2.45 GHz, 8 α-helical residues were maintained. At a higher frequency of 4.9 GHz, only an average of 4.9 helical residues were retained. This illustrates that at a field strength of $10^8$ V/m, increasing frequency of the electric field leads to increasing damage to the secondary structure of isolated insulin chain-B. At the same time, the lower frequency has a similar stabilising effect which may restrict normal protein behaviour.
Figure 6.8: 100 ps running average of the B9–B19 helical content retained as a function of time for the isolated chain-B systems under oscillating electric field of different frequencies. Red lines indicate the all-helical content, and the green lines indicate the α-helical content. (continued on next page)

In order to look at the changes to the central B9–B19 helix in terms of the number of helical residues, the plots of all-helical and α-helical residues of each insulin chain-B system under oscillating electric field as a function of time are presented in Figure 6.8. The plots have been averaged over 100 ps for clarity.
Figure 6.8: 100 ps running average of the B9–B19 helical content retained as a function of time for the isolated chain-B systems under oscillating electric field of different frequencies. Red lines indicate the all-helical content, and the green lines indicate the α-helical content. (continued)

Figure 6.9: 100 ps running average of the retained B9–B19 helical backbone hydrogen bonds as a function of time for isolated chain-B systems under electric field of different frequencies with strength equal to or above $10^8$ V/m.
As observed in the static and oscillating field systems of 2.45 GHz, the all-helical residues are maintained at 80% (8.8 residues) throughout the simulation in all systems except for O2H, O2D, and O3D, where the significant disruption to the secondary structure was observed. The α-helical content is generally reduced to 30% (3.3 residues) by the end of the simulation, due to the presence of α- to π-helix transitions. However, in systems O3H and O5D, the α-helical content was retained to the end of the simulation.

Figure 6.9 shows the retained hydrogen bonds between the backbone atoms of residues B9–B19 of systems O2H, O2D, O3H, and O3D. The plots have been averaged over 100 ps for clarity. The plots demonstrated the correlation between the loss of hydrogen bonds between the backbone atoms with the loss of the helical region.

### 6.3.6 Solvent accessible surface area

The SASA was calculated using the program X-PLOR for the whole data collection stage after the first 1 ns. For clarity, the resulting SASA and variance are expressed as a percentage of the SASA of the N300A system, and are presented in Table 6.5.

<table>
<thead>
<tr>
<th>System name</th>
<th>Percentage SASA (relative to N300A)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>O2_-</td>
<td>116.9 ± 8.8</td>
</tr>
<tr>
<td>O3_-</td>
<td>104.9 ± 3.1</td>
</tr>
<tr>
<td>O4_-</td>
<td>101.9 ± 6.0</td>
</tr>
<tr>
<td>O5_-</td>
<td>107.2 ± 3.6</td>
</tr>
<tr>
<td>N300A</td>
<td>100.0 ± 5.8</td>
</tr>
<tr>
<td>N400</td>
<td>107.7 ± 6.1</td>
</tr>
</tbody>
</table>

**Table 6.5:** Percentage of solvent accessible surface area throughout the data collection stage after the first 1 ns, relative to the N300A simulation.

The results presented in Table 6.5 agree very well with the results obtained on the systems with applied oscillating field of frequency 2.45 GHz, presented in Section 5.3.7. The only exception is the system at $10^8$ V/m. As previously discussed, the frequency of oscillating electric field applied at this strength correlates with the extent of damage sustained by insulin chain-B. At the lowest frequency applied (1.225 GHz, system O3H),
the SASA was minimally affected, with only 4.9% increase relative to the SASA of the N300A system. At field frequency of 2.45 GHz (system O3), the SASA increased by 7.7%. At the highest frequency of 4.9 GHz (system O3D), the increase in SASA is 8.1% relative to the SASA of the N300A system. It is also notable that at this field strength, as the frequency was increased, the variance in SASA also increased.

6.4 Chain-B in monomeric environment

6.4.1 Secondary structure

Figure 6.10 shows the secondary structure evolution of insulin chain-B under monomeric conditions as obtained using the STRIDE algorithm. As previously observed with the oscillating field of frequency 2.45 GHz, the plots reveal little disruption to the secondary structure of monomeric insulin chain-B under applied oscillating field stress for field strengths of $10^8$ V/m and lower. This finding is also in agreement with the results obtained from the simulations of monomeric chain-B under static electric field stress. There were no α- to π-helix transitions observed in the systems presented in this section.

At field strength of $5 \times 10^8$ V/m, the application of oscillating field of frequency 1.225 GHz (system OM2H) was shown to disrupt the secondary structure of monomeric insulin chain-B, whereas a frequency of 4.9 GHz (system OM2D) did not show any obvious signs of damage. This relates to the presence of chain-A (see discussion in Section 4.4.2), which contributes its two helical dipoles to the total dipole moment of the protein. In the case of system OM2H, the applied electric field oscillations were such that insulin chain-B experienced more time under the influence of the field before it changed direction. The overall effect is that it experiences stretching of the helical region, similar to what is experienced by system SM2 under the application of static field. In contrast, in system OM2D, the oscillating field changes direction rapidly enough to not cause significant damage to the secondary structure. This is because the whole protein is only exposed for a brief amount of time to a strong instantaneous electric field, before the field changes direction. Unlike the isolated chain-B in the field of the same strength (system O2D) the
Figure 6.10: Secondary structure evolution of the monomeric chain-B systems under oscillating electric field of different frequencies.

presence of disulfide bonds confers increased resistance to damage in system OM2D, even at such large electric field strength.
6.4.2 Dipole moment distribution

The dipole moments in the $z$-direction for the B9–B19 helix of the monomeric insulin chain-B over the data collection stage are plotted in Figure 6.11. The dipole moments have been normalised to allow direct comparison between the individual plots.

Figure 6.11: B9–B19 helical dipole moment distribution in the $z$-direction of insulin chain-B in the presence of chain-A, under the application of oscillating fields of different frequencies. Each plot has been displaced by 0.1 to improve clarity. Positive dipole moment signifies positive correlation with the electric field.

The plots again show either a double peak, where the helix is continuously realigning with respect to the field direction (systems OM2H, OM2D); or a skewed distribution where the molecule is not fast enough to align with the electric field (the remaining systems). The skewed distribution is more prevalent in the lower electric field strength, as is observed in Figure 6.11, because the electric field is not strong enough to cause near-instantaneous response of the protein as the field changes direction.

6.4.3 Root mean square displacements

The RMSDs of the B9–B19 helix backbone atoms from each simulation were calculated relative to the start of the data collection stage. These are plotted in Figure 6.12 and have been averaged over 100 ps for clarity. The behaviour of the systems under oscillating field frequency of 1.225 GHz was observed to be practically identical to the behaviour of
the equivalent systems under oscillating field frequency of 2.45 GHz, presented in Section 5.4.3. In contrast, at field frequency of 4.9 GHz, system OM2D was observed to have relatively stable RMSD of 0.66 Å. Nonetheless, variations in the instantaneous RMSD value were observed, which indicate that the relative stability occurs because the whole protein is only briefly exposed to the strong instantaneous electric field before the field changes its direction. In between the direction changes, the whole protein only experiences minimal instantaneous electric field stress, and hence has an opportunity to repair the conformational changes that might have occurred.

The average RMSD values for the last 1 ns of the simulations are presented in Table 6.6. Compared to the RMSDs of the reference system, all the systems at or under $10^8$ V/m showed good agreement. The RMSDs of those systems also agree with the RMSDs of the monomeric insulin chain-B systems under static fields or oscillating fields of frequency 2.45 GHz. As discussed previously in section 5.4.3, this finding illustrates the stability offered by the presence of chain-A and the accompanying disulfide bonds tethering both chains together.

<table>
<thead>
<tr>
<th>System name</th>
<th>RMSD [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>H</em></td>
</tr>
<tr>
<td>OM2_</td>
<td>2.7 ± 0.2</td>
</tr>
<tr>
<td>OM3_</td>
<td>0.59 ± 0.07</td>
</tr>
<tr>
<td>OM5_</td>
<td>0.53 ± 0.06</td>
</tr>
<tr>
<td>NM300</td>
<td>0.56 ± 0.09</td>
</tr>
<tr>
<td>NM400</td>
<td>0.7 ± 0.1</td>
</tr>
</tbody>
</table>

Table 6.6: B9–B19 helical backbone RMSD of each monomeric chain-B simulation under oscillating electric field of different frequencies, averaged over the last 1 ns.

At a field strength of $5 \times 10^8$ V/m, the lower electric field frequency of 1.225 GHz (system OM2H) yielded a slightly higher RMSD compared to the RMSD of the system under the electric field frequency of 2.45 GHz (system OM2). However, the RMSD and variance of system OM2H were observed to be of the order exhibited by the monomeric insulin chain-B system under the application of static electric field of the same effective strength. The similarity of the effect of the 1.225 GHz oscillating field on monomeric insulin chain-B and the effect of the static field is related to the longer exposure of the
whole protein to the applied field as the oscillating field nears its maximum. This is in contrast to the oscillating field frequency of 4.9 GHz, where the whole protein only experiences brief periods of strong instantaneous electric field as the oscillating field nears its maximum. In the case of system OM2D, these brief periods of strong instantaneous
electric field are not enough to cause any structural damage to the B9–B19 backbone.

### 6.4.4 Radius of gyration

Table 6.7 presents the average $r_g$ of each simulation discussed in this section, after the first 1 ns. Chain-A has been ignored in this calculation, so that only the $r_g$ of chain-B was considered. The results generally agree with the results from the application of oscillating field of frequency 2.45 GHz, presented in Section 5.4.4. The large $r_g$ of system OM2H is expected, as the application of oscillating field strength of $5 \times 10^8$ V/m with frequency of 2.45 GHz has been shown to disrupt the secondary structure conformation of monomeric insulin chain-B. However, system OM2D showed similar $r_g$ to the NM300 system, while having a relatively large variance, as discussed previously in Sections 6.4.1 and 6.4.3.

<table>
<thead>
<tr>
<th>System name</th>
<th>$r_g$ [Å]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>_H</td>
</tr>
<tr>
<td>OM2_-</td>
<td>12.1 ± 1.2</td>
</tr>
<tr>
<td>OM3_-</td>
<td>10.9 ± 0.2</td>
</tr>
<tr>
<td>OM5_-</td>
<td>10.53 ± 0.25</td>
</tr>
<tr>
<td>NM300</td>
<td>11.2 ± 0.2</td>
</tr>
<tr>
<td>NM400</td>
<td>10.55 ± 0.35</td>
</tr>
</tbody>
</table>

**Table 6.7:** Radius of gyration of monomeric chain-B systems under applied oscillating electric field of different frequencies, averaged over data collection stage, after the first 1 ns.

### 6.4.5 Helical content

In Table 6.8, the retained number of $\alpha$-helical residues, averaged over the whole data collection range, is presented. Similarly to the monomeric insulin chain-B under static field and oscillating field stresses, only the number of $\alpha$-helical residues are presented due to the absence of $\alpha$- to $\pi$-helix transitions.

The results indicate that the application of oscillating electric field of frequencies 1.225 GHz and 4.9 GHz generally yields the same effect to a protein’s $\alpha$-helical content as the oscillating field of frequency 2.45 GHz, presented in Section 5.4.5. The loss of
CHAPTER 6. INSULIN SIMULATIONS:  
FREQUENCY DEPENDENT ELECTRIC FIELD STRESS

<table>
<thead>
<tr>
<th>System name</th>
<th>Retained α-helical residues</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>_H</td>
</tr>
<tr>
<td>OM2_</td>
<td>8.3 ± 4.3</td>
</tr>
<tr>
<td>OM3_</td>
<td>15.8 ± 0.7</td>
</tr>
<tr>
<td>OM5_</td>
<td>13.8 ± 1.4</td>
</tr>
<tr>
<td>NM300</td>
<td>16.0 ± 0.2</td>
</tr>
<tr>
<td>NM400</td>
<td>13.5 ± 1.6</td>
</tr>
</tbody>
</table>

**Table 6.8:** Retained number of α-helical residues for monomeric insulin chain-B under oscillating electric field of different frequencies, averaged over the data collection stage.

α-helical content is observed in system OM2H, in agreement with the observation of system OM2. However, system OM2D retained most of its helical content, in contrast to the application of oscillating electric field with frequencies 1.225 GHz and 2.45 GHz. This provides further confirmation of the observations presented previously, where the application of high frequency electric field caused less damage than the application of oscillating electric field of lower frequencies. This is because high frequency oscillating electric field is only sufficient to align the proteins, but is not enough to cause structural damage which normally arises due to the stretching of the helical regions.

Figure 6.13 shows the percentage of retained helical content in residues B9–B19 of chain-B, relative to the start of the data collection stage. The plots have been averaged over 100 ps for clarity. In this figure, both all-helical (consisting of α-, π-, and 3_{10}-helices) and α-helical content of the B9–B19 region of monomeric insulin chain-B are presented. The figure highlights the stability of the systems under the applied field of strength lower than $5 \times 10^8$ V/m. System OM2H was shown to exhibit similar progressive loss of the B9–B19 helical region to system OM2 (compare Figure 6.13a and 5.17b) as the simulation progresses, and correlates well with the plot of retained backbone hydrogen bonds (Figure 6.14a). After 2.5 ns, the helical content was reduced to an average of 50%. In contrast, system OM2D retained all of its B9–B19 helical content to the end of the simulation. However, some variations are observed at around 4 ns and 8 ns. This variation indicates that some disruption to the secondary structure does occur throughout
CHAPTER 6. INSULIN SIMULATIONS:
FREQUENCY DEPENDENT ELECTRIC FIELD STRESS

Figure 6.13: 100 ps running average of the helical content retained as a function of time for the monomeric chain-B systems under oscillating electric field of different frequencies. Red lines indicate the all-helical content, and the green lines indicate the \( \alpha \)-helical content.

The simulation, but was quickly repaired by the protein. These variations arise from the fluctuations in the retained backbone hydrogen bonds, as shown in Figure 6.14b.
6.4.6 Solvent accessible surface area

Table 6.9 shows the percentage change of SASA for monomeric insulin chain-B systems under oscillating electric field stress of different frequencies, relative to the NM300 system. The values are calculated for the whole data collection stage after the first 1 ns. Chain-A has been ignored in this calculation, so that only the SASA of chain-B was considered.

<table>
<thead>
<tr>
<th>System name</th>
<th>Percentage SASA (relative to NM300)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>OM2_H</td>
<td>102.1 ± 5.2</td>
</tr>
<tr>
<td>OM3_H</td>
<td>93.7 ± 2.9</td>
</tr>
<tr>
<td>OM5_H</td>
<td>89.0 ± 3.9</td>
</tr>
<tr>
<td>NM300</td>
<td>100.0 ± 1.8</td>
</tr>
<tr>
<td>NM400</td>
<td>94.1 ± 4.4</td>
</tr>
</tbody>
</table>

Table 6.9: Percentage of solvent accessible surface area of chain-B under monomeric conditions throughout the data collection stage after the first 1 ns, relative to the NM300 simulation.

The SASA of the systems considered in this section agree very well with the results from the systems under oscillating field frequency of 2.45 GHz, presented in Section 5.4.6. Below the field strength of $5 \times 10^8$ V/m, the oscillating field systems with all three frequencies showed closer SASA values to those obtained from the thermally stressed system, which is indicative of the compact state adopted by systems under stress.
System OM2H shows a large SASA and variance, which is within the value obtained by the application of oscillating field frequency of 2.45 GHz. However, system OM2D showed smaller SASA with large variance, indicating that the protein has a similar conformation to the NM400 system, albeit with higher mobility and flexibility under the oscillating field conditions applied here.

6.5 Conclusion

In this chapter, we have explored and compared the effects of oscillating electric fields of frequencies 1.225 GHz and 4.9 GHz of various strengths on the behaviour of both isolated and monomeric insulin chain-B.

The results at low field strengths agree well with the results obtained from the application of oscillating electric fields of frequency 2.45 GHz, confirming that the oscillatory nature of the electric field has more destabilising effect on insulin chain-B compared to static fields of the same effective strength.

At high fields strengths, the dipole moment of the protein was observed to interact more closely with the applied field of frequency 1.225 GHz compared to the field of frequency 4.9 GHz. This is believed to be because at a lower frequency, the protein is exposed to longer periods of high strength instantaneous electric field as the oscillation reaches its maximum. Consequently, some of the systems showed a similar restriction to the protein structure to that observed in the static field simulations. In contrast, at high frequency, the protein is exposed to shorter periods of high strength instantaneous electric field.

Similarly to the application of oscillating electric field of frequency 2.45 GHz, the application of field frequencies of 1.22 GHz and 4.9 GHz on proteins with more than one helical region was observed to interfere with the cooperative effect of the dipole moments. This effect is most evident at the high field strength of $5 \times 10^8$ V/m of frequency 4.9 GHz, where structural damage was not observed. This is in contrast to the field frequency of 1.225 GHz and 2.45 GHz, where loss of secondary structure was observed after 3 ns. The reason why the application of field frequency of 4.9 GHz does not result in any signs of
secondary structure damage is that the oscillation occurs very quickly so that the protein only experiences brief periods of high instantaneous electric field strength. Between the periods with high instantaneous electric field strength, the protein experiences relaxation periods, which allow it to repair the damaged secondary structure.

It is important to note that even with no signs of structural damage, the application of high field strength oscillating field of any frequency gave rise to the characteristic dipole moment distribution with either a double peak or a skewed distribution. This means that the alignment still took place, and indicates that the alignment process is separate from the disruption process. In order for a protein to be damaged by the oscillating electric field, it has to align first with respect to the field, before the electric field can stretch its helical residues and cause structural damage. This finding may be exploited as a simulation tool for alignment of simple helical proteins by applying strong, intermittent electric field. The application of such applied fields will enable the proteins to align themselves with respect to the electric field, and at the same time allow the proteins to repair the structural damage. This idea will be further explored in Section 7.2.
Chapter 7

Conclusions and Future Work

“All good things must come to an end”

*English proverb*

7.1 Conclusions

In this thesis, we have explored the behaviour of the insulin chain-B peptide under various conditions: ambient and under stress. The stresses studied include chemical stress, which was simulated by removing chain-A along with its connecting disulfide bonds from chain-B; thermal stress, which was simulated by elevating the temperature from 300 K to 400 K; and electric field stress, which was simulated by subjecting the system to static and oscillating electric fields of various strengths and frequencies. A range of structural and conformational analysis methods were employed in order to gain insight into the response of insulin chain-B to these forms of stresses.

Multiple simulations of isolated chain-B under ambient conditions were performed and found to reproduce the features observed in the NMR studies, including generally satisfying NMR experimental NOE distance restraints better than the crystal structure. Our results also revealed the existence of multiple pathways in the “normal” dynamics of isolated chain-B, including one that led to the formation of a T-like state. The T-state is
believed to be the active state of insulin, and has the B1–B8 residues forming an extended conformation which is directed away from the core of the insulin molecule. The T-like state encountered in this study has the B1–B8 residues pointing away from the main B9–B19 helix of chain-B, albeit still retaining the helical structure. In both of these pathways, the glycine residue B8 was found to play a key role in initiating the structural change.

The application of thermal stress was shown to increase the conformational dynamics of both isolated and monomeric insulin chain-B by increasing the mobility and flexibility of the termini regions. This was evident from the conformational analysis which showed a conserved B9–B19 helical core, while at the same time secondary structure disruption occurred at both termini, which changed the overall size of the whole protein.

Compared to the isolated form, insulin chain-B under monomeric conditions was observed to be very stable, owing to the presence of the disulfide bonds which tether it to chain-A. The presence of these disulfide bonds is important in order to preserve the secondary structure of the insulin monomer and also inhibits the helix-breaking process of insulin chain-B within the simulation timeframe. These bonds were strong enough to withstand elevated temperature and electric field of strengths up to $10^8 \text{ V/m}$ with minimal disruption to the secondary structure.

The application of static electric field was shown to affect the behaviour of the protein as it aligns with respect to the electric field direction. This alignment arises because of the coupling between the electric field and the protein’s helical dipole moment. For simple proteins, with few long helical regions, this alignment may also induce the stabilisation of the $\alpha$-helix, thus hindering flexibility and can possibly prevent important conformational changes such as the transition to an active state or binding of other molecules. This transition was inhibited by the application of electric field over $5 \times 10^7 \text{ V/m}$.

The application of oscillating fields, on the other hand, was shown to have a more destabilising effect on a protein’s secondary structure. Complete loss of secondary structure upon the application of oscillating fields occurs at a lower field strength compared to static electric fields, suggesting the rapid change in electric field direction does more damage to the secondary structure than the application of a continuous electric field. This
finding suggests a different mechanism of damage caused by oscillating electric fields. Whereas in static fields the damage to a protein arises from the restriction of flexibility and interference with normal dynamic behaviour, in oscillating fields, the damage appears to stem from the rapid alignment process that the protein undergoes with the changing electric field strength and direction. As the field gets stronger, the protein undergoes faster alignment with the field with very minimal lag time when the field changes direction. In order to do so, the protein has to move very rapidly—a process similar to that of the thermal stress response. In addition to the rapid alignment process, the protein also experiences constant stretching and relaxation of the helical region as the electric field oscillation reaches its peaks and zeros, which accelerates the unfolding process of the protein.

The effect of the oscillating fields on a protein was shown to be invariant of frequency at field strengths lower than $10^8$ V/m. However, at field strength of $10^8$ V/m or higher, there were some frequency-dependent effects observed. The dipole moment of the protein was observed to interact more strongly with the applied field of frequency 1.225 GHz compared to the field of frequency 4.9 GHz. This is likely due to the protein being exposed to longer periods of instantaneous high strength electric field as the oscillation reaches its maximum at a lower oscillating field frequency. Conversely, the protein is exposed to shorter periods of instantaneous high strength electric field at high frequency oscillating field. Consequently, some of the systems showed a similar behaviour to that observed in the static field simulations.

When helical fragments are in the vicinity of other helical fragments, their dipole moments can affect one another, resulting in variations in the field strength necessary to align the protein with respect to the field. For monomeric insulin under the application of static electric field stress, all three helical dipole moments—arising from the helical region in chain-B and the two helical regions in chain-A—interact with the applied field. This results in a lower field strength necessary to cause disruption to the secondary structure. In the monomeric chain-B, this field strength was $5 \times 10^8$ V/m, whereas in the isolated chain-B, it was $10^9$ V/m.
Under the application of oscillating electric fields, the presence of more than one helical region on a protein was also observed to interfere with the cooperative effect of the dipole moments. This resulted in relatively less damage to the protein structure compared to the equivalent system under the static electric field stress, the effect being more pronounced as the field strength is increased. At a field strength of $5 \times 10^8$ V/m, a frequency dependence was observed, with complete loss of secondary structure at field frequencies of 1.225 GHz and 2.45 GHz, but no observable structural damage at field frequency of 4.9 GHz. This is due to the rapid oscillation of the electric field, so that the protein only experiences brief periods of instantaneous high strength electric field. Between the periods with high instantaneous electric field strength, the protein experiences relaxation periods, which allow it to repair the damaged secondary structure.

Overall, the application of electric field was observed to cause alignment of the helical regions, which is characterised by a directional preference observed in the dipole moment distribution. The alignment is stronger with the application of stronger fields. While prolonged exposure to high field strengths damages the protein by stretching the helical residues, the application of short-term static field or oscillating field was observed to cause less structural damage. This indicates that the alignment process is separate from the conformational disruption process. In order for a protein to be damaged by the oscillating electric field, it first has to align with respect to the field, before the electric field can stretch its helical residues and cause structural damage. This finding can be exploited as outlined in Section 7.2.

In proteins with little or no helical regions, the damage in vivo is most likely due to the movement of the centre of mass induced by the application of the electric field. This movement of centre of mass is also experienced by the heat shock proteins. When this happens, the heat shock proteins may have difficulties in binding to the unfolded protein, resulting in the accumulation of partially unfolded proteins.
7.2 Future work

A protein may form amorphous aggregates *via* partially folded intermediates as part of its normal function. Alternatively, the deposition of insoluble aggregates of misfolded proteins, amyloid fibrils, results in a range of human diseases, such as Alzheimer’s and variant Creutzfeldt-Jakob disease [309]. Little is known of the mechanism of fibril formation. It is believed to be a multistage process driven by hydrophobic interactions where a variety of intermediate structures are formed [310]. The precursors to fibrillation are the partially folded intermediates, which interestingly, appear to be more toxic than the amyloid fibrils. Their subsequent development into mature fibrils may be a form of protection. Fibrillation may be enhanced by the local environment, such as changes in the metal ion concentration, ionic strength, presence of chaotrope agents, temperature, pH conditions, organic solvents, or cosolvents. Other factors include mutations, transmitted prion proteins, or simply the inevitable aging process.

Insulin belongs to the family of unrelated proteins that is very prone to form amyloid fibrils under slightly destabilising conditions [311, 312]. Fibrillation in insulin is a result of inherent properties such as self-association, crosslinking and denaturation. These properties have created difficulties in production, storage and therapeutic use of insulin [274]. Fibril formation of insulin is believed to follow a series of steps [313]—monomerisation, formation of partially folded intermediate or intermediates, nucleation, and fibril growth. A number of models have been put forward to explain the mechanism of fibrillation [264, 267, 312–315]. All the models involve significant conformational changes in the fibrillation process; however, the exact structural mechanism continues to remain ambiguous.

The molecular basis of the formation of fibrils in insulin, which is believed to be a multi-step process, is poorly understood. Insulin fibrils are suggested to arise from aggregation of the monomer state [315], so conditions promoting monomerisation also promote fibrillation. The monomer undergoes a conformational change and nucleation occurs after the assembly of several protein monomers. The rate of fibril formation can be dependent
on the lag time required to form a stable nucleus. The conditions that are known to promote fibrillation are pH 2 and temperature of 60°C. However, fibrillation in amyloidogenic peptides such as insulin, is enhanced by a range of factors. By running simulations under a variety of conditions, we can explore the factors that destabilise the normal self-association pathway and cause the perturbations in the intermediate structures. The effect of pH levels on amyloid formation is complicated [315]. Low pH enhances fibrillation, although this is at least partly attributed to acidic conditions promoting monomeric insulin to form from the hexameric form, however, a conformational change is also believed to occur. The rate of fibrillation may also be dependent on the type of acid [264]. We aim to clarify the complicated effects of ionic strength on the level of fibrillation [311, 313].

Various models have been put forward to explain the structural basis of the conformational transition. These models vary in detail, but are based on movements of the termini regions. A recent study of insulin under amyloidogenic conditions suggested an intermediate conformation where the N-termini of chains-A and B detach from the core [267]. Movements of the termini regions expose the helical portions of the protein. Previous work in amyloidogenic peptides suggests a transition between α-helix to β-strand as a step in amyloid formation [316]. This trend has also been observed in insulin where Fourier Transform Infrared (FTIR) spectroscopy studies show that soluble insulin and initial aggregates are mainly helical, but when fibrils are clearly formed, the structure is β-sheet [317]. It is highly probable, that there is in fact an ensemble of partially folded intermediates.

Mutations may also accelerate the formation of fibrils—experimental work on protein mutants has shown that structure is highly dependent on mutations [318]. Different β-sheet structure in the fibrils was associated with varying primary structure. It has also been shown that species is a relevant factor—bovine insulin is more prone to form fibrils than other species [315]. It has been suggested that amyloid formation may actually be a generic property of polypeptide chains [319]. The simulation of mutant structures may provide information on whether fibrillation is sequence specific, and the types of features that predispose a particular sequence to form fibrils. For example, there is a question
whether hydrophobic or charged residues destabilise the protein or not.

This thesis has presented the insulin molecule under various environments and stresses. The results at ambient conditions showed that we were able to reproduce experimentally observed structures and inherent conformational flexibility. Based on the findings made in this project, a three-year Australian Research Council (ARC) project has been granted to investigate the molecular mechanism of the protein aggregation process, with particular emphasis on the conditions for formation of fibrillar aggregates. We will attempt to determine the precise nature of the conformational changes which occur during the process. The partially folded intermediates are of particular interest, not only because they are believed to be the precursors to the formation of fibrils, but also because of the evidence suggesting a link between the intermediates and toxicity. We will further develop, test, and apply theoretical modelling techniques to simulate the aggregated states of the target protein insulin under a variety of conditions known to affect the rate of fibrillar aggregation. We will then expand and generalise the understanding developed to other fibrillation-prone proteins. Directly following the outcomes of this PhD project, an intermittent strong electric field is going to be applied to a system containing more than one insulin molecules in order to accelerate the alignment process which precedes the formation of $\beta$-sheets. As this thesis has suggested, the application of intermittent electric fields can be used as a simulation tool to align a protein via interaction with its helical region. Electric field conditions can be designed that will only align the proteins without causing permanent damage, therefore enabling simulation studies of the protein aggregation process.

In order to systematically describe and understand the mechanism, our study will explore the range of possible conformational states, including their relative populations. We aim to investigate the movement of the termini regions and the role of the helical regions in $\beta$-strand formation. We will examine the initial stages of aggregation, and the hydrophobic forces that are believed to drive aggregation. This may also provide insights into the structure of the nucleus and clarify such issues as whether the $\beta$-strands making up the fibrils are parallel or antiparallel in structure.
In summary, the procedures and knowledge developed in this project have led to a continuing study of insulin behaviour under various conditions which will shed light on such an important and topical issue as protein fibrillation mechanisms.
Bibliography


[114] D. Parker, M. Rivera, T. Zor, A. Henrion-Caude, I. Radhakrishnan, A. Kumar, L. H. Shapiro, P. E. Wright, M. Montminy, and P. K. Brindle. Role of secondary


BIBLIOGRAPHY


[213] A. Rauk, D. A. Armstrong, and D. P. Fairlie. Is oxidative damage by β-amyloid and prion peptides mediated by hydrogen atom transfer from glycine α-carbon to


[266] A. Budi, S. Legge, H. Treutlein, and I. Yarovsky. Effect of external stresses on


