INVESTIGATING THE MUTATIONS AND STRUCTURE OF SHORT PART OF COLLAGEN TYPE I IN OSTEOGENESIS IMPERFECTA TYPE III

ASAL SAFI SAMGHABADI

A thesis submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Science (Biotechnology)

Faculty of Bioscience and medical engineering

U n i v e r s i t i T e c h n o l o g i M a l a y s i a

August 2013

To my beloved family

ACKNOWLEDGMENT

I would like to begin by sincerely thanking my supervisor, Associated Prof Dr Mohd Shahir Shamsir for his constant support, guidance and mentorship over the course of this thesis. He gave me the freedom to define my thesis statements. I am very grateful to having the opportunity for work on this thesis.

I am heartily thankful to my parents, for their unconditional supports and encouraging me to continue my education.

ABSTRACT

The extracellular matrix for connective tissues displays an abstruse alloy of changeable members of various protein families describing structural validity and different physiological functions. The most important component of extracellular matrix is collagen. Collagens' large family in humans is consisting of at least 27 members (described as being type I to XXVII) with 42 distinguished polypeptide chains. Collagen type I is the most plentiful and extensively expressed collagen in human which is comprising of 2 α1 chains and one α2 chain and their gens are located at the different loci,*COL1A1* and *COL1A2* respectively. Mutation at these loci causes various connective tissue disorders such as Osteogenesis Imperfecta(OI). Osteogenesis Imperfecta (OI) or Vrolik's syndrome is a heterogeneous group of inherited conditions which has different types but, Type III is individualized among the other classifications because this type is the "Progressive Deforming" type. This study reports a comparative investigation on a short sequence of α1 and α2 chains near the c-terminal positions by the means of mutated and normal cases. Our focus in this study was on two point mutations near the end of each chain which in α1 are Gly973Ser & Gly1009Ser and in α2 are Gly997Asn & Gly1006Ala by extracting our desired parts of α1&α2 sequences from UniProt and modeling them by using EsyPred3D with applying the most identical pdb file obtained from rcsb & BLAST as template. After visualization and doing mutation by PyMOL, the differences in their sequence have been investigated by Jalview. Our obtained results from simulation of normal and mutated structures of α1&α2 by GROMACS software demonstrated that substitution of Gly by Ser in α1 and by Ala in α2 affected on the function of protein, on the other hand, substitution of Gly by Asn in α2 effected on the structure of protein.

ABSTRAK

Matriks extracellular (atau matriks luar sel) untuk tisu penghubung mempamerkan ahli aloi berubah yang sukar difahami serta terdiri daripada pelbagai kumpulan protein yang menerangkan kesahihan struktur dan fungsi fisiologi yang berbeza. Kolagen jenis I adalah kolagen terbanyak yang diekspress secara meluas dalam badan manusia dan ia terdiri daripada beberapa rantai 2 α1 dan satu rantai α2 serta. Mutasi pada lokus ini menyebabkan pelbagai gangguan dalam tisu penghubung seperti Osteogenesis Imperfecta (OI). Osteogenesis Imperfecta (OI) atau Vrolik sindrom adalah satu kumpulan heterogeneous yang diwarisi dan terdiri daripada pelbagai jenis tetapi, Jenis III telah diklasifikasikan sebagai individu berbanding dengan jenis-jenis yang lain kerana ia adalah jenis yang "Progresif Deformasi" atau “Progresif yang mencacatkan bentuk”. Kajian ini melaporkan tentang penyelidikan mengenai urutan singkat rantaian α1 dan α2 yang berada berhampiran c-terminal dan dikaji dengan membuat perbandingan antara kes mutasi dan normal. Fokus kajian ini adalah pada dua titik mutasi yang berhampiran dengan hujung setiap rantaian di mana, pada α1 adalah Gly973Ser & Gly1009Ser dan α2 adalah Gly997Asn & Gly1006Ala, dengan mengektrak bahagian urutan α1 & α2 yang dikehendaki dari UniProt dan membina model melalui EsyPred3D dengan menggunakan fail pdb yang sama (diperolehi daripada RCSB & BLAST) sebagai template. Selepas visualisasi dan melakukan mutasi melalui PyMOL, perbezaan dalam urutan telah dikaji denagn menggunakan Jalview. Hasil yang diperolehi dari simulasi struktur normal dan mutasi daripada α1 & α2 melalui perisian GROMACS menunjukkan bahawa penggantian GLY kepada SER dalam α1 dan kepada ALA dalam α2 menjejaskan fungsi protein manakala penggantian GLY kepada Asn dalam α2 memberi efek kepada struktur protein.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF FORMULA

LIST OF ABBREVIATIONS

A1	Alpha 1
A2	Alpha 2
OI	Osteogenesis Imperfecta
COL1A1	Collagen type I alpha 1
COL1A2	Collagen type I alpha 2
AD	Autosomal Dominant
AR	Autosomal Recessive
NCBI	National Center for Biotechnology Information
PDB	Protein Data Bank
BLAST	Basic Local Alignment Search Tool
MSA	Multiple Sequence Alignment
GROMACS	GROningrn Machine for Chemical Simulation
A1m1	Alpha 1 with first mutation
A1m2	Alpha 1 with second mutation
A2m1	Alpha 2 with first mutation
A2m2	Alpha 2 with second mutation
ECM	Extra Cellular Matrix
A^{o}	Angstrom
Cα	Carbon alpha
C-terminal	Carboxy-terminal
N-terminal	Amino-terminal
FACIT	Fibril Associate
Kb	Kilo base pair
cDNA	cyclic DNA

CM	Centimeter
UniProt	Universal Protein resource
UniRef	UniProt Refrence Cluster
UniProtKB	UniProt Knowledgebase
UniPrac	UniProt Archive
UniMES	UniProt Metagenomic and Environmental Sequrnces
SIB	Swiss Institute of Bioinformatics
EBI	European Information Resource
PIR	Protein Information Resource
PIR-PSD	Protein Information Resource-Protein Sequence Database
TrEMBL	Translated EMBL Nucleotide Sequence Data Library
3D	3 Dimension
NMR	Nuclear Magnetic Resonance
ID	Identity Document
DNA	Deoxyribonucleic Acid
RNAse	Ribonucleic Acidase / Ribosomal Ribonucleic Aci
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
A	Alanine
P	Proline
B	Aspartate or Aspargine
Q	Glutamine
C	Cysteine
R	Arginine
D	Aspartate
S	Serine
E	Glutamate
T	Throeonine
F	Phenylalanine

U	Selenocysteine
G	Glycine
V	Valine
H	Histidine
W	Tryptophane
I	Isoleucine
Y	Tyrosine
K	Lysine
Z	Glutamate or Glutamine
L	Leucine
M	Methionine
N	Asparagine

CHAPTER 1

INTRODUCTION

1.1 Background

Collagen is a family of natural protein which can be found in animals, especially in the corpus and connective tissues of vertebrates (Müller, 2003). Collagen is the primary ingredient of the connective tissues, and is the most numerous types of protein found in mammals (Di Lullo, Sweeney, Körkkö, Ala-Kokko, & San Antonio, 2002). The elongated form of collagen fibrils accounted between 25% and 35% of the total amount of protein in the whole body of most animals. Collagen is the main constructive protein which organizes molecular cables that causes the strengthening in tendons and the vast sheets that compose the skin and interior organs. The addition of mineral crystals to collagen allows the formation of bones and teeth. Collagen found in skeleton protects the softer tissues and keeps them connected.

Although collagen possesses numerous irreplaceable obligatory functions, it is a protein with relatively simple structure. Collagen is made of a triple helix which contain two identical α1 chains and an extra α2 chain with a only a small differences in its chemical combination to differentiate them (Orgel, San Antonio, & Antipova, 2011) , Each subunits of collagen molecule consisted of repeating triple motifs of Gly-X-Y

where the combination and presence of glycine amino acids in the structure of α chains and in each third amino acid allows the triple chains structure to self-assemble into a right handed triple helix which is then supported by hydrogen bonds and other ionic interactions (Hulmes, 2002).

So far, 28 different types of collagen molecules have been identified which types I to V are five common types. Some collagen types are continuously expressed, while the rest of them have limitation and each type has its own function. Collagen type 1 is the most common type of collagen with the highest rate of expression in humans, making it the most extensively studied and characterized (Prockop, 1995). One of the most popular diseases which caused by mutation in collagen type 1 is Osteogenesis Imperfecta.

Osteogenesis Imperfecta (OI), sometimes known as **brittle bone disease**, or "Lobstein syndrome" (Kid)) is an inheritable bone disease. People who are afflicted with this disorder exhibit imperfect connective tissue due to the inability to produce and resulting in the shortage of the Type 1 collagen (Frank Rauch & Glorieux, 2004). This shortage is caused by the substitution of glycine amino acid to a bigger amino acid in the structure of triple helix, causing the formation of inflated regions inside of collagen structure, which effect both the intra molecular interaction and the neighboring molecular interactions (Gautieri, Uzel, Vesentini, Redaelli, & Buehler, 2009), This resulted in the formation of incorrect collagen structure between collagen fibrils during the formation of the bone and hydroxyapatite crystal and causing brittleness.

In the case of genetic disorder, OI is caused by autosomal dominant disturbance of collagen type1. Most of this disorder are caused by mutations in both COL1A1 and COL1A2 genes which encode the α1 and α2 chains of type 1 collagen respectively (Drögemüller et al., 2009)

There are five types of Osteogenesis Imperfecta classification:

a) Type I: AD (autosomal dominant), normal structure, blue sclerae, loss hearing, little deformity
b) Type II: AD, long bone fractions, lethal form,
c) Type III: AD & AR (autosomal recessive): server form, blue sclerae, loss hearing, oncoming deforming, shorter length, multiple breaks.
d) Type IV: AD, softly server, variable short length, normal sclerae.
e) Type V: caused by hypertrophic calluses.(Pagon et al., 2005)
f) In this study, we focused on osteogenesis Imperfecta type III

1.2 Problem Statement:

Currently all the studies on the two classes of mutations in collagen reported in Osteogenesis Imperfecta are biochemical in nature. There are no detailed studies on the structural conformation and dynamics of the differences between the mutated and the wild type variants.

1.3 The aim of study

The aim of this study is to investigate the possible effects of two point mutations on the tertiary structure of collagen type I as well as study the importance of positions of mutations on stability of protein and how they can be related to the symptoms of Osteogenesis Imperfecta type III. The data for this project will be based on experimental

results stored in bioinformatics databases such as the NCBI (http://www.ncbi.nlm.nih.gov), PDB (http://www.rcsb.org) and Gene bank (www.ncbi.nlm.nih.gov/genbank/). The data from these databases will then be analyzed using bioinformatics and visualization software such as BLAST (blast.ncbi.nlm.nih.gov/), PYMOL (www.pymol.org/) and the Jalview (www.jalview.org/) for multiple sequence alignment (MSA) then molecular dynamics simulation software such as GROMACS (www.gromacs.org/) to compare the stability and flexibility of the structures of normal and mutant types of α1 and α2 chains in collagen type I. The study aims to identify the differences in: sequences, structures, impact of mutations and stability of both chains in two mutant and normal type condition.

1.4 Objectives

The objectives of this study were:

1. Determining the differences of the sequence, structure and conformation between the wild-type and mutant type structure.
2. Structural and dynamics comparison and analysis of the implication of the differences in amino acid sequence in Osteogenesis Imperfecta Type III
3. Investigate the impacts of two point mutations on stability of both α1 and α2 chains of collagen type I in Osteogenesis Imperfecta type III.

1.5 Scope of Study

The data for this project will be sourced by bioinformatics databases that store experimental information such as the NCBI (http://www.ncbi.nlm.nih.gov), the Protein Data Bank (www.rcsb.org), the UniProt (http://www.uniprot.org) and the Genbank (http://www.ncbi.nlm.nih.gov/genbank/).

1.6 Significance of the study

This study will help identify and characterize the structural effects of point mutations and how it influences the structure and conformational changes via molecular dynamic simulations. The obtained output enables researchers to gain useful information regarding the collagen type I stability and its importance on various related diseases such as Osteogenesis Imperfecta type III. The findings of this study can provide information for research in collagen type I and its associated diseases.

REFERENCES:

Altschul, Stephen F, Gish, Warren, Miller, Webb, Myers, Eugene W, & Lipman, David J. (1990). Basic local alignment search tool. ***Journal of molecular biology, 215*****(3), 403-410.**

Asara, John M, Schweitzer, Mary H, Freimark, Lisa M, Phillips, Matthew, & Cantley, Lewis C. (2007). Protein sequences from mastodon and Tyrannosaurus rex revealed by mass spectrometry. ***Science, 316*****(5822), 280-285.**

Astbury, William Thomas. (1938). The fourth Spiers memorial lecture. X-ray adventures among the proteins. ***Transactions of the Faraday Society, 34(350),*** **378-388.**

Baljet, B. (2002). Aspects of the history of Osteogenesis imperfecta (Vrolik's syndrome). ***Annals of Anatomy-Anatomischer Anzeiger, 184*****(1), 1-7.**

Barsh, Gregory S, Roush, Christine L, Bonadio, Jeffrey, Byers, Peter H, & Gelinas, Richard E. (1985). Intron-mediated recombination may cause a deletion in an alpha 1 type I collagen chain in a lethal form of osteogenesis imperfecta. ***Proceedings of the National Academy of Sciences, 82*****(9), 2870-2874.**

Bateman, John F, Lamande, Shireen R, & Ramshaw, John AM. (1996). 2 Collagen Superfamily. ***Extracellular matrix*****, 22(34), 345-456.**

Bhattacharjee, Arnab, & Bansal, Manju. (2005). Collagen structure: the Madras triple helix and the current scenario. ***Iubmb Life, 57*****(3), 161-172.**

Bilezikian, John P, Raisz, Lawrence G, & Martin, T John. (2008). ***Principles of bone biology, two-volume set*** **(Vol. 1): Academic Press, 23(78), 32-56.**

Boutet, Emmanuel, Lieberherr, Damien, Tognolli, Michael, Schneider, Michel, & Bairoch, Amos. (2007). Uniprotkb/swiss-prot ***Plant Bioinformatics*** **(pp. 89-112): Springer, 44(98), 45-89.**

Brinckmann, Jürgen. (2005). Collagens at a glance *Collagen* (pp. 1-6): Springer, 56(65), 67-98.

Buckley, Mike, Walker, Angela, Ho, Simon YW, Yang, Yue, Smith, Colin, Ashton, Peter, Penkman, Kirsty. (2008). Comment on" Protein Sequences from Mastodon and Tyrannosaurus rex Revealed by Mass Spectrometry". *Science, 319*(5859), 33-33.

Burgeson, Robert E, & Nimni, Marcel E. (1992). Collagen types. Molecular structure and tissue distribution. *Clinical orthopaedics and related research, 282(56)*, 250-272.

Byers, Peter H, Wallis, Gillian A, & Willing, Marcia C. (1991). Osteogenesis imperfecta: translation of mutation to phenotype. *Journal of medical genetics, 28*(7), 433-532.

Chu, Mon-Li, Williams, Charlene J, Pepe, Gugliemina, Hirsch, Jeffrey L, Prockop, Darwin J, & Ramirez, Francesco. (1983). Internal deletion in a collagen gene in a perinatal lethal form of osteogenesis imperfecta. *Nature, 304*(5921), 78-80.

Cole, William G. (1994). Collagen genes: mutations affecting collagen structure and expression. *Progress in nucleic acid research and molecular biology, 47(56)*, 29-80.

Colige, Alain, Sokolov, Boris P, Nugent, Paul, Baserga, Renato, & Prockop, Darwin J. (1993). Use of an antisense oligonucleotide to inhibit expression of a mutated human procollagen gene (COL1A1) in transfected mouse 3T3 cells. *Biochemistry, 32*(1), 7-11.

Dalgleish, Raymond. (1997). The human type I collagen mutation database. *Nucleic acids research, 25*(1), 181-187.

Dalgleish, Raymond. (1998). The human collagen mutation database 1998. *Nucleic acids research, 26*(1), 253-255.

Di Lullo, Gloria A, Sweeney, Shawn M, Körkkö, Jarmo, Ala-Kokko, Leena, & San Antonio, James D. (2002). Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. ***Journal of Biological Chemistry, 277*****(6), 4223-4231.**

Drögemüller, Cord, Becker, Doreen, Brunner, Adrian, Haase, Bianca, Kircher, Patrick, Seeliger, Frank, Leeb, Tosso. (2009). A missense mutation in the SERPINH1 gene in Dachshunds with osteogenesis imperfecta. ***PLoS genetics, 5*****(7),341-456.**

Frenkel, SALLY R, Toolan, BRIAN, Menche, DAVID, Pitman, MARK I, & Pachence, JAMES M. (1997). Chondrocyte transplantation using a collagen bilayer matrix for cartilage repair. ***Journal of Bone & Joint Surgery, British Volume, 79*****(5), 831-836.**

Gajko-Galicka, Anna. (2002). Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans. ***ACTA Biochemica Polonica-Eniglish Edition, 49*****(2), 433-442.**

Gautieri, Alfonso, Uzel, Sebastien, Vesentini, Simone, Redaelli, Alberto, & Buehler, Markus J. (2009). Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils. ***Biophysical journal, 97*****(3), 857-865.**

Gelse, K, Pöschl, E, & Aigner, T. (2003). Collagens—structure, function, and biosynthesis. ***Advanced drug delivery reviews, 55*****(12), 1531-1546.**

Hess, Berk, Kutzner, Carsten, van der Spoel, David, & Lindahl, Erik. (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. ***Journal of chemical theory and computation, 4*****(3), 435-447.**

Hulmes, David JS. (2002). Building collagen molecules, fibrils, and suprafibrillar structures. ***Journal of structural biology, 137*****(1), 2-10.**

Kid, X-Ray OI Type V. Osteogenesis imperfecta, 42(87), 98-123.

Kielty, Cay M, & Grant, Michael E. (2003). The collagen family: structure, assembly, and organization in the extracellular matrix. ***Connective Tissue and Its Heritable Disorders: Molecular, Genetic, and Medical Aspects, Second Edition*****, 44(87), 159-221.**

Laakso, Risto. (2005). Protein structure analysis, 32(45), 45-98.

MARK, Helga, AUMAILLEY, Monique, WICK, Georg, FLEISCHMAJER, Raul, & TIMPL, Rupert. (1984). Immunochemistry, genuine size and tissue localization of collagen VI. *European Journal of Biochemistry, 142*(3), 493-502.

Müller, Werner EG. (2003). The origin of metazoan complexity: Porifera as integrated animals. *Integrative and Comparative Biology, 43*(1), 3-10.

Myllyharju, Johanna, & Kivirikko, Kari I. (2004). Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics, 20*(1), 33-43.

Nassa, Manisha, Anand, Pracheta, Jain, Aditi, Chhabra, Aastha, Jaiswal, Astha, Malhotra, Umang, & Rani, Vibha. (2012). Analysis of human collagen sequences. *Bioinformation, 8*(1), 26.

Orgel, JPRO, San Antonio, JD, & Antipova, O. (2011). Molecular and structural mapping of collagen fibril interactions. *Connective Tissue Research, 52*(1), 2-17.

Pagon, Roberta A, Bird, Thomas D, Dolan, Cynthia R, Stephens, Karen, Adam, Margaret P, Steiner, Robert D, Byers, Peter H. (2005). Osteogenesis Imperfecta, 45(98), 120-140.

Pauling, Linus, & Corey, Robert B. (1951a). Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proceedings of the National Academy of Sciences of the United States of America, 37*(5), 235-451.

Pauling, Linus, & Corey, Robert B. (1951b). The structure of fibrous proteins of the collagen-gelatin group. *Proceedings of the National Academy of Sciences of the United States of America, 37*(5), 272.

Prockop, JD. (1995). Collagens: molecular biology, diseases, and potentials for therapy. *Annual review of biochemistry, 64*(1), 403-434.

Rauch, F, Travers, R, Parfitt, AM, & Glorieux, FH. (2000). Static and dynamic bone histomorphometry in children with osteogenesis imperfecta. *Bone, 26*(6), 581-589.

Rauch, Frank, & Glorieux, Francis H. (2004). Osteogenesis imperfecta. *The Lancet, 363*(9418), 1377-1385.

Reeder, Janet, & Orwoll, Eric. (2006). Adults with osteogenesis imperfecta. *New England Journal of Medicine, 355*(26),1098-1231.

Rich, Alexander, & Crick, FHC. (1961). The molecular structure of collagen. *Journal of molecular biology, 3*(5), 483-IN484.

Sasisekharan, V. (1962). Stereochemical criteria for polypeptide and protein structures. *Collagen (N. Ramanathan, ed.)89(65)*, 39-78.

Sato, Keiji, Yomogida, Kentaro, Wada, Takayuki, Yorihuzi, Tetuya, Nishimune, Yoshitake, Hosokawa, Nobuko, & Nagata, Kazuhiro. (2002). Type XXVI collagen, a new member of the collagen family, is specifically expressed in the testis and ovary. *Journal of Biological Chemistry, 277*(40), 37678-37684.

Schuppan, Detlef, Somasundaram, Rajan, & Just, Martin. (1991). *The extracellular matrix: a major signal transduction network.* Paper presented at the Cellular and Molecular Aspects of Cirrhosis: Proceedings of the International Conference on" Cellular and Molecular Bases of Liver Cirrhosis" Held in Rennes (France) on July 3-5, 1991.

Schwede, Torsten, Kopp, Jürgen, Guex, Nicolas, & Peitsch, Manuel C. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic acids research, 31*(13), 3381-3385.

Schweitzer, Mary Higby, Suo, Zhiyong, Avci, Recep, Asara, John M, Allen, Mark A, Arce, Fernando Teran, & Horner, John R. (2007). Analyses of soft tissue from Tyrannosaurus rex suggest the presence of protein. *Science, 316*(5822), 277-280.

Shoulders, Matthew D, & Raines, Ronald T. (2009). Collagen structure and stability. ***Annual review of biochemistry, 9(65),78,*** **929.**

Starr, Stephanie R, Roberts, Timothy T, & Fischer, Philip R. (2010). Osteogenesis imperfecta: primary care. ***Pediatrics in Review, 31*****(8), e54-e64.**

Taipale, Jussi, Miyazono, Kohei, Heldin, Carl-Henrik, & Keski-Oja, Jorma. (1994). Latent transforming growth factor-beta 1 associates to fibroblast extracellular matrix via latent TGF-beta binding protein. ***The Journal of cell biology, 124*****(1), 171-181.**

Van der Rest, M, & Garrone, R. (1991). Collagen family of proteins. ***The FASEB Journal, 5*****(13), 2814-2823.**

Von der Mark, Klaus. (1999). ***Structure, biosynthesis and gene regulation of collagens in cartilage and bone*****: Academic Press, Orlando, 7(9), 90-110.**

Vuorio, Eero, & De Crombrugghe, Benoit. (1990). The family of collagen genes. ***Annual review of biochemistry, 59*****(1), 837-872.**

Willing, Marcia C, Deschenes, Sachi P, Slayton, Rebecca L, & Roberts, Erik J. (1996). Premature chain termination is a unifying mechanism for COL1A1 null alleles in osteogenesis imperfecta type I cell strains. ***American journal of human genetics, 59*****(4), 799-980.**

Wu, Cathy H, Apweiler, Rolf, Bairoch, Amos, Natale, Darren A, Barker, Winona C, Boeckmann, Brigitte, Lopez, Rodrigo. (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. ***Nucleic acids research, 34*****(suppl 1), D187-D191.**