MOLECULAR DYNAMIC SIMULATION STUDIES OF Q212H, V203G AND N173K MUTATIONS IN PRION DISEASES

SARA AMIDIAN

A dissertation submitted in partial fulfillment of the requirements for the award of the degree of Master of Science (Biotechnology)

Faculty of Biosciences and Medical Engineering
Universiti Teknologi Malaysia

APRIL 2014
To my beloved Mohammad, father and mother, sister and brother
ACKNOWLEDGEMENT

I would like to express the deepest appreciation to my advisor Assoc. Prof. Dr. Shahir Shamsir who continually and convincingly conveyed a spirit of adventure in regard to research and scholarship. Without his guidance and persistent help this dissertation would not have been possible.

My heartfelt appreciation goes to my beloved husband and best friend, for being greatly tolerant and supportive perpetually and for the sleepless nights we were working together before deadlines.

I thank my fellow Bioinformatics lab mates for the stimulating discussions and knowledge sharing sessions, and for all the fun we have had in the last two years.

I would also like to thank the Malaysian Ministry of Education for their financial support granted through Malaysian International Scholarship.

Last but not the least, I would like to thank my parents for their constant support and encouragement.
ABSTRACT

The Prion protein (PrP) roots from a collection of diseases identified as transmissible spongiform encephalopathies (TSEs), which are caused by conversion of PrP from its normal cellular form (PrP<sub>C</sub>) to a misfolded, oligomeric isoform (PrP<sub>Sc</sub>). Prion diseases are considered fatal and until now no known treatment has been reported. Prion diseases in humans are grouped based on whether they are sporadic, inherited, or acquired. In the case of inherited prion disease, which is also known as familial prion disease, abnormal prion proteins are produced through a genetic mutation. So far, about 40 point mutations have been discovered. The globular domain (aa125 to 228) of PrP plays a critical role in its folding and stability and many of pathogenic mutations are located on this part. V203G, Q212H and N173K are three PRNP mutations, which were reported recently, but it remained questionable whether the mutations were causal of the disease. In this research, we preformed Molecular Dynamic Simulation and structural analysis on the three previously mentioned unknown-disease-related mutations (V203G, Q212H and N173K) and prion disease related mutations V203I and Q212P as positive control for V203G and Q212H respectively and neutral polymorphism N171S as a negative control for N173K. We investigated to see how similar the unknown-disease-related mutations act compared to their controls, to verify whether the mutations were causal of the disease. 50 ns of molecular dynamic simulations were performed for all mutations and wild type, using GROMACS 4.6.3 software and GROMS96 force field. Changes in RMSD, RMSF, salt bridges, secondary structure and Solvent accessible surface area were explored by analyzing the trajectories. The results revealed similar dynamic behavior between Q212H, V203G, N173K and other prion pathogenic mutations; all three under study mutations showed a decrease in the protein’s overall stability, an increase in HB and HC region flexibility, a major loss in salt bridges in the HA and HB region, changes in the electrostatic surface of PrP and made the protein more exposed to solvent, which are all common dynamic behaviors among pathogenic mutations.
**ABSTRAK**

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DECLARATION</td>
<td>ii</td>
</tr>
<tr>
<td></td>
<td>DEDICATION</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENT</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td></td>
<td>ABSTRAK</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>TABLE OF CONTENTS</td>
<td>vii</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xi</td>
</tr>
<tr>
<td></td>
<td>LIST OF ABBREVIATION</td>
<td>xiii</td>
</tr>
</tbody>
</table>

1 INTRODUCTION
1.1 Background of the problem 1
1.2 Objectives of the Study 4
1.3 Scope of the Study 4
1.4 Problem Statement 5
1.5 Significance of the Study 5

2 LITERATURE REVIEW
2.1 Prion 6
2.2 Structure of Prion Protein 6
2.3 Structure of PrPSc 8
2.4 Prion Diseases 9
2.5 Models of Prion Propagation 13
2.6 Study of the effect of pathogenic mutations through Molecular Dynamic Simulation 14
3 METHODOLOGY 16
3.1 Selection of PDB file 16
3.2 Applying Mutation 16
3.3 Running Molecular Dynamic Simulation 17
  3.3.1 Preparation Stage 17
  3.3.2 Production Stage 21
  3.3.3 Analyzing Stage 21

4 RESULTS AND DISCUSSION 23
4.1 Selection of Mutations 23
  4.1.1 Selecting Q212P as a Positive Control for Q212H 24
  4.1.2 Selecting V203I as a Positive Control for V203G 24
  4.1.3 Selecting N171S as a Negative Control for N173K 25
4.2 Dictionary of Secondary Structure of Protein (DSSP) 25
4.3 Root-Mean-Square Deviation (RMSD) 27
  4.3.1 RMSD Comparisons of WT, Q212P and Q212H 27
  4.3.2 RMSD Comparison of WT, V203I and V203G 28
  4.3.3 RMSD Comparison of WT, N171S and N173K 29
4.4 Cα Root-Mean-Square Fluctuation (RMSF) 30
  4.4.1 RMSF Comparison of WT, Q212P and Q212H 30
  4.4.2 RMSF Comparison of WT, V203I and V203G 32
  4.4.3 RMSF Comparison of WT, N171S and N173K 34
4.5 Salt Bridge 37
4.6 Solvent Accessible Surface Area (SASA) 38
4.7 Electrostatic Potential Surface 39
5 CONCLUSION

5.1 Concluding Remarks 44
5.2 Future Work 45

REFERENCES 46
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Classification of human prion diseases (Gambetti et al. 2003)</td>
<td>9</td>
</tr>
<tr>
<td>4.1</td>
<td>Average values of hydrophobic, hydrophilic and total solvent accessible surface area of all mutants and WT</td>
<td>39</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE NO.</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Schematic of the prion protein. The flexible N-terminus contains the octarepeat domain. The structured C-terminal domain contains two β-strands (arrows) and three α-helices (boxes). The gray circles stand for the N-linked glycosylations. The C-terminus is with a GPI anchor, which connects the protein to the cell surface (Spevacek 2012)</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Full-length huPrP cartoon representation (Giachin et al. 2013)</td>
<td>8</td>
</tr>
<tr>
<td>2.2</td>
<td>The HuPrP’s secondary structure representation along with every polymorphisms and pathogenic mutations presently recognized (GSS in red, fCJD in black, polymorphisms in green, FFI in blue and unclassified phenotype in gray) (Giachin et al. 2013)</td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>Models of prion replication: a) The template assistance model b) The seeded polymerization model (Surewicz &amp; Apostol 2011)</td>
<td>14</td>
</tr>
<tr>
<td>3.1</td>
<td>Stages of MD simulation</td>
<td>17</td>
</tr>
<tr>
<td>4.1</td>
<td>Graphical representation of secondary structure analysis for the WT and mutant PrP, based on the DSSP algorithm.</td>
<td>26</td>
</tr>
<tr>
<td>4.2</td>
<td>RMSD comparison of WT, Q212P and Q212H</td>
<td>27</td>
</tr>
<tr>
<td>4.3</td>
<td>RMSD comparison of WT, V203I and V203G</td>
<td>28</td>
</tr>
<tr>
<td>4.4</td>
<td>RMSD comparison of WT, N171S and N173K</td>
<td>29</td>
</tr>
<tr>
<td>4.5</td>
<td>RMSF Comparisons of WT, Q212P and Q212H</td>
<td>31</td>
</tr>
<tr>
<td>4.6</td>
<td>Left: Cartoon representation of Q212P mutant showing Proline position (P212) as a sphere. Right: Cartoon representation of Q212H mutant showing Histidine position (H212) as a sphere</td>
<td>32</td>
</tr>
</tbody>
</table>
4.7 Left: Cartoon representation of V203I mutant showing Isoleusine’s position (I203) as a sphere. Right: Cartoon representation of V203G mutant showing Glycine’s position (G203) as a sphere
4.8 RMSF comparison of WT, V203I and V203G
4.9 Cartoon representation of N171S polymorphism, showing Serine’s position (S171) as a sphere
4.10 RMSF comparison of Wt and N171S
4.11 Cartoon representation of N173K mutation, showing Lysine’s position (K173) as a sphere
4.12 RMSF comparison of WT and N173K
4.13 Salt bridges involve in HA, HB and HC are plotted over 50 ns of simulation time. Presence of salt bridge is shown in white
4.14 Electrostatic potential surface comparison of WT and Q212P
4.15 Electrostatic Potential Surface Comparison of WT and Q212H
4.16 Electrostatic Potential Surface Comparison of WT and V203I
4.17 Electrostatic Potential Surface Comparison of WT and V203G
4.18 Electrostatic Potential Surface Comparison of WT and N171S
4.19 Electrostatic Potential Surface Comparison of WT and N173K
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PrP</td>
<td>Prion Protein</td>
</tr>
<tr>
<td>PrPc</td>
<td>Cellular Prion Protein (Normal Isoform)</td>
</tr>
<tr>
<td>PrPSc</td>
<td>Scrapie Prion Protein (Pathogenic Isoform)</td>
</tr>
<tr>
<td>HuPrP</td>
<td>Human Prion Protein</td>
</tr>
<tr>
<td>GPI</td>
<td>Glycophosphatidylinositol</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>aa</td>
<td>Amino Acid</td>
</tr>
<tr>
<td>S</td>
<td>Sheet</td>
</tr>
<tr>
<td>H</td>
<td>Helix</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>TSE</td>
<td>Transmissible Spongiform Encephalopathy</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamic</td>
</tr>
<tr>
<td>WT</td>
<td>Wild Type</td>
</tr>
<tr>
<td>CJD</td>
<td>Creutzfeldt-Jakob Disease</td>
</tr>
<tr>
<td>iCJD</td>
<td>Iatrogenic Creutzfeldt-Jakob Disease</td>
</tr>
<tr>
<td>fCJD</td>
<td>Familial Creutzfeldt-Jakob Disease</td>
</tr>
<tr>
<td>FFI</td>
<td>Fatal Familial Insomnia</td>
</tr>
<tr>
<td>GSS</td>
<td>Gerstmann-Straussler-Scheinker Disease</td>
</tr>
<tr>
<td>sCJD</td>
<td>Sporadic Creutzfeldt-Jakob Disease</td>
</tr>
<tr>
<td>vCJD</td>
<td>Variant Creutzfeldt-Jakob Disease</td>
</tr>
<tr>
<td>SB</td>
<td>Salt Bridge</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-Mean-Square Deviation</td>
</tr>
<tr>
<td>RMSF</td>
<td>Root-Mean-Square Fluctuation</td>
</tr>
<tr>
<td>SASA</td>
<td>Solvent Accessible Surface Area</td>
</tr>
<tr>
<td>nm</td>
<td>Nanometer</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Background of the problem

The prion protein (PrP) roots from a collection of diseases identified as transmissible spongiform encephalopathies (TSEs) (Vila-Viçosa et al. 2012). Neurodegenerative disorders with sporadic, infectious and genetic forms are common diseases found in both human and animals which are caused by conversion of PrP from its normal cellular form (PrP$^C$) to a misfolded, oligomeric isoform (PrP$^{Sc}$) (Xu et al. 2012) (De Simone et al. 2013; Thirumalai 2013). In fact, prion diseases are misfolded proteins, which accumulate inside the central nervous system and produce an infectious disease. This has been remained a debate and is considered the main idea of the Prion Hypothesis, because it contrasts compared to other known infectious agents (virus/bacteria/fungus/parasite) which must contain nucleic acids (either DNA, RNA, or both). Prion diseases are considered fatal and until now no known treatment has been reported. Some examples of TSEs include fatal familial insomnia in humans, bovine spongiform encephalopathy in cattle, Scrapie in sheep, and Creutzfeldt–Jakob disease, Gerstmann–Sträussler–Scheinker syndrome and kuru (Vila-Viçosa et al. 2012; Thirumalai 2013).

PrP is monomeric and soluble in its normal cellular form, however its pathogenic form (PrP$^{Sc}$) contains a high β-structure content which may accumulate and form amyloid fibrils (Vila-Viçosa et al. 2012). PrP$^{Sc}$, resulting from the misfolding
of PrP<sup>C</sup>, gathers extra PrP<sup>C</sup> to form deposits of accumulated PrP<sup>Sc</sup> in the affected individual’s brain (Spevacek 2012). Different causes have been proposed for the conversion of PrP<sup>C</sup> into PrP<sup>Sc</sup>, however the mechanism of the conversion still remains unanswered (Vila-Viçosa et al. 2012).

The chemical composition of amino acid sequences of PrP<sup>C</sup> and PrP<sup>Sc</sup> are identical but their difference lies in their conformations specifically in terms of secondary structure (Vila-Viçosa et al. 2012). The structure of PrP<sup>Sc</sup> has the tendency to assemble into insoluble aggregates, hence NMR spectroscopy and X-ray crystallography are not able to determine the structure of this protein so the structure still remains unresolved. According to limited number of experiments on PrP<sup>Sc</sup> using the CD spectroscopy and Fourier transform, it was observed that compared to PrP<sup>C</sup>, PrP<sup>Sc</sup> are less helical and consist of more β-structures. Hence, it is logical to conclude that in the transition from PrP<sup>C</sup> to PrP<sup>Sc</sup> there exists a conversion of α-helices to β-structures (Xu et al. 2012). PrP<sup>C</sup> α-helix and β-sheet content is approximately 42% and 3% respectively, while PrP<sup>Sc</sup> α-helix content decreases to ∼30% and its β-sheet content about 43% (Mishra 2010; Vila-Viçosa et al. 2012).

PrP is located on chromosome 20 in humans and encoded in the PRNP gene, and it is expressed as a 254 amino acid protein, in which the first 22 amino acids are removed after they signal to translocate the nascent protein to the endoplasmic reticulum (Spevacek 2012). With the addition of a glyco-phosphatidyl-inositol (GPI) anchor to produce a 209 amino acid protein, PrP (amino acids (aa) 23-231), the C-terminal 23 amino acids are cleaved (Spevacek 2012). Several structures of human PrP (huPrP) exist in the Protein Data Bank. Based on these data, the huPrP includes an N-terminal region with no distinct structure (aa 23–124) and globular domain in the C-terminal region (aa 125–228) composed of three α-helices (HA: aa 144–156; HB: aa 174–194; HC: aa 200–228) and short two-stranded antiparallel β-structure (S1–S2) (aa 128–131 + aa 161–164) (Figure 1.1). It has a disulfide bond between C179 and C214 and two Nglycosylation sites (N181 and N197). The N-terminal region is characterized by octarepeats (aa 51–91) that seem to develop a structure in the presence of Cu2+ or other metals (Vila-Viçosa et al. 2012).
Figure 1.1  Schematic of the prion protein. The flexible N-terminus contains the octarepeat domain. The structured C-terminal domain contains two β-strands (arrows) and three α –helices (boxes). The gray circles stand for the N-linked glycosylations. The C-terminus is with a GPI anchor, which connects the protein to the cell surface (Spevacek 2012)

Currently, Prion diseases in humans are grouped based on whether they are sporadic, inherited, or acquired. In majority of cases prion disease are sporadic; meaning, they develop spontaneously with no known reason. On the other hand, prion disease is rarely inherited due to a faulty gene, and less often acquired through medical procedures, transfusions, or contaminated food.

In the case of inherited prion disease, which is also known as familial prion disease, abnormal prion proteins are produced in the body caused by a genetic mutation (fault in the gene that codes for the prion protein). The abnormal prion proteins are predisposed to undergo the change in shape, which induces the production of rogue proteins.

It is known that conformation transition from PrP$^C$ to PrP$^{Sc}$ is induced with the replacement of some amino acids in PrP. Human familial prion diseases are associated with about 40 point mutations of PRNP, and most of these mutations are located in the globular domain of the protein (Guo, Ren, et al. 2012b). This purely structural rearrangement between PrP$^C$ and PrP$^{Sc}$ provides an opportunity for investigation from a dynamics point of view. Experimentally, it is unfeasible to distinguish the actual conformational transition or folding/unfolding process between these two states. As mentioned before the hydrophobic nature of PrP$^{Sc}$ prohibit successful structure determination by x-ray crystallography, or NMR spectroscopy, molecular dynamic
(MD) simulation has become one of the few available methods for elucidating PrP\textsuperscript{C} to PrP\textsuperscript{Sc} conversion (Shamsir et al. 2008).

MD simulation has obvious advantages to study the structural conversion, as it cannot only provide plentiful dynamic structural information but also model the required environment for conversion easily. Actually, this method has been applied successfully to explore how the mutations affect the conversion between PrP\textsuperscript{C} and PrP\textsuperscript{Sc} (Guo, Ning, et al. 2012a).

1.2 Objectives of the Study

The purpose of this study is to:

1) Identify prion mutations as controls for Q212H, V203G and N173K mutations.
2) Run molecular dynamics simulation for the mutations and their controls
3) Analyse the effects of mutations on the protein structure, stability and intermolecular interactions.

1.3 Scope of the Study

Applying 50 ns of MD simulations in order to investigate three unknown-disease-related PrP mutations (Q212H, V203G, N173K) and an additional MD simulation run for three controls consisting of two pathogenic mutations (Q212P, V203I) taken as positive controls and a neutral polymorphism (N171S) as negative control and a MD simulation of WT as a basis for comparison of all mutations.
1.4 Problem Statement

By running MD simulation for the identified pathogenic mutations (positive controls) and polymorphism (negative control) in human PrP along with the WT PrP we could explore the effect of mutations on WT PrP. By this means, MD simulation could be performed on Q212H, V203G and N173K (the unknown-disease-related mutations) and the results could be compared with the controls. This will help us to explore the dynamic similarity between Q212H, V203G, N173K and prion pathogenic mutations.

1.5 Significance of the Study

Mutations in prion protein gene can produce diverse clinical phenotypes. Hence, diagnosing whether a mutation in prion protein is causal of the disease, becomes difficult and need further examination, which is time consuming and costly. By using MD simulation, present study explores mutations that are not known whether they are pathogenic. This research could serve as a step forward to acquiring a more rapid, economical and reliable approach for a definite distinction of non-pathogenic PrP mutations from pathogenic mutations.
REFERENCES


Thirumalai, J.C.A.D., 2013. Helices 2 and 3 are the initiation sites in the PrPC → PrPSC transition. pp.1–32.


