Xè Premi PRBB al millor treball de recerca en Ciències de la Salut i de la Vida

2015

Treball guanyador del 3r premi

The brittle and the unbreakable: an approach to mutation in collagen type I genes

Natalia Espasandín Hueter

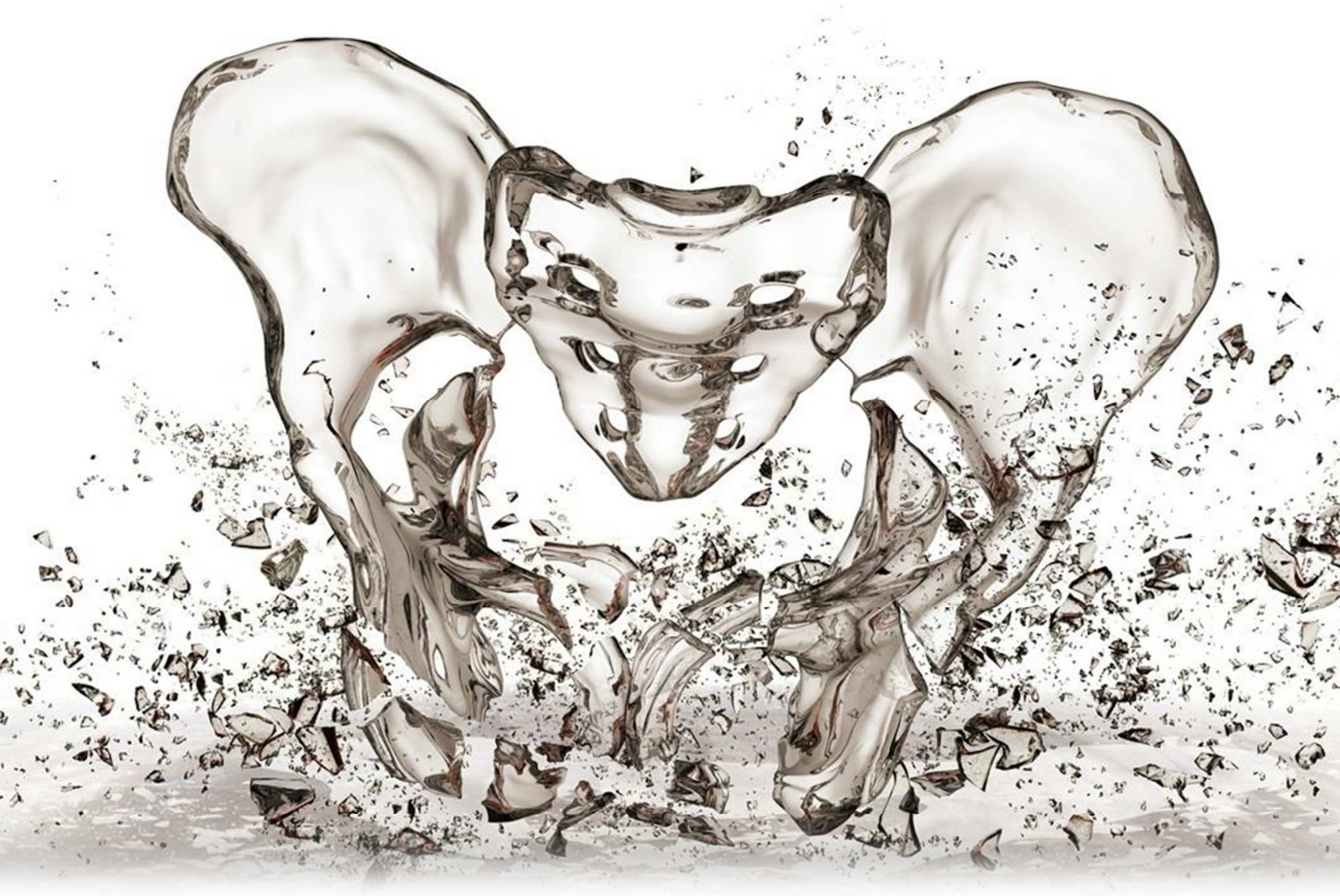
Tutora: Cristina Llombart Clara

Institut Puig Cargol (Sant Antoni de Calonge)

THE BRITTLE AND THE UNBREAKABLE

AN APPROACH TO MUTATIONS IN COLLAGEN TYPE I GENES

TREBALL DE RECERCA CURS 2014 - 2015



ALUMNA	TUTORA	CURS
NATALIA ESPASANDÍN HUETER	CRISTINA LLOMBART CLARA	2014 - 2015

Generalitat de Catalunya
Departament d'Ensenyament
Institut Puig Cargol

Pla de Calonge, s/n
17252 Sant Antoni de Calonge
Tel. 972 652 598
Fax 972 653 466
b7000706@xtec.cat
http://www.xtec.cat/iespuigcargol-calonge/

THE BRITTLE AND THE UNBREAKABLE

An approach to mutations in collagen type I genes

TREBALL DE RECERCA

Alumna: Natalia Espasandín Hueter

Tutora: Cristina Llombart Clara

Institut Puig Cargol

Departament de ciències

Curs acadèmic 2014 – 2015

Data d'entrega: 12 de gener de 2015

ABSTRACT

Osteogenesis imperfecta (OI) is a genetic disorder characterized by extreme bone fragility and fracture proneness. It is caused by collagen type I alterations, and the genotype-phenotype correlations have been studied in this research project. The data was taken from the LOVD database and analysed searching different relationships between the mutations and the resulting severity of the disorder. It has been determined that there are different regional models that correlate with ligand binding sites for chaperones and enzymes. The role of different cellular processes and non-collagenous proteins was also assessed. Although almost all the mutations correlated in some way with the OI subtype, further investigation and research is needed in order to improve the quality of life of these patients through better genetic counselling and development of gene therapy.

To Eric and Luis Evangelista, we will always remember you.

To all those who are fighting against rare diseases.

"We used to think our fate was in our stars. Now we know,
in large measure, our fate is in our genes."

James D. Watson

– ACKNOWLEDGEMENTS –

I would like to express my greatest gratitude to the people who have helped and supported me throughout my research project. I am grateful to my tutor, Cristina Llombart, for her continuous support for the project from initial advice in the early stages of conceptual inception and through ongoing advice and encouragement to this day.

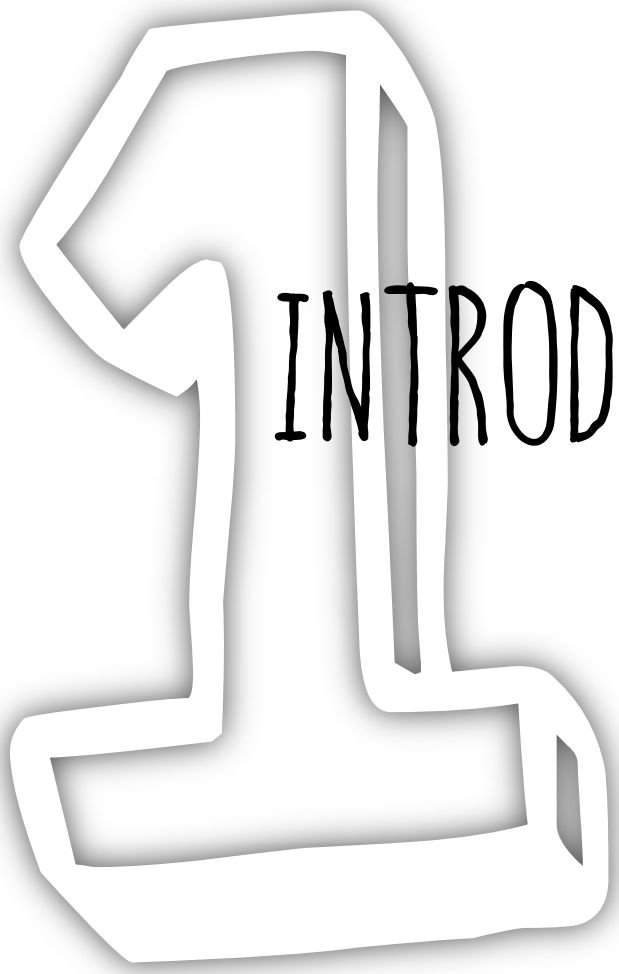
I would like to express my great appreciation to Dr Silverman for his valuable and constructive suggestions during the planning and development of this research project. I am equally grateful to all the health care professionals and researchers that contributed with their part to this research project.

I would also like to thank all the patient associations for their support, especially Federación Española de Enfermedades Raras, Federación Asperger España, Asociación Huesos Cristal España and Brittle Bone Society.

A special thank of mine goes to Eric Evangelista and his father, Dr Evangelista. Although they are no longer with us, they have always animated me to realise this research project. They have been my inspiration and motivation for the subject and have and will always give me enough strength to never give up.

Last but not least, I want to thank my family and friends for their undivided support and interest who inspired me and encouraged me to go my own way, without whom I would be unable to complete my project.

- CONTENTS -



INTRODUCTION

1. 1. SUBJECT CHOICE

The subject for this research project arose from a simple question I asked myself some months ago. But before I say which question it is, I must explain its background.

There are more than three million people in Spain living with a rare disease. We know about six thousand different rare diseases and most of them are genetic and severely life-threatening. I'm one among those three million people, and so was my late friend Eric.

We have got a collagen-synthesis defect. I have got a rather mild form but he had the most severe type of our disease. The only difference between us was that I lost about half of the gene, while he only had a single base substitution.

Nonetheless, I did not come up with the question until I was in a biology class and the teacher said: 'the most aberrant mutations are deletions and insertions while substitutions are in most cases from minor importance'. From this latter sentence came my question:

*Why did he had to pass away if he **just** had a substitution?*

* * *

Doing this research paper I want to find an answer to my question. I know it is a big deal for such a young girl, but I will try it.

At the beginning I wanted to do it through the study of my own condition, but then I realised that it would be uncomfortable for me and very difficult since only a few cases are reported. I looked for another possibility and I found out that this 'genetic phenomenon' also occurs in many other collagen-related disorders. Then I decided to study our sister-disease, osteogenesis imperfecta.

I could have chosen any other disorder but osteogenesis imperfecta is the most similar to mine: "brittle on the outside, unbreakable on the inside."

1. 2. OBJECTIVES

The main objective of this research project is to find out why mutations in collagen genes are so singular and differ from many other genes. Nonetheless, this objective can be subdivided in others that comprise different parts of the investigation. Moreover, there are other minor purposes that want to be achieved doing this research project.

The main theoretical objectives are:

- Approach rare diseases to people who never have heard about them.
- Learn more about osteogenesis imperfecta and related disorders.
- Study how some surveillance pathways work in eukaryotic cells.
- Know more about collagen, the most abundant protein in humans.
- Find out why different mutations in one single gene can have such different forms of phenotypical expression and lead to different forms of osteogenesis imperfecta.
- Find a genotype-phenotype correlation in human collagen type I genes.
- Postulate a mechanism to predict the phenotype of possible mutations.

The objectives listed above are more related to the content of the project. However, there are also some other purposes that want to be reached with this assay:

- Learn how the nomenclature for genetic mutations works.
- Acquire a larger knowledge about genetics, inheritance and molecular biology.
- Learn how to browse and use some of the world largest protein and gene databases.

1. 3. QUESTIONS

Before commencing to study the genotype-phenotype correlation, there are some questions I posed to guide me through the research. The questions listed below are the ones I will try to answer at the end of this research project:

Q_1: Which factors contribute to the definition of a mARN as aberrant, and thus lead to its degradation?

Q_2: Why are point mutations, such as glycine substitutions, usually more severe than frameshifts or exon skippings?

Q_3: Why do some glycine substitutions lead to more severe phenotypes than others?

Q_4: Is there a spatial relationship between the location of the mutation and its severity?

Q_5: Is there a severity difference whether the mutation is on the alfa-1 chain or alfa-2 chain, the two collagen monomers?

Q_6: Is there a relationship between the mutation and special findings, like blue sclera or hearing loss, common symptoms seen in osteogenesis imperfecta?

Q_7: Which is the phenotypical outcome of non-glycine substitutions?

Q_8: Is the hydroxylation process of collagen during its synthesis affected by the number of nucleotides involved in the mutation?

1. 4. HYPOTHESIS

Collagen is a protein that has a very elegant and singular structure. In its synthesis, many different enzymes act together in order to stabilize the protein and ensure its functionality. From these two facts, and also from what I know about collagen, I posed some hypothesis which I will try to verify after the research.

H_1: Mutations late downstream probably escape to degradation due to the non-recognition of the sequence as aberrant. This could occur because the recognition mechanisms do not work properly when the mutation is far away from the start codon.

H_2: Glycine substitutions are probably more serious because of:

a) They contribute significantly to chain stabilisation, more than other amino acids.

b) They escape from degradation, because they are not recognized as mutant mARN.

H_3: The severity of glycine substitutions probably depends on the neighbouring amino acids, such as its electrical charge or hydroxylation state.

H_4: Mutations near to the chain end are presumably less severe than the ones at the beginning, since there is a larger region that can normally assemble.

H_5: Mutations on the alfa-2 chain, a subunit of the collagen chain, are probably more severe, since there is only one of them in each collagen molecule.

H_6: There is presumably a spatial relationship between different, type-specific symptoms in osteogenesis imperfecta.

H_7: Substitutions of amino that are not encoded as glycine probably do not have negative effects on the functionality of the protein.

H_8: Mutations that involve a large number of nucleotides could have a negative effect on the posterior hydroxylation and glycosylation, since the enzymes could have recognition problems.

1. 5. METHODOLOGY

This research project requires a clear, rigorous and organized methodology. This fact is important, since the objectives are still far away right now.

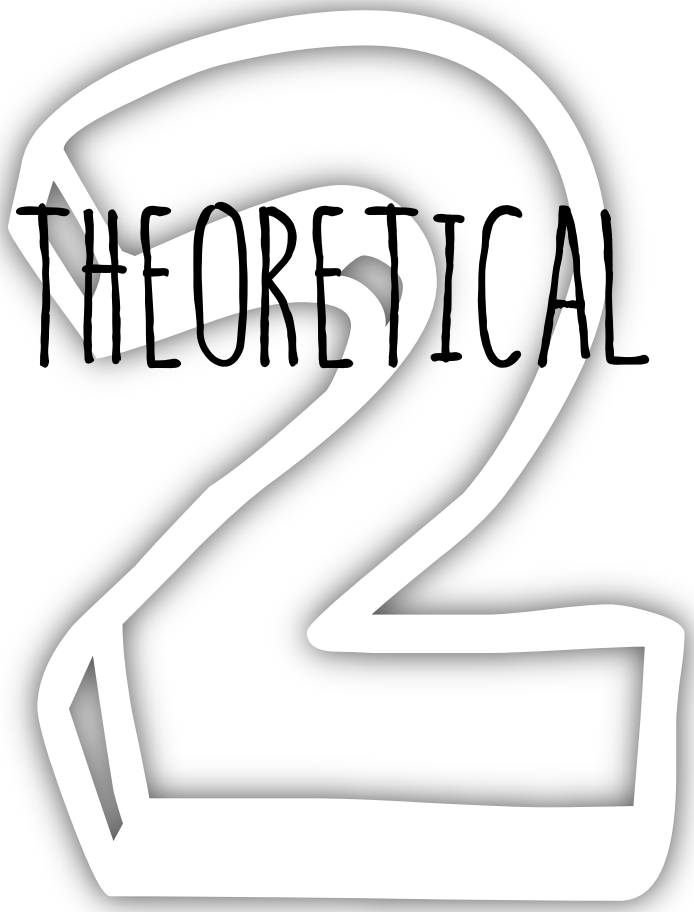
First of all, a wide research and review of the bibliography will be performed. The main information sources will be scientific publications and databases. The theoretical framework will include all the necessary information to conduct and understand the research, without being excessive or include irrelevant facts that deviate from the topic.

The main section of the project, the fieldwork, involves an extensive analysis of the collected information in order to get an answer to the posed question.

The main part of the fieldwork includes a recompilation of mutations in the COL1A1 and COL1A2 genes. However, mutations in these genes responsible for Ehlers-Danlos syndrome, silent/synonymous mutations or polymorphisms will not be included. This data will be taken from the National Centre of Biotechnology Information database. The collected information will be classified according to the type of mutation and according to the resulting phenotype, in order to ease the posterior analysis.

The objective is to find a genotype-phenotype correlation through this data. In order to achieve it, the role of different enzymes, chaperones and intrinsic/extrinsic factors will be studied. After the analysis of the database information, the correlation established will be used to postulate a method for the phenotype prediction.

Since the interpretation of the data will be a long-lasting and challenging process, I have got the help from some scientists and professionals who can support and guide me through the research: Dr Silverman, physician at Hospital Vall d'Hebron (Barcelona), and Dr Malfait, researcher at Centre for Medical Genetics, Ghent University (Belgium).



THEORETICAL FRAMEWORK

2. 1. COLLAGEN[1] STRUCTURE AND STABILITY

2. 1. 1. Generalities

Collagen is the most abundant protein in mammals making up to one third of the total amount of proteins in the body. Collagen is the main structural component of connective tissue and it is found in tendons, bone, cornea, skin, cartilage, ligaments, blood vessels and muscle, among others [54][60][2].

There are a lot of different collagen types and nowadays we know about thirty different types composed of at least forty-five different polypeptide chains. Although every type is different we can classify them into different groups according to its quaternary structure. The most common categories are the fibrillar and the network-forming types. There are also the FACITs (fibril-associated collagens with interrupted triple helices), the MACITs (membrane-associated collagens with interrupted triple helices) and finally the MULTIPLEXINs (multiple triple-helix domains and interruptions) [60].

The fibrillar type is found in almost all the body and it is the major compound of the extracellular matrix (ECM). Collagen is mostly excreted by cells called fibroblasts, found in the stroma of animal connective tissue, and osteoblasts, found in bone tissue [116].

The defining feature of type-I collagen is an elegant structure in which three parallel polypeptides, known as alpha chains, assemble into a larger right-handed triple helix called tropocollagen. These collagen-fibres join together and form larger and thicker associations, the microfibrils, which also get together and form the collagen fibres (**Figure 1)** [54][61][116].

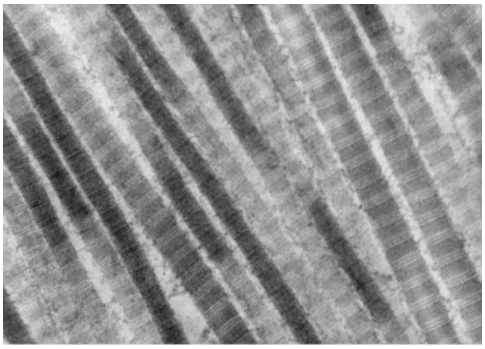


Figure 1 | Transmission electron microscope image of a connective tissue area that shows fibres of collagen type I.

[1] In the whole research project, when collagen is mentioned, it refers especially to collagen type I if not otherwise specified.

[2] Numbers in brackets are the corresponding ones to the references found at the end of the project.

2. 1. 2. Chemical structure and folding shape

The primary structure of collagen type I differs from most of the other proteins synthetized by the human body. The amino acid sequence of collagen is characterized by the repetition of a Gly – Xaa – Yaa structure, in which the Xaa amino acid is mostly proline and the Yaa amino acid usually hydroxyproline (**Figure 2**) [54][60]. The presence of glycine is essential as its small size is needed to fit into the centre of the triple helix [116][117].

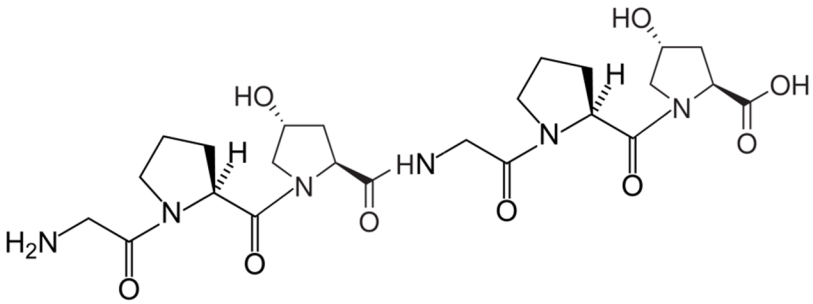


Figure 2 | Molecular representation of the Gly – Xaa – Yaa disposition in collagen type I. In this case the shown amino acids are glycine, proline and hydroxyproline forming the most common sequence in this protein.

Collagen is a protein with a high amount of hydroxylated residues. Prolines and lysines of the polypeptide chains are hydroxylated by different enzymes in a post-translational process, resulting in hydroxyprolines (Hyp, O) and hydroxylysine (Hyl) [2][60][118]. The hydroxylation of proline is very important since it increases dramatically the thermal stability of the triple helices [119]. However, this phenomenon only occurs if the Hyp is on the Yaa position and not in the Xaa position. Non-hydroxylated tropocollagen molecules melt at a temperature of 27°C whereas chains with a large amount of Hyp's on Yaa positions do it at a temperature of approximately 42°C [60] (*for further information see **Table 10**, page 68*).

Lysine hydroxylation occurs mostly in the N-terminus and C-terminus. This hydroxylation is important for the formation of covalent bindings between different tropocollagen molecules when they assemble into larger fibres [116].

The secondary structure of type-I collagen is a left-handed helix in which every twist is formed by the typical Gly – Xaa – Yaa triplets. This structure is known as α-chain and it is very important for the posterior stabilisation of the molecule [2][54][60][116]. Collagen is not considered a protein having a tertiary structure since the secondary structure comprises the whole molecule.

The quaternary structure of collagen is formed by the association of three α-chains, giving rise to a right-handed triple helix stabilized by hydrogen bonds between the different amino acids (**Figure 3**) [116].

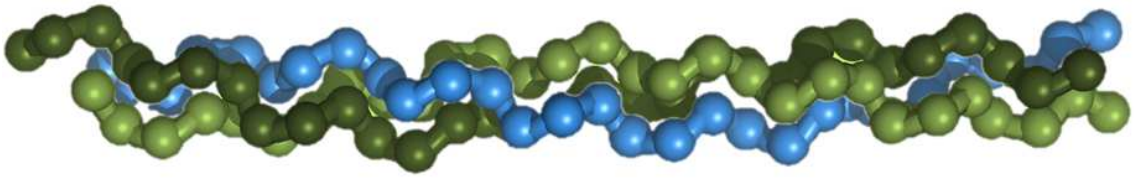


Figure 3 | Sphere model showing the triple-helix conformation of collagen type I. The three α-chains assemble into a major right-handed helix known as tropocollagen. These three α-chains are stabilized by hydrogen bonds between the different amino acid residues. Every tropocollagen molecule is about 300 nm long and 1.5 nm in diameter and is formed by two α-1 chains (green) and one α-2 chain (blue) [60][116].

The interchain hydrogen bonds that stabilize the triple-helix are located between the glycine amino acids and the other residues in the Xaa and Yaa positions. The most frequent bonds are $N - H_{(Gly)} \cdots O = C_{(Xaa)}$, as shown in **figure 4** [60].

Another possible chain-stabilizing bond has been found through crystallographic studies; the $C^{\alpha} - H_{(Gly/Yaa)} \cdots O = C_{(Xaa/Gly)}$ hydrogen bond. It could contribute to the stabilisation of the triple-helix although it is a phenomenon that occurs spontaneously and can be interrupted through high temperatures [60].

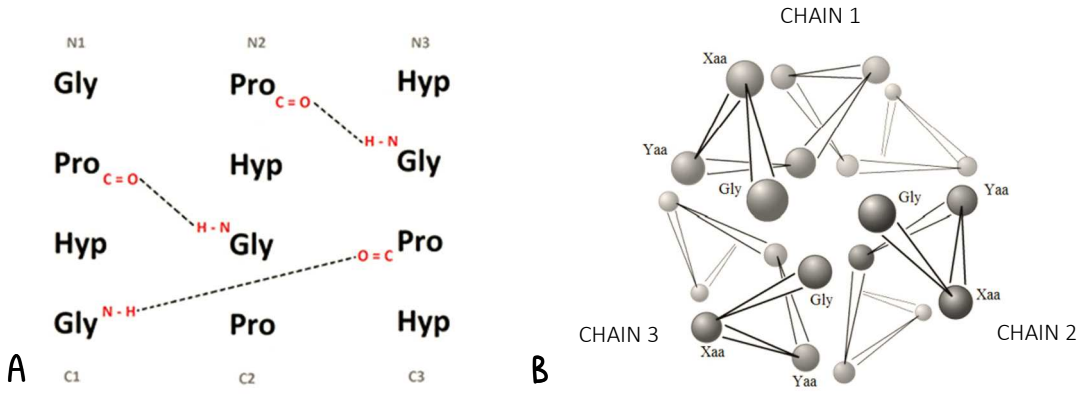


Figure 4 | a) The interchain $N - H_{(Gly)} \cdots O = C_{(Xaa)}$ hydrogen bonds are shown. Glycine is the hydrogen bond donor and the Xaa amino acid, in this case proline, is the hydrogen bond acceptor. **b)** View along the tropocollagen axis showing the disposition of the glycine (Gly), X and Y amino acids. Note the central disposition of the glycine residues.

Within each tropocollagen chain we can distinguish different parts. Most of them disappear during post-translational modifications in the rough endoplasmic reticulum or in the Golgi apparatus (**Figure 5**) [116]. The importance of these structures is discussed in further sections.

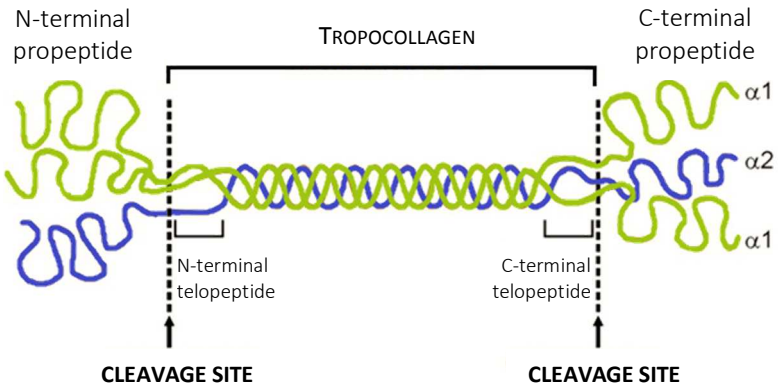


Figure 5 | Different parts of a non-mature collagen chain are represented. At both chain ends there are sequences known as propeptides that are cleaved by procollagen peptidases. There are also terminal telopeptides which contribute to the stabilisation of the molecule linking both ends in order to avoid the opening of the chain [60] [116].

Collagen is a protein which has a higher quaternary structure characterized by the association of many tropocollagen chains that form large fibres. This union forms a bigger structure known as collagen fibrils, and the tropocollagen molecules are disposed following a strict pattern. The tropocollagen subunits assemble together staggered from each other about 67 nm [116]. This distance is known as D-period and it comprises a gap and an overlap section (**Figure 6**) [54].

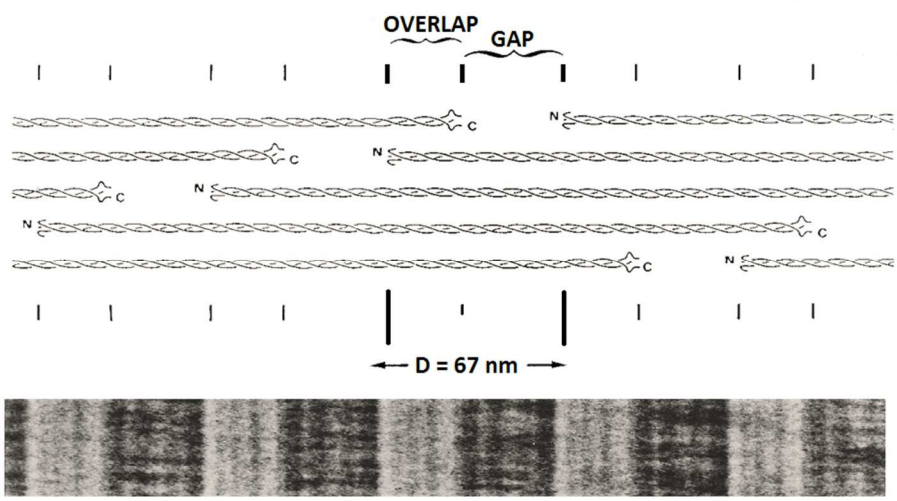


Figure 6 | Illustration representing the disposition of the tropocollagen molecules when they assemble. The gap and overlap regions are symbolized and the D-period is compared to a transmission electron microscope image of a collagen section.

2. 1. 3. Collagen synthesis

Collagen type I is a heterotrimeric protein formed by two α-1 and one α-2 chain. These two types of α-chains differ a little bit from each other in the amino acid sequence, but both share the same secondary structure.

The genes responsible for the collagen synthesis are the COL1A1 and COL1A2. The first one encodes α-1 chain while the second one encodes α-2 chain. First of all, these genes are turned on and the transcription begins. The main steps of collagen synthesis are represented in **Figure 7** [54][55][60][77][78][116].

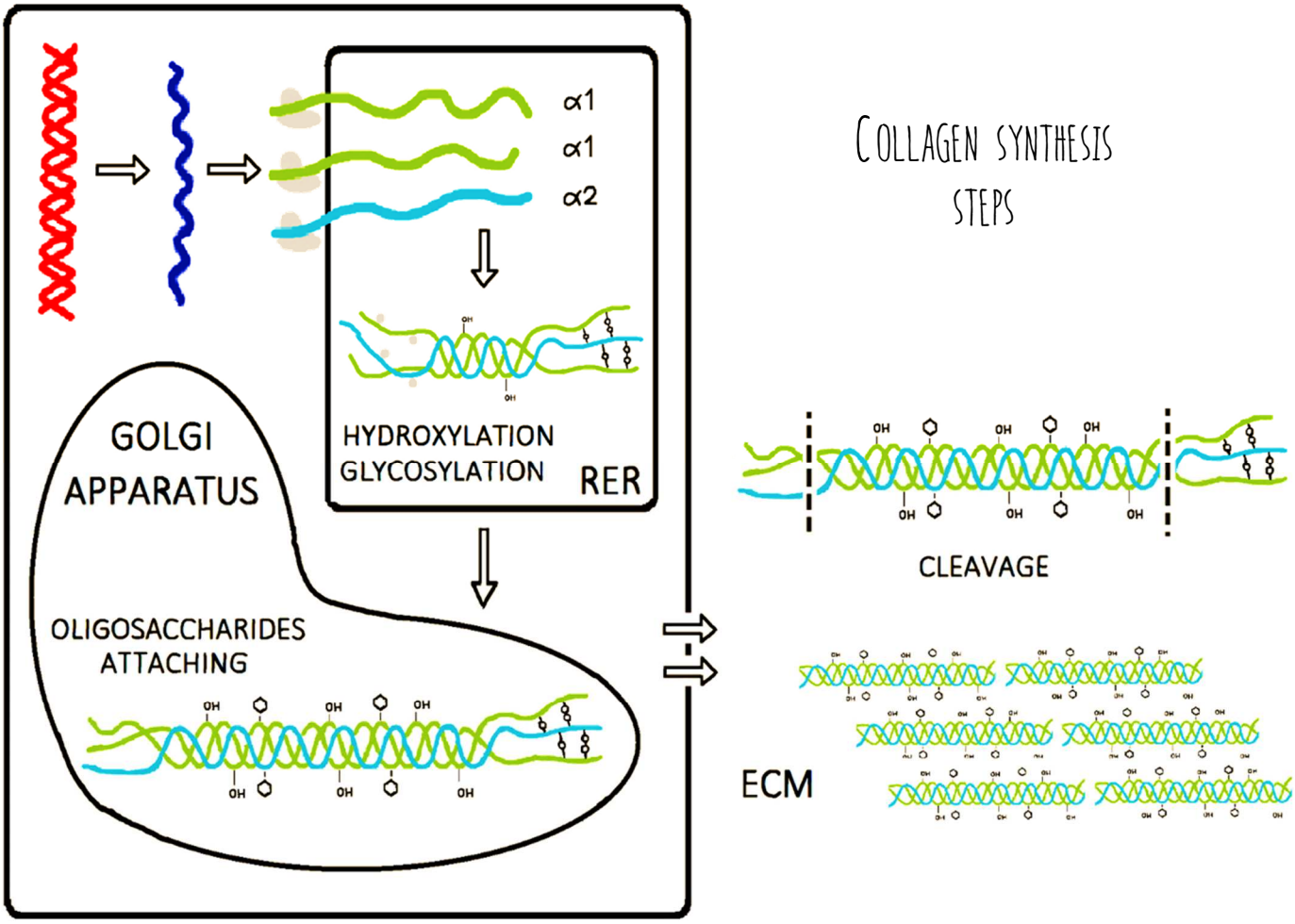


Figure 7 | The different steps of collagen synthesis are represented. The DNA is transcribed into mRNA. The alpha chains are synthetized and assembled into triple helix in the ER lumen. Hydroxylation and glycosylation then occur. The collagen molecules are transported into the Golgi apparatus where oligosaccharides are attached. Lastly, the collagen molecules are secreted, cleaved and incorporated into ECM.

2. 1. 4. Collagen related disorders

In the following table (**Table 1**) there are some collagen-related disorders caused by mutations in different genes [60]. Most of them are characterized by tissue fragility and the main symptoms include joint and vascular alterations, cartilage defects, bone deformities and skin brittleness. Although there are many other genes encoding collagen chains, they are not well studied and the related disorders are not defined yet.

Collagen type	Encoding genes	Pathology
I	COL1A1, COL1A2	Ehlers-Danlos syndrome type VII [artrochalasia][3] [85], Infantile cortical hyperostosis[4] [116], Osteogenesis imperfecta
II	COL2A1	Various spondyloepimetaphyseal dysplasias, Stickler syndrome
III	COL3A1	Ehlers-Danlos syndrome type IV [vascular]
IV	COL4A1, COL4A2, COL4A3, COL4A4, COL4A5, COL4A6	Alport syndrome
V	COL5A1, COL5A2, COL5A3	Ehlers-Danlos syndrome types I & II [classical]
VI	COL6A1, COL6A2, COL6A3	Bethlem myopathy, Ullrich congenital muscular dystrophy
VII	COL7A1	Dystrophic epidermolysis bullosa
VIII	COL8A1, COL8A2	Posterior polymorphous corneal dystrophy 2
IX	COL9A1, COL9A2, COL9A3	Multiple epiphyseal dysplasia types II, III & VI
X	COL10A1	Schmid metaphyseal chondrodysplasia

Table 1 | Different collagen-related disorders are listed, indicating in each case the affected collagen-type and the responsible encoding genes [60].

[3] Ehlers-Danlos syndrome types VIIA and VIIB are characterized by congenital bilateral hip luxation, mild osteoporosis, muscle hypotonia, loose joints and minimal skin involvement.

[4] Infantile cortical hyperostosis (aka Caffey disease) is an inflammatory, self-limited disorder in children that causes bone changes and soft tissue swelling. The disease resolves on its own after approximately six months. The genetic penetrance is variable.

2. 2. OSTEOGENESIS IMPERFECTA

2. 2. 1. Description

Osteogenesis imperfecta, also known as Lobstein syndrome or brittle bone disease, is a genetic disorder characterized by bones that prone to fracture. It is caused by different mutations and the result is an abnormal synthesis of collagen type I in connective tissue [81][108].

It is classified as a rare disease and its incidence is about 1 in 20,000 live births [122]. There are sixteen different types that can be distinguished although there are four main types [109]. In the following table there is the most used classification proposed by Sillence in 1979:

Table 2 | The four main (classical) types of osteogenesis imperfecta are shown in the table, the classification proposed by Sillence [122]. These ones are caused by mutations in collagen type I genes.

Type	Severity	Genes	Inheritance[6]
I	mild	COL1A1, COL1A2	autosomal dominant
II	lethal	COL1A1, COL1A2	autosomal dominant
III	severe	COL1A1, COL1A2	autosomal dominant
IV	moderate	COL1A1, COL1A2	autosomal dominant

Table 3 | The six major new types of osteogenesis imperfecta are shown and the responsible encoding genes.

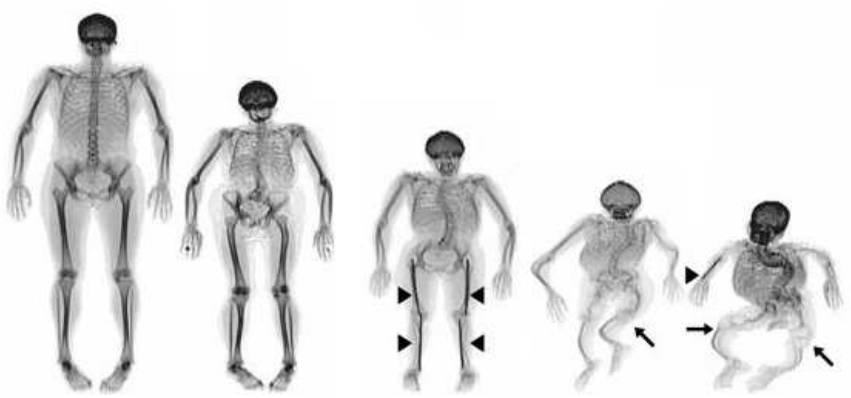
Type	Severity	Genes	Inheritance[6]
V	moderate	IFITM5	autosomal dominant
VI	moderate	SERPINF1	autosomal recessive
VII	moderate	CRTAP	autosomal recessive
VIII	severe/lethal	LEPRE1	autosomal recessive
IX	severe/lethal	PPIB	autosomal recessive
X	severe	SERPINH1	autosomal recessive

The main types of OI are the types I, II, III and IV. These are the ones related to the *COL1* genes, responsible for the encoding of the type-I collagen polypeptide chains. The other types (V-X) are associated to other genes and are not well studied yet [109].

[6] Although the inheritance of all types of osteogenesis imperfecta is autosomal-linked and dominant, in types II and III almost all the mutations are new whereas in types I and IV about 40% are inherited from one progenitor [122].

Osteogenesis imperfecta is characterized by fragile bones which break easily, skeletal deformities (especially in long bones and scoliosis), hearing loss, blue sclera, dentinogenesis imperfecta and joint hypermobility [81][108][122].

Even if most of the listed symptoms are present in all the types, it varies from person to person and there are also type-specific findings and signs (**Figures 8 - 9**).



TYPE I	**TYPE I**	**TYPE IV**	**TYPE III**	**TYPE III**
Female	Female	Female	Female	Female
38 yr	63 yr	35 yr	27 yr	40 yr
171 cm	137 cm	124 cm	94 cm	84 cm

Figure 8 | Full-body X-rays from different patients with OI. Note the difference of height and bone deformities between the affected persons, even between the persons affected by the same OI subtype.

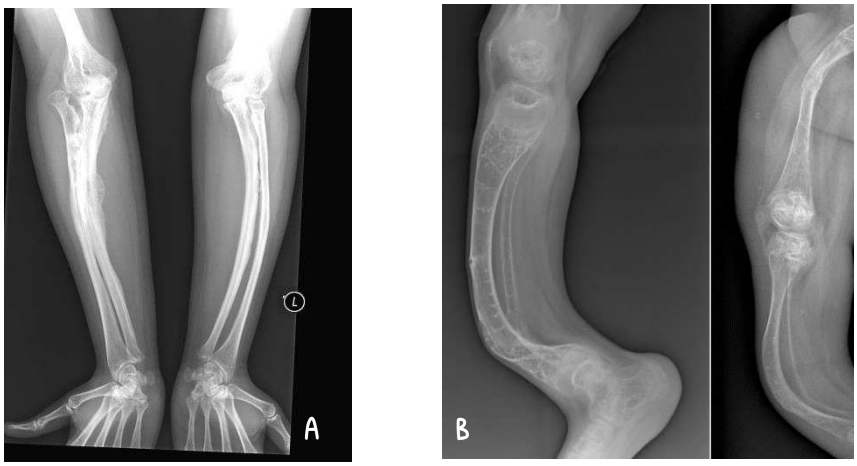


Figure 9 | a) X-ray showing forearm and wrists of an OI patient. Note the bowing of both radius and ulna and the presence of callus forming around the healing fractures. **b)** Lower extremities X-rays showing severe antecurvatum of femur, tibia and fibula. There can be also observed the abnormal translucency of the bones, seemingly as if they were made up of glass. This phenomenon occurs as a consequence of the low bone density (osteoporosis), the poor bone mineralisation and high resorption rates seen in osteogenesis imperfecta [122]. [21]

2. 2. 2. Symptoms

Osteogenesis imperfecta type I [7]

Osteogenesis imperfecta type I is considered the mildest type although the severity varies significantly between the affected persons [89][109][122].

- Fractures: most of the fractures in type I OI occur before puberty and in later years with a significant diminution between them.
  The average fracture rate is about 10 – 70 more or less.
- Bone deformity: in this type the skeletal deformities and/or scoliosis are usually mild or totally absent.
- Height: can be below average for age but adults may be similar compared to their unaffected family members.
  The average height for OI type I patients is usually below the 50th percentile but above the 3rd percentile.
- Sclera: the sclera is usually grey-blue coloured.
- Injuries: affected people can be prone to skin bruises and joint dislocations.
- Dentinogenesis imperfecta: is usually present.

Osteogenesis imperfecta type II [8]

Osteogenesis imperfecta type II is considered the most severe type since it is lethal during perinatal period. Fractures occur during foetal development and often cause severe skeletal malformations. Thoracic deformities cause an underdevelopment of the lungs and intracranial haemorrhages are in the most cases fatal [81][90][122].

All the cases of type-II OI are caused by spontaneous mutations and it can usually been diagnosed prenatally through and abdominal echography [122]. Ultrasound examination from 14 weeks gestational age can diagnose this type of OI, and the most common findings are short and fractured long bones, underossification of the cranium with an abnormal clarity of intracerebral structures, and barrel-shaped thorax deformation [109].

[7] OMIM entry #166200
[8] OMIM entry #166210

Osteogenesis imperfecta type III [9]

Osteogenesis imperfecta type III is considered the most severe, non-lethal type of OI. The main clinical features are fractures with minimal trauma or spontaneous without a clear cause. Due to these fractures, upper and lower limbs are usually severely deformed [91][109][122].

- Fractures: bones break very easily, usually healing quite rapidly and causing severe deformations. Microfractures while walking or doing other quotidian activities are common. Fractures in the womb are also often seen, making it possible for prenatal diagnosis. Patients can have more than one-hundred fractures in a lifetime.
- Bone deformity: in this type the skeletal deformities and scoliosis are severe with limbs that are curved and impossible walking. Scoliosis is often rapidly-progressing, and spine and long bones usually need rodding in order to stabilize and reduce the deformities (**Figure 10a**).
- Height: is notably below the average for age. The average height for OI type III patients is below the 3rd percentile; many of them do not overcome 120 cm.
- Sclera: the sclera, as seen in type-I OI, is usually grey-blue coloured (**Figure 10b**). Nonetheless, in many cases it can turn normally-coloured (white) after puberty.
- Injuries: affected people usually have skin brittleness and can have joint dislocations associated to the fractures.
- Dentinogenesis imperfecta: is usually present.

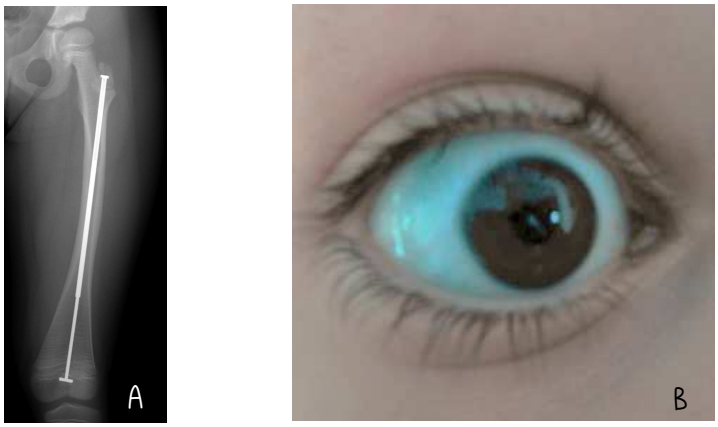


Figure 10 | a) X-ray showing a femur which has been stabilized through the insertion of a long intramedullary rod. **b)** Photograph showing an eye with blue-coloured sclera. This phenomenon occurs due to the thinness of the sclera which lets see the underlying choroidal veins. [21]

[9] OMIM entry #259420

Osteogenesis imperfecta type IV [10]

Osteogenesis imperfecta type IV is considered the moderate type of OI. The number of fractures and deformities are not as severe as seen in type III but more significant as in type I. This type is the most heterogeneous since it includes all the patients that do not fit in the other types. Nonetheless, persons with type-IV OI usually do not have blue sclera, unlike in types I and III [92][109][122].

- Fractures: bones break easily but spontaneous fractures without trauma are not often. Patients with this type of OI hardly ever have more than one-hundred fractures.
- Bone deformity: bone and spine deformities are usually seen but they are not as severe as in type III. The patients are usually able to walk and do quotidian activities without serious restrictions and improve puberty.

  Height: is below the average for age. They are usually below the 3rd percentile but are not much smaller as unaffected people.
- Sclera: the sclera, in contrast to the other types, is generally normal-coloured.
- Injuries: affected people usually have a soft skin that bruises easily and joint hypermobility.
- Dentinogenesis imperfecta: rarely seen.

Non-classical OI types

Osteogenesis imperfecta types V, VI, VII, VIII, IX and X are much less frequent than the classical types. They are caused by other genes and do not affect directly the collagen structure [109].

- OI type V: it has the same clinical expression as type IV but it is characterized by the 'V triad': radio-opaque band adjacent to growth plates, hyperplastic calluses, and calcification of the radio-ulnar interosseous membrane [93][122].
- OI type VI: it has the same clinical features as type IV but it is distinguished histologically by 'fish-scale' bone appearance [94][122].
- OI types VII, VIII, IX and X: these four types are very rare and are found in only a few families, usually in consanguineous family members that have descendants [95][98][122].

[10] OMIM entry #166220

Radiological findings

Osteogenesis imperfecta causes different alterations in bones (**Figures 11a** and **11b**). The most characteristic radiological sign is the abnormal translucency of the bones due to a low density and high resorption (**Figure 12a**). The bone deformities are usually seen in long bones and spine, and are the result of the repeating fractures that heal causing the bowing of bones [122]. After a break, people affected by osteogenesis imperfecta usually form a hypertrophic callus around the fracture focus. A special feature of osteogenesis imperfecta type V is the calcification of the radio-ulnar interosseous membrane, as seen in **Figure 12b** [93][122].

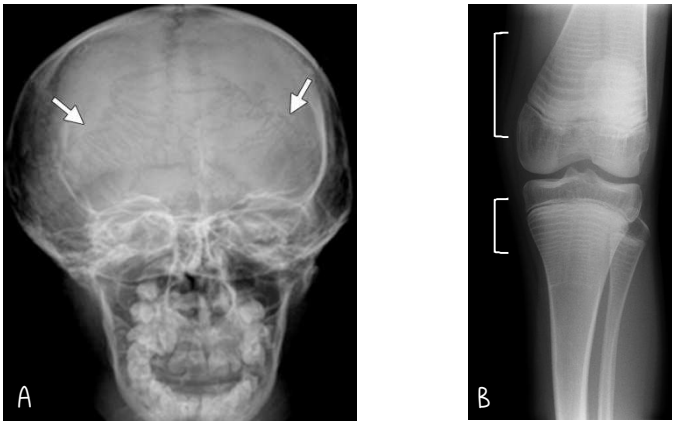


Figure 11 | a) Cranium X-ray from a patient with OI. The arrows point to the wormian bones often seen in osteogenesis imperfecta. Wormian bones (aka intrasutural bones) are extra bone pieces that form within a suture in the skull. **b)** Knee X-ray showing the zebra sign. Note the horizontal lines located at the distal and proximal segment of femur and tibia respectively, near the growth plate. This phenomenon occurs during bisphosphonate therapy applied in more severe types of OI, and they correspond to areas with a major bone density [112]. [21]

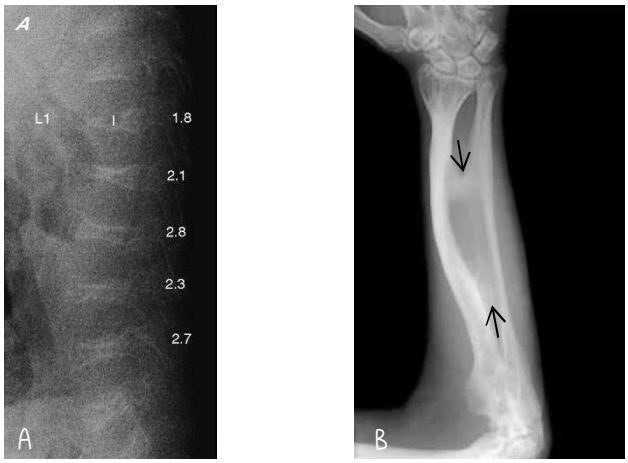


Figure 12 | a) Lateral spine X-ray showing the characteristic low bone density seen in osteogenesis imperfecta. Note the radiotranslucency of the vertebrae. **b)** Forearm X-ray showing the radio-ulnar interosseous membrane calcification, distinctive feature for osteogenesis imperfecta type V [93]. [21]

2. 2. 3. Treatment

There is no cure for osteogenesis imperfecta. The main goals of treatment in these patients are to decrease the number of fractures, correct deformities and reduce bone pain. There are three main treatment methods: bisphosphonates, physiotherapy and surgical treatment [122].

Bisphosphonates

Bisphosphonates are a group of drugs used to increase bone mass. Formerly, they were used to treat osteoporosis but in 1998 a clinical trial demonstrated the effectiveness of bisphosphonates in patients with osteogenesis imperfecta. They used pamidronate (pamidronic acid) and the number of fractures and pain was significantly reduced. It is usually administrated intravenous since its effectiveness is higher than if done orally [122].

Physiotherapy

Physical therapy is very important since it helps strengthen the muscles and maintain normal motion of joints. Long casting-periods weaken the bones and reduce joint mobility, two factors that must be prevented in patients with OI. When casting is required, immobilisation of large body parts is avoided and the cast is usually removed earlier or replaced in order to prevent loss of bone mass. Crutches, orthoses and other physical aids are also usually required [122].

Intramedullary rodding

Surgical treatment is commonly performed in severe cases of osteogenesis imperfecta in order to correct deformities, improve bone strength and avoid fractures. During the late 1940s, Harold A. Sofield, MD, at Shriners Hospitals for Children (Chicago), developed a method in which large stainless steel rods are inserted into large bones to stabilise them. Nowadays, this method is used almost all over the world, and it has been improved so that modern intramedullary rods can be introduced into growing bones since they can enlarge themselves while the bone lengthens (telescopic rods) [122].

Spinal fusion is also usually required to correct spine deformities such as scoliosis or kyphosis.

2.3. INTRODUCTION INTO THE GENETIC BACKGROUND OF OSTEOGENESIS IMPERFECTA

The mutations leading to osteogenesis imperfecta are located on the COL1A1 and COL1A2 genes. Several mutations on both genes have been identified and most of them lead to osteogenesis imperfecta phenotypes. Nonetheless, mutations in the N-terminus region causes a disease called Ehlers-Danlos syndrome type artrochalasia, [116] and therefore these ones will not be studied.

2. 3. 1. Collagen type I, alpha 1 & 2 genes

Table 4 | The domains of type I collagen are listed including the number of amino acids comprising each of them. The final mature tropocollagen molecule only includes the triple helical domain [56].

Domain	COL1A1	COL1A2
N-terminus		
Signal peptide	22	22
N-propeptide	139	57
Collagen domain		
N-telopeptide	17	11
Triple helix	1014	1014
C-telopeptide	26	15
C-terminus		
C-propeptide	246	247

The COL1A1[11] gene is located on the long arm of chromosome 17, at position 21.33 (17q21.33) [77]. The COL1A2[12] gene is located on the long arm of chromosome 7, at position 22.1 (7q22.1) [78].

Both genes have got a triple helical domain which comprises 1014 amino acids so that there are 338 Gly–Xaa–Yaa repetitions. This triple helical domain is divided in 43 exons (from a total of 52). Each exon is formed by 45, 54, 99 or 108 base pairs [54].

[11] OMIM entry *120150

[12] OMIM entry *120160

2. 3. 2. Expression pathways of collagen mutations

In osteogenesis imperfecta there are two different outcomes in collagen synthesis when mutations are present: haploinsufficieny and dominant negative effect [25].

In normal conditions, when there is a mutant allele, the other one can replace its function and there is no pathological outcome. In most cases, half of protein product is enough to permit the cell work normally [25].This does not occur when a mutation in collagen genes is present.

Haploinsufficiency

Haploinsufficiency is a situation in which a cell cannot work properly when only half the amount of protein is produced. The mutation leading to this phenomenon is called null allele (aka amorphic mutation), which means that these allele does not produce a functional protein.

The outcome of this loss-of-function mutation, applied to collagen genes, means that only fifty percent of the normal amount of α-1 or α-2 chains are produced [68]. In this case half of tropocollagen molecules are normal while the other half is composed of loose polypeptide chains (**Figure 13**).

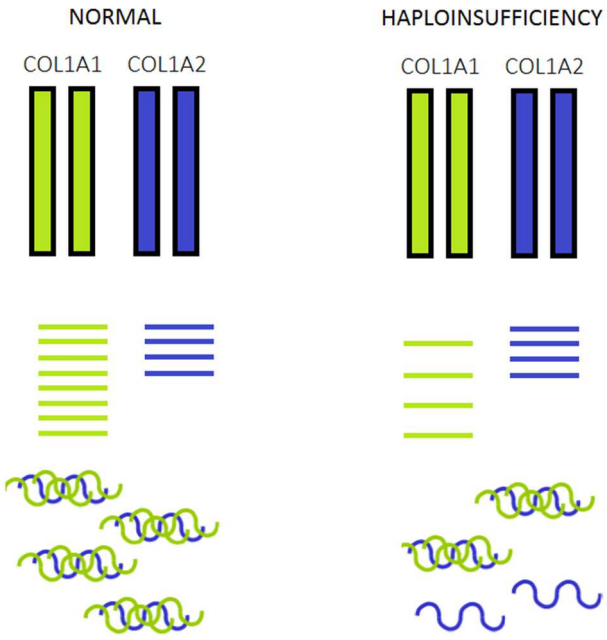


Figure 13 | One possible haploinsufficiency situation is represented. In normal conditions a tropocollagen molecule is formed by two α-1 (green) and one α-2 (blue) chain. This means there is a natural equilibrium in which the COL1A1 alleles produce twice as much chains as COL1A2 alleles do. When a mutation produces a null allele only half the amount of protein is produced and the equilibrium is interrupted. The result is a lack of polypeptide chains, and it can be that either α-1 or α-2 chains cannot assemble into normal tropocollagen molecule.

Dominant negative effect

Dominant negative effect, also known as antimorphic mutations, is a phenomenon in which a mutant protein antagonises the function of the normal protein.

Translated into the collagen molecule, this means that when a mutant chain is introduced the whole protein destabilizes losing its functionality (**Figure 14**).

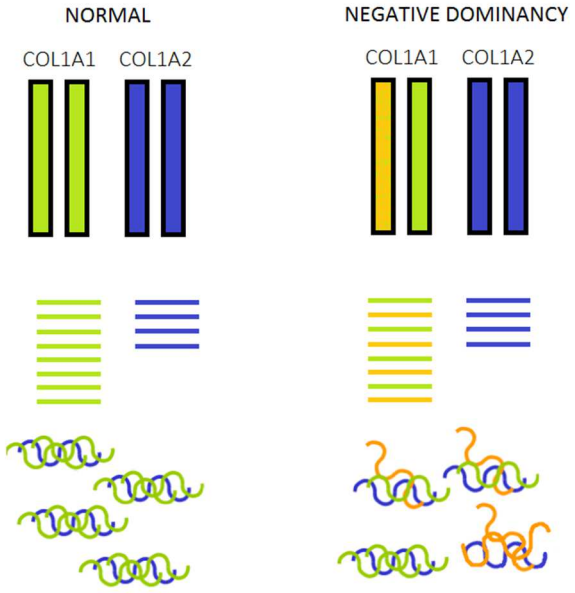


Figure 14 | One possible dominant negative situation is represented. In normal conditions a tropocollagen molecule is formed by two α-1 (green) and one α-2 (blue) chain. However, when there is an allele that produces a mutant chain (orange), this chain is introduced into mostly every tropocollagen molecule. The presence of one mutant chain weakens the whole molecule making that almost 100 % of the tropocollagen molecules are not normal and thus not functional.

The main difference between both situations is that in haploinsufficiency the mutation leads to the absence of protein product from one allele, whereas in dominant negative effect there is a protein product but it is not functional. The phenotypical outcome of haploinsufficiency, a quantitative defect of collagen, is much less worse than negative dominancy, a qualitative defect. The first one causes OI type I [89] while the second one causes OI types II, III and IV [122].

This research project pretends to study which mutations lead to haploinsufficieny and which to negative dominancy and therefore to the different types of osteogenesis imperfecta.

FIELDWORK

3. 1. DATA RECRIUTMENT

To conduct this study data was obtained from the *LOVD Osteogenesis Imperfecta Variant Database*[13]. This database contains all the known DNA variants causing osteogenesis imperfecta with its respective phenotypes and complementary information. The mutations included in this study are glycine substitutions, splice sites, frameshifts and nonsense mutations (PTC). All the mutations comply with the nomenclature guidelines set by the Human Genome Variation Society (HGVS) [17].

The number of total mutations is listed in the following table.

Table 5 | The table below shows the number of mutations that are comprised in the study.

	Glycine substitutions	Splice site alterations	PTC	Frameshifts	Total
COL1A1	306	139	115	178	738
COL1A2	300	61	4	6	371
Total	606	200	119	184	1109

The selection process of mutations was performed under strict criteria. The respective inclusion/exclusion conditions are the following ones:

Inclusion criteria

- Patients with a confirmed diagnosis of OI through genetic testing.
- OI cases caused by mutations in the encoding genes for collagen type I.
- Disease subtype must be reported without ambiguity.

Exclusion criteria

- Larger mutations involving more genes (micro-/chromosomal deletion syndromes)
- Patients having more than one disease-causing mutation.
- Disease causing mutations in other genes than COL1A1 and COL1A2.
- Patients carrying mutations for other diseases.
- Lack of reliable references.

[13] **LOVD** (**L**eiden **O**pen **V**ariation **D**atabase) is an open-source database developed in the Leiden University Medical Center (Netherlands). Sequence variants of COL1A1 and COL1A2 genes used in this research project are available at < http://www.le.ac.uk/ge/collagen/ >.

In order to ease the study there are some applications and software that have been used. These tools can simulate experimental procedures and since there was not the possibility to conduct them, they have been used to replace the laboratory work.

PyMol

PyMol is a molecular visualization software used in protein research, computational chemistry, and biology. Protein Data Bank (PDB) is a molecular structure database and there are 3D representations available for different macromolecular species. The corresponding files are downloaded and opened with, for example, PyMol.

Collagen Stability Calculator

The Collagen Stability Calculator is an application used to determine the melting temperature of short, collagen-like peptides or generate a relative thermal stability of collagen chains. The software is based on data obtained from host-guest peptides and integrated to produce an algorithm to predict thermal stability [49].

Reading-frame checker

The reading-frame checker is an application included in the LOVD database used to predict the consequences of exon duplications and deletions. The information is based on direct translation of the mRNA. It cannot be used in case of splice site alterations [17].

Coding DNA Reference Sequence

The cDNA Reference Sequence from LOVD database is a file with the exon and intron sequence, also indicating the possible premature stop codons in the +1 and +2 reading frames. In order to ease the mutation description every tenth nucleotide and amino acid are highlighted [17].

Human Splicing Finder website [14]

The Human Splicing website is an online application which can be used to analyse splice sites, mutations, sequences, and multiple mRNA transcripts. The sequence can be chosen through different identification names and numbers from different databases (RefSeq Peptide ID, Gene Name, Ensembl Gene ID, etc).

[14] Human Splicing Finder website: < http://www.umd.be/HSF3/ >

3. 2. PHENOTYPE MODULATION

Disease-causing mutations in collagen-encoding genes have a complete penetrance, meaning that everyone who carries these mutations will develop a pathological phenotype.

The study of collagen mutations is simplified because epigenetic regulation and environmental factors are not significant, therefore not influencing the clinical outcome. Nonetheless, the phenotype variability in osteogenesis imperfecta varies from almost asymptomatic to perinatal lethal, and this makes difficult to find a clear phenotype-genotype correlation only through a few criteria [10][40][122].

At first, OI types caused by other genes than the collagen-encoding ones were ignored since in this research project only the COL1A1/2 genes are studied. However, during the investigation, it was noted that they could have a key role in modulating the severity of the resulting phenotype.
Most of the non-collagen related OI-types are caused by altered molecules, chaperones and enzymes that interact with collagen but do not work properly [33]. This makes that the resulting collagen molecule is not formed correctly without having a structural alteration. This fact suggests that it could also be on the contrary; certain mutations in the collagen molecule could interfere with the activity of these proteins, thus increasing the severity of the disease.

Another factor that contributes to the phenotype severity is cellular surveillance systems. These mechanisms function as control pathways that ensure correct cell function and protein synthesis. Some mutations and misfolded proteins can activate these surveillance systems having various consequences ranging from protein degradation to apoptosis.

The main highlights that have been studied are:

- Enzymes, molecules and chaperones that interact with collagen.
- Cellular surveillance pathways
- Mutations throughout the triple helix and the C-terminal

These mechanisms and factors are exposed and discussed in the following sections in order to provide an insight into the modulation of the OI phenotype.

3. 3. FRAMESHIFTS AND PREMATURE STOP CODONS

Frameshifts and premature stop codons (PTCs) are the less common mutations. At first, these two mutations types required to be studied separately but then it was noted that their effect is the same: truncating the protein. It is supposed that nonsense mutations and out-of-frame indels should have major effects on the body since the molecule function is abolished by these alterations.

In collagen, in almost all the cases this does not happen. Unlike other proteins collagen is much more affected by structural alterations than by the absence of protein product [56]. Therefore, collagen molecule takes advantage of a cellular surveillance system known as nonsense-mediated mRNA decay (NMD).

The pre-mRNA is a direct transcript from the DNA sequence and therefore contains introns. In order to get rid of the no-coding regions a maturation process occurs known as splicing. During splicing, introns are eliminated and exons are joined forming exon-junction complexes (EJC) and mature mRNA (**Figure 15**) [25][45].

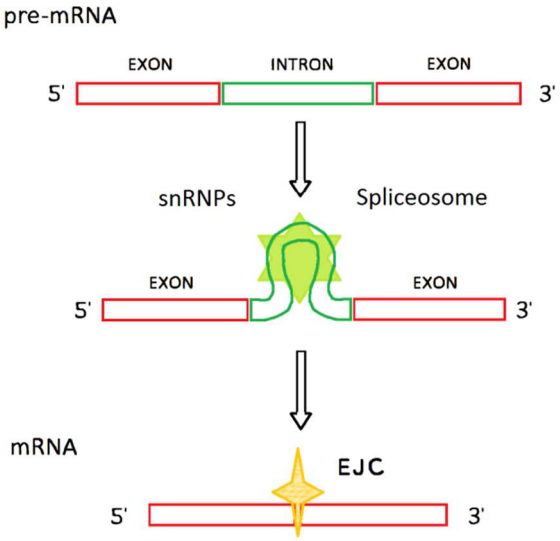


Figure 15 | Representative image of transcription and mRNA maturation. Introns are cut off by the spliceosome and small nuclear ribonucleic particles (snRNPs). When exons are linked the EJCs attach to the splice junction and are transported to the cytoplasm together with the mRNA.

After the splicing process the mRNA is transported to the cytoplasm where translation initiates. Ribosomes then add amino acids to the growing polypeptide chain. If during the translation the ribosome arrives to a stop codon but there are still EJCs present, the protein is triggered for NMD. When NMD is activated the aberrant mRNA transcript is rapidly degraded, as well as the protein (**Figure 16**) [3][25][45].

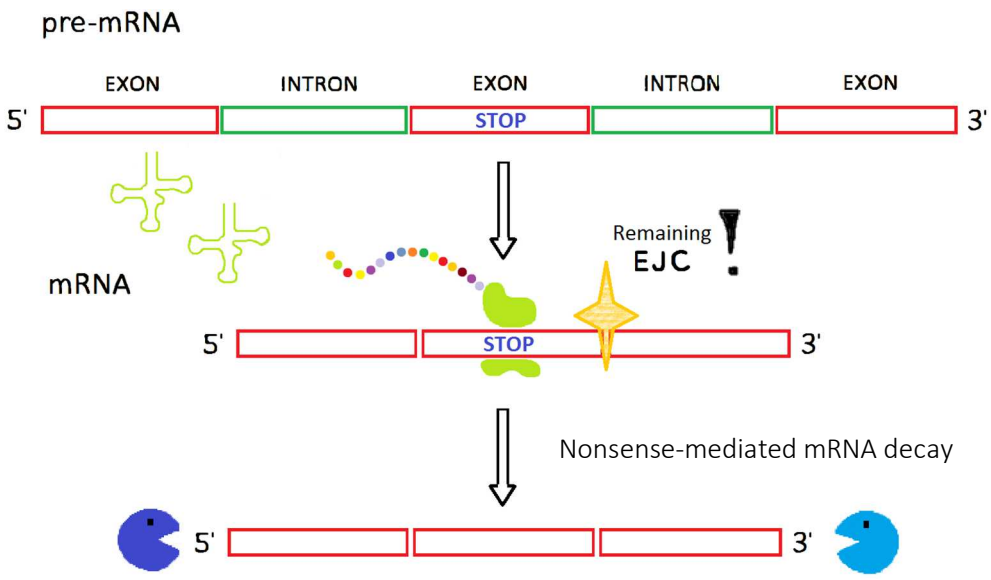


Figure 16 | Representative image of nonsense-mediated mRNA decay. The presence of a remaining EJC activates the molecules that participate in NMD and the aberrant mRNA is destroyed by different enzymes through deadenylation and decapping. The protein is degraded by proteasomes in the cytoplasm through proteolysis [3][45].

The final outcome of NMD is the degradation of both mRNA and the protein [3]. The result is a quantitative deficiency of the protein since all the products of the mutant allele are destroyed.

Translated into osteogenesis imperfecta phenotype, this means that only half amount of alpha chains are produced since the aberrant ones are degraded. This fact is supported when we look at the mutations chart. Almost all the resulting phenotypes are OI type I, consistent with a quantitative defect.

Nonetheless, beyond residue number 1190 severe and lethal phenotypes appear. This fact indicates that the truncated chain is included into the tropocollagen molecule and thus fails to trigger NMD. Studies have revealed that premature stop codons in the last 55 nucleotides of the second last exon and the whole last exon do not activate the surveillance pathway [3]. This could explain the severe phenotypes resulting in nonsense mutations near 3' end of mRNA.

In comparison to collagen, dystrophin is a structural protein [80] sensible to quantitative defect. When a truncated protein is produced it is rapidly degraded and this situation results in a very severe disease known as Duchenne muscular dystrophy. In contrast, punctual mutations such as in-frame indels and substitutions do not trigger NMD and thus are not degraded. This latter results in a milder disease called Becker muscular dystrophy [79].
With this contrast we can see that even in proteins with comparable and similar functions, alterations can be expressed in many different ways.

3. 4. GLYCINE SUBSTITUTIONS

The most common OI-causing mutations are glycine substitutions [40]. This type of mutation changes the obligate glycine residue in the Gly – Xaa – Yaa triplet for a larger amino acid, thus disrupting the folding of the triple helix [9]. When a nucleotide substitution occurs in the glycine-coding triplet, eight possible residues result: aspartic acid, glutamic acid, tryptophan, cysteine, serine, alanine, valine and arginine [10][23].

Proα1 and proα2 chains

An overall relationship can be found observing the number of lethal mutations in each chain. As seen in **Table 6** there is a significant difference between the lethality rates of mutations in each chain. This means that the probability of a substitution being lethal is significantly higher in alpha-1 chain than in alpha-2.

Table 6 | The table below shows the number of lethal mutations in each alpha chain and the respective percentages. From a total of 176 lethal mutations, 114 are in COL1A1 and 62 in COL1A2.

	COL1A1	COL1A2
Number of mutations	306	300
Number of lethal mutations	114	62
Percentage lethal	37.25 %	20.67 %

In both alpha chains there are the same numbers of glycines, so this fact cannot contribute to this difference. A possible cause for this occurrence is that a mutant alpha-1 chain can be introduced twice in one tropocollagen molecule and having a double destabilising effect (**Figure 17**).

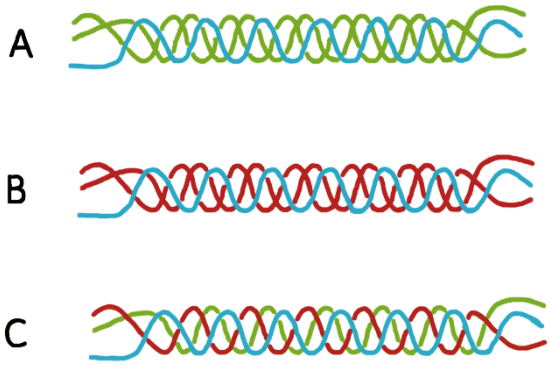


Figure 17 | This image represents different ways the alpha-1 chain can be introduced into collagen molecule. Colour green is alpha-1, blue is alpha-2 chain and red is mutant alpha-1. **A)** Normal tropocollagen heterotrimer. **B)** Collagen trimer containing two mutant alpha-1 chains. **C)** Collagen trimer containing both normal and mutant alpha-1 chains.

The data listed above is not consistent with that reported by Marini et al. (2009)[40]. In their review they stated a lower percentage, of 35,6 % and 18,9 % respectively, in both alpha chains. Although the average difference between the chains is in both cases about 2 % there could not be found a clear cause for this discrepancy. One hypothesis is the divergence in the amount of mutations studied in the essays since this could influence the results. Another possibility is the growing genetic tests done on children that die in perinatal period. They are performed in order to assess the possibility of recurrence of a genetic disease[15], and the higher amount of reported lethal mutations could be the cause of this discrepancy (Silverman, personal communication).

C- and N-terminal regions

The alpha chains constituting collagen have got two ends: N-terminus and C-terminus, which make the molecule very soluble and easy to transport through the cell [122]. There is an evident difference between mutations at these two sides and also along the chain.

Folding of collagen begins with the nucleation of the C-terminus. The triple helix propagates towards N-terminus and diverse crosslinks and bonds form in order to stabilise this structure (**Figure 18**) [122]. Mutations at C-terminus do not display the same characteristics as mutations in the triple helix.

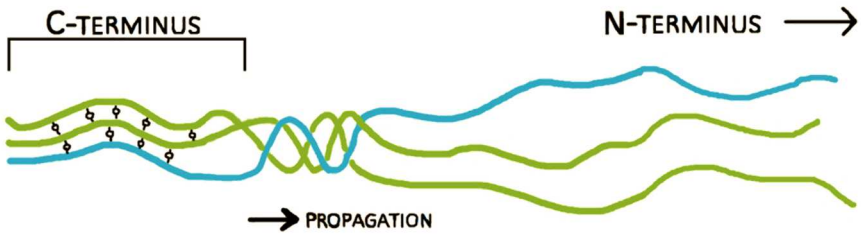


Figure 18 | Representation of the C- and N-terminal regions and the propagation direction of triple helix folding. The black dots at C-terminus denote the disulphide bonds.

The C-terminus regions of the three alpha chains align in order to begin helix folding [32]. Disulphide bonds form to stabilize the molecule and nucleation occurs towards N-terminal [54].

[15] In OI, although all the lethal cases are caused by spontaneous mutations, there is a low number of families in which recurrence occurs. This phenomenon is caused because of germline (gonadal) and/or somatic mosaicism, meaning that one parent's sperms or eggs carry the mutation without having symptoms or a mild form of them [108].

When mutations are present at the C-terminus end, initial assembly can be delayed or even avoided and no collagen molecules form.
In these types of mutations, high overmodification of the collagen molecule occurs due to a prolonged exposure to modifying enzymes. The assembly of collagen molecules is extremely hindered, and retention and accumulation of unfolded proteins follows in the lumen of the ER. Nonetheless, once the folding has begun (if it begins), a stable collagen molecule can be formed be introduced into extracellular matrix.

Surprisingly, mutations at C-terminus have shown to be either lethal or mild. The ones that affect highly conserved residues, such as cysteines or asparagines [46][47] essential for intra- and interchain bonding, are usually more severe. Substitutions for cysteine have also showed to have severe consequences. This could be due to the spontaneous formation of disulphide bonds that can alter the initial trimerization of the three proα chains.

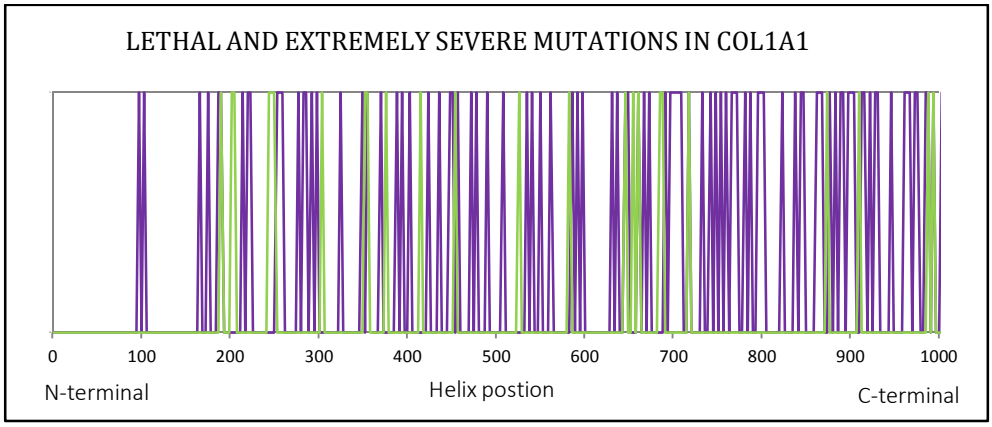
Another region that showed to have lethal mutations is found between residues 1277 – 1285 in proα1 chain and residues 1181 – 1189 in proα2 chain, in both cases found at the C-terminus (exon 50). This highly conserved region with the sequence I **D** P **N** **Q** G **C** [T/N] [M/L] **D** A I K V has shown to be a lethal cluster. The Asp, Asn, Gln, Cys and Asp residues (highlighted in bold in the sequence) are responsible for the binding of calcium ions essential for the mineralisation of bone [114][115]. The substitution of these residues resulted in OI type II phenotypes, perhaps due to the impossibility of incorporating calcium ions and form properly hydroxyapatite crystals.

In normal conditions both N- and C-terminus are cleaved off by proteinases. Mutations in ADAMTS2[16] that abolish the function of the procollagen N-proteinase cause Ehlers-Danlos syndrome type VIIC, characterized by severe joint hypermobility and slight osteopenia. The non-cleavage of the N-terminus and thus its incorporation into ECM causes a minimal involvement of bone [37]. This is consistent with the fact that mutations in exon 6 and upstream, including the procollagen N-proteinase recognition site, cause a phenotype identical to EDS type VIIC (EDS types VIIA and VIIB, see **Table 1**).

In contrast, the non-cleavage of the C-terminus causes a phenotype comparable with OI type III. Mutations that abolish the function of the procollagen C-proteinase (aka bone morphogenetic protein 1; BMP1[17][83]) cause the incorporation of the globular C-terminus into the ECM. This causes OI type XIII, a severe recessive form of the disease with extreme bone fragility and deformation of limbs.

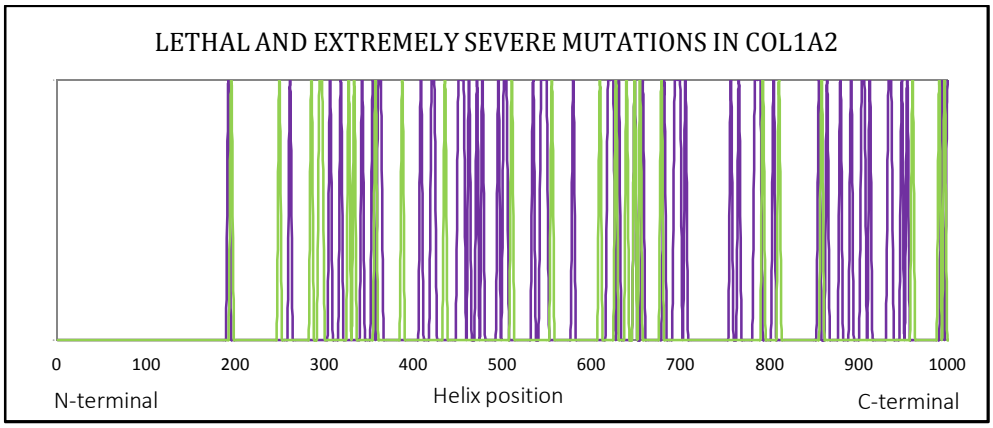
[16] OMIM entry *604539
[17] OMIM entry *112264

When observing the mutations in the triple helical domain from the training set it results that lethal mutations occur more often at the C-terminal end than at the N-terminal end. As shown in **Graph 1**, among the first 200 residues of the proα1 chain only a few lethal mutations are present whereas towards C-terminus the density of lethal mutations increases.



Graph 1 | Representation of the distribution of lethal glycine substitutions along the proα1 chain. The green lines represent extremely severe mutations and the purple lines represent lethal mutations.

The same schema can be observed in proα2 chain. Although the number of lethal glycine substitutions is fewer than in proα1 there is also a tendency of lethality towards C-terminus as can be seen in **Graph 2**.

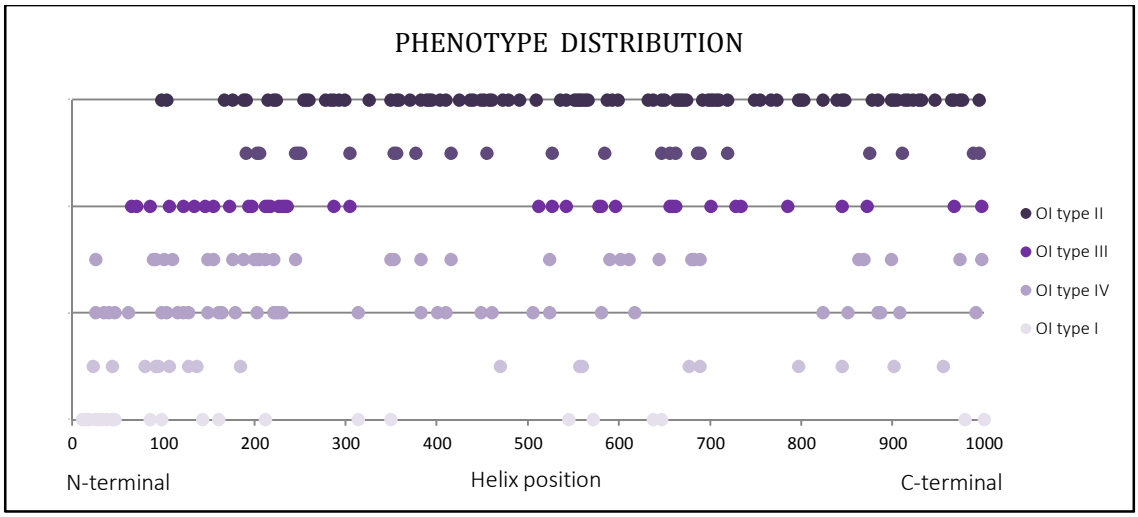


Graph 2 | Representation of the distribution of lethal glycine substitutions along the proα2 chain. The green lines represent extremely severe mutations and the purple lines represent lethal mutations.

This increasing distribution of lethal mutations towards C-terminus can be explained because of the propagation direction of helix folding. When a mutation occurs near C-terminus there is a very small region that can normal self-assemble. Depending on whether renucleation occurs N-terminal to the mutation site, [31] stable tropocollagen molecules will form or not.

The increasing number of lethal mutations can also be observed in **Graph 3**. There are represented all the mutations identified with its corresponding phenotype.

It can be seen that mutations causing the mild type of OI are distributed N-terminal on the helix. This is consistent with the lower disruption of helix folding. In contrast, it can be seen that severe and lethal mutations are in the C-terminal region.



Graph 3 | Representation of the mutation distribution along the triple helix. Dark purple dots indicate more severe phenotypes (lethal mutations at the top, mild ones at the bottom). The dots between lines represent intermediate phenotypes between two subtypes of OI. As can be seen, the density of lethal mutations is higher at the C-terminal site while the density of mild OI type is higher at N-terminal.

As observed, there is a substantial genotype-phenotype correlation depending on the helix position of the mutation. However, it is not consistent and it has a lot of exceptions. It is clear that mutations disrupting helix folding result in more severe phenotypes, but there is no evidence that high chain overmodification and delay of helix folding correlate directly with the severity of the disorder.

Identity of substituting amino acid

As explained above, when glycine is substituted by another residue, eight different amino acids can result: aspartic acid, glutamic acid, cysteine, serine, alanine, valine, arginine and tryptophan. The identity of these eight amino acids has got a relative strong correlation with the severity of the disease.

Glycine is the amino acid with the smallest side chain (single hydrogen) [117], needed to fit into the centre of the triple helix. If this hydrogen is replaced by a charged and/or bulkier side chain, the triple helix is interrupted [12][14]. **Figure 19** shows a representative mechanism of how this occurs, which can be compared with the effect of a wedge.

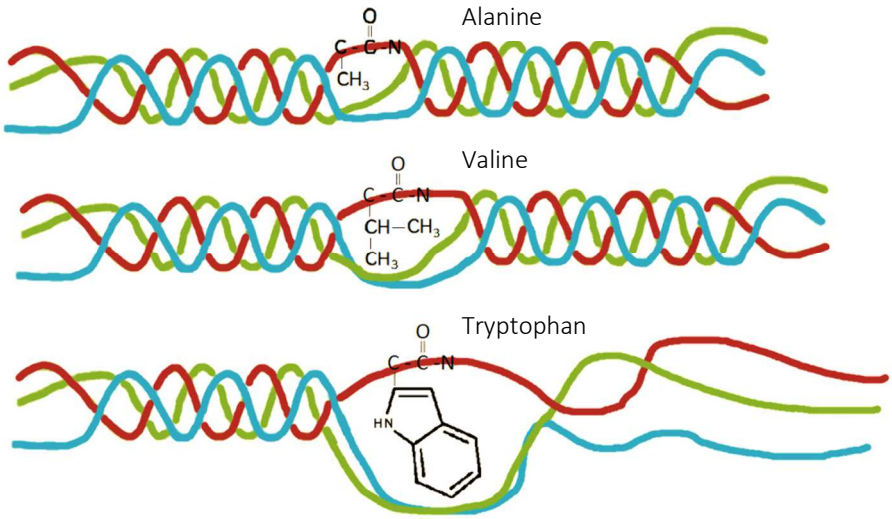


Figure 19 | Representation of tropocollagen molecule containing a glycine substitution. The upper molecule contains an interruption in the triple helix caused by a residue with a small side chain. The middle one contains a larger residue, but renucleation can still occur. The lover molecule has an interruption caused by a larger residue, but the molecule can no longer fold into its native structure.

Although eight amino acids can result, only seven are reported as substituting for glycine. Tryptophan has not been included since it is only reported twice, in both cases resulting in OI type III phenotype. The lethal mutations found in the training set are listed in **Table 7**.

	proα1			**proα2**		
	Nº mutations	Nº Lethal	% Lethal	Nº mutations	Nº Lethal	% Lethal
Alanine	32	3	9.4 %	10	0	0.0 %
Serine	90	20	22.2 %	63	5	7.9 %
Arginine	46	25	54.3 %	46	9	19.6 %
Valine	31	22	71.0 %	53	13	24.5 %
Glutamic acid	8	5	62.5 %	22	6	27.3 %
Cysteine	51	15	29.4 %	35	7	20.0 %
Aspartic acid	48	34	70.8 %	69	22	31.9 %
Tryptophan	0	0	0.0 %	2	0	0.0 %
Total	306	114	37.3 %	300	62	20.7 %

Table 7 | The lethal mutations found in both chains are listed. The first column shows the total number of mutations found for each amino acid. The second column lists the number of lethal mutations for each amino acid. The third column is the percentage of lethal mutations for each amino acid.

A clear relationship can be found between the identity of the substituting amino acid and the lethality. Firstly, we can see that the overall lethality in proα2 chain is less than in proα1. The amino acid gradient for lethality results to be Ala < Ser < Cys < Arg < Val ≤ Glu < Asp. Tryptophan is not included because the number of mutations identified is irrelevant and statistically insignificant. To study the phenotypic outcome of substitutions for tryptophan, a greater number of samples are needed.

The aberrancy of each substituting amino acid correlates with the chemical properties of the residues. As seen in **Figure 20**, long, branched and charged side chains result to disrupt more the triple helix. The correlation found is consistent with the one posed by Beck et al. (2000)[4].

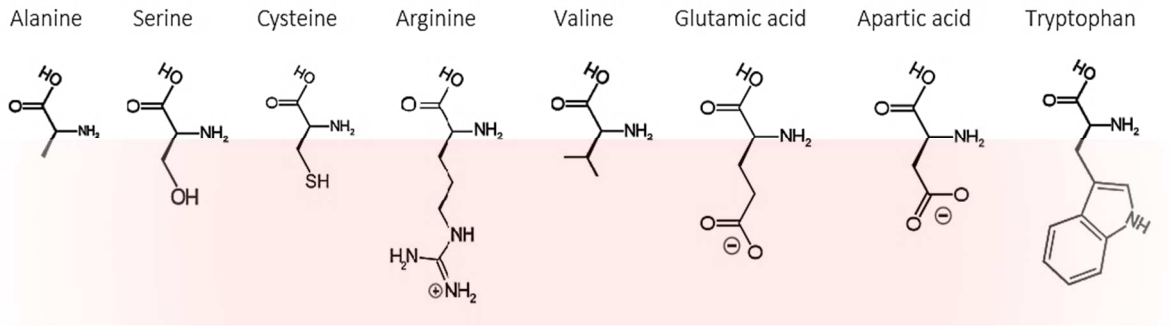
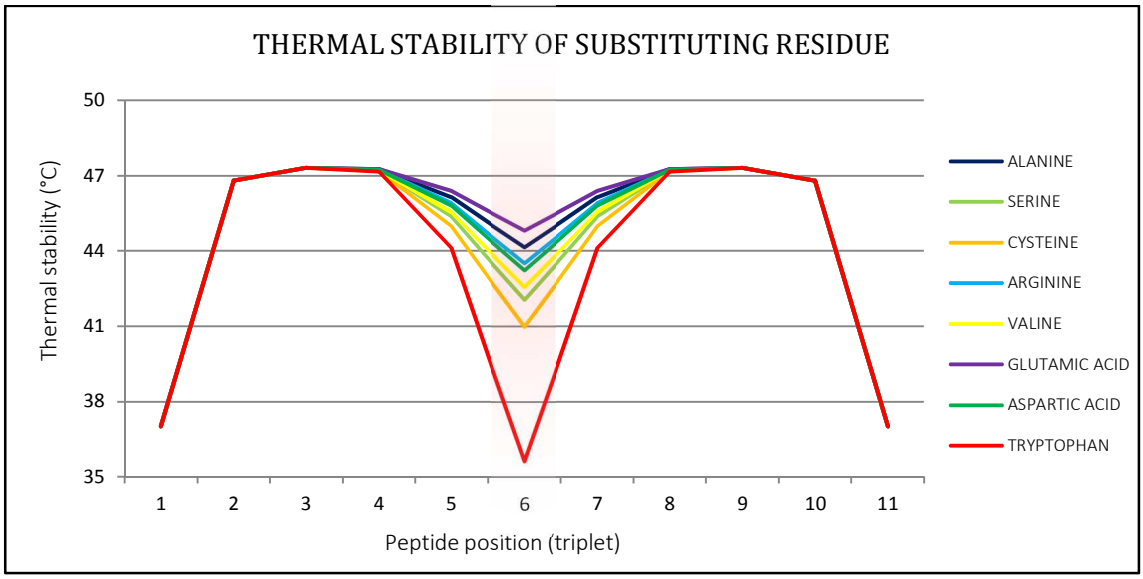


Figure 20 | Substituting amino acids, ordered by increasing lethality. The different side chains are in the red-shadowed area.

The charged chains of aspartic acid and glutamic acid have shown to have the most aberrant results. The little side chains of alanine, serine and cysteine can be fit into the tiny spaces and do not disrupt the triple helix as pronounced as the other residues.

To determine better the disruption of the triple helix each substituting amino acid was evaluated in a host-guest training set (**Graph 4**). The melting temperature of each residue was calculated, but it resulted that there is no relationship between lower thermal stability and greater severity or disruption.



Graph 4 | Thermal stability of substituting amino acids in host-guest peptides $(GPO)_5 – GXaaO – (GPO)_5$. The guest residue was incorporated in the Xaa position (6^{th} triplet) flanked with high-stabilising sequences (GPO) at both terminal ends. There is no relationship between the thermal stability and the lethality of the residues.

The host-guest training set resulted in no relationship between the melting temperature of the amino acids and its severity when they substitute a glycine residue. Nonetheless, we can observe that tryptophan has a great destabilizing effect. Even clustered between the high-stabilizing GPO sequences the melting temperature of the peptide is below normal body temperature.

The low percentage of tryptophan mutations is probably due to the low possibility of a single-base substitution result in a tryptophan-coding triplet. In both COL1A1 and COL1A2 there are only 26 possible GGG triplets that could become an AGG triplet and thus code for tryptophan. However, the G → A transition is relative common, more than G → C and G → T transversions. The small amount of reported substitutions for tryptophan could also be due to a high destabilizing effect of this residue leading to early spontaneous abort of embryos carrying these mutations. This could lead to a non-detection of these mutations resulting in that low number of reported tryptophan substitutions.

The fact that tryptophan could be severely destabilizing can be supported by three evidences:

1. Tryptophan has a bulky hydrophobic side chain containing an indole functional group. This could interfere with assembly, more than aspartic acid or glutamic acid because it is very difficult to accommodate such a big side chain in so a tiny space.

2. Host-guest peptides containing a tryptophan residue show very low melting temperatures. The average difference of thermal stability between the other residues and tryptophan is greater as could be seen in **Graph 4**.

3. The two reported substitutions for tryptophan resulting in the type III form of OI are found in high stability regions, meaning that even in these regions its outcome is very severe. This fact supports the idea that tryptophan substitutions are unconditionally lethal along the chain in normal- and less-stable regions (almost all the chain length).

These three facts could explain the low occurrence of these mutations. If this hypothesis is true, the lethality gradient of residues would be Ala < Ser < Cys < Arg < Val ≤ Glu < Asp < **Trp**.

However, this classification is not perfectly accurate. It has a relative high sensibility in detecting the lethal mutations although the specificity is low because of the numerous exceptions and regions where almost all the substitutions have lethal outcomes. It can be said that it is a good initial method to assess the lethality of glycine substitutions even though other factors must be considered.

3. 5. COMPLEX REGIONAL MODELS

The fact that some lethal mutations are clustered in specific regions along the chain suggests the existence of a regional model. In anterior chapters it has been seen that there is a regional model based on the increasing severity towards C-terminal, but it is not rigorous and it has many lacks.

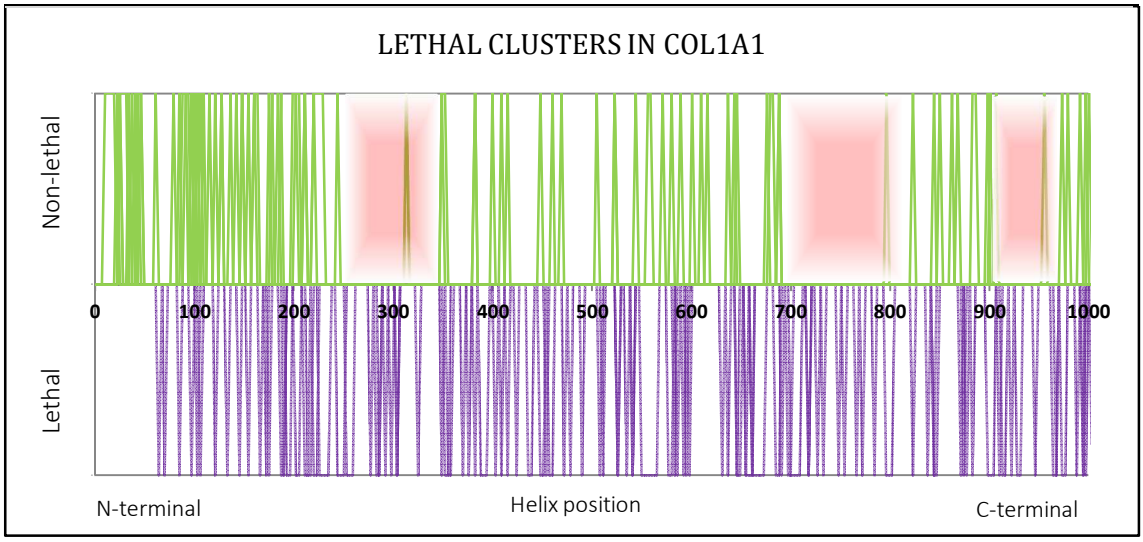
While analysing the mutations in the training set it has been noted that the lethal mutations are grouped in regions where almost every substitution leads to OI type II or II/III.

COL1A1 gene

In the proα1 encoding gene there are three regions that show an elevated number of exclusively lethal glycine substitutions. These clusters correspond to helix positions

1. 250 → 350
2. 690 → 830
3. 885 → 980

As can be seen in **Graph 5**, between the anterior mentioned helix positions there are almost only lethal mutations.



Graph 5 | Representation of the lethal clusters in COL1A1. The regions shadowed in red are the ones in which only lethal mutations are found, giving rise to this regional model.

These three lethal regions overlap with important ligand binding sites (LBS) in the collagen molecules [20][62]. In **Figure 21** the lethal clusters are represented on the collagen D-period with the major ligand binding regions (MLBR). There are three main MLBRs and lethal regions 2 and 3 overlay with MLBR 2 and MLBR 3 [40]. This suggests that mutations in these regions can abolish the binding site and thus interfere with ligand binding, giving rise to a very severe OI phenotype.

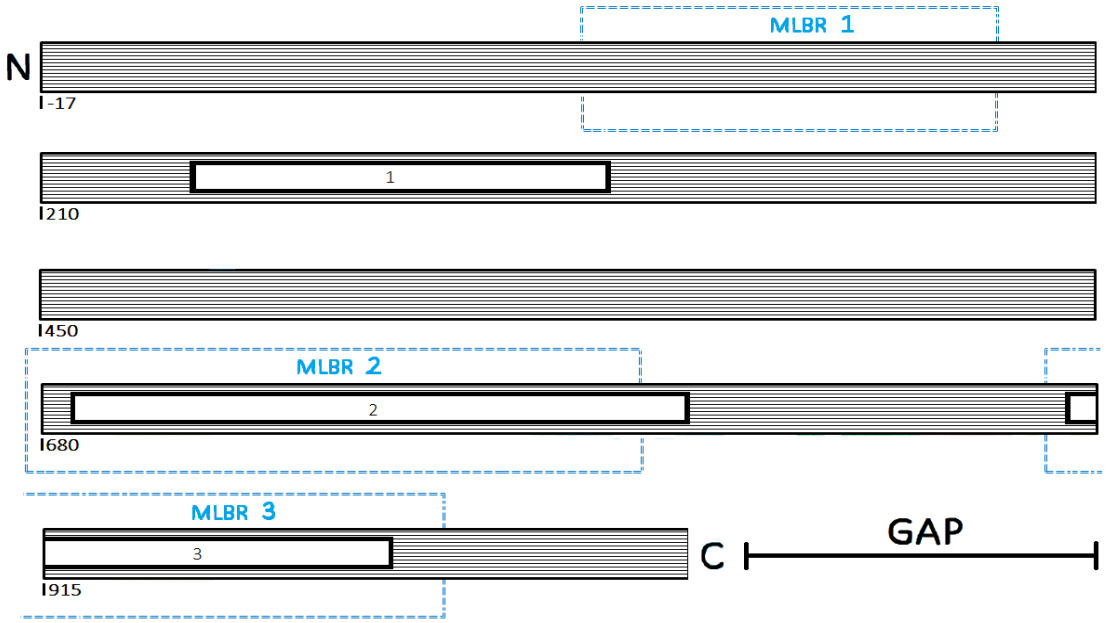


Figure 21 | Representation of the lethal clusters in COL1A1 on the D-period. The blue dotted boxes represent the MLBR of collagen type I. The white boxes represent the lethal clusters. As can be observed, two of the three exclusively lethal regions overlay with MLBRs 2 and 3.

MLBR 2 comprises ligands like integrin, fibronectin, phosphophoryn and cleavage sites for matrix metalloproteinases. MLBR 3 includes important residues for intermolecular crosslinking (Lys^{930}) and also ligand binding sites like decorin and various glycosaminoglycans [20][62].

It can also be observed that all three lethal regions are found in the overlap and not in the gap region. This could also influence the severity of the disease. In **Graph 5** it can be observed that non-lethal mutations are located between the lethal clusters and therefore in the gap region. One hypothesis is that the mutations in the gap region could be better accommodated into the fibril than the ones in the overlap region. The greater space availability can be an advantage for bulky mutated molecules. The disturbance of aberrant tropocollagen molecules on the whole fibril would be lesser since in the gap region they have a greater freedom of movement. This is supported by the fact that near C-terminus there is a large region (positions 830 → 880) with non-lethal mutations which matches with the gap region. At this position, following the severity gradient towards C-terminus, it would be more logical that these mutations end up in severe/lethal OI types rather than in moderate and mild OI types.

COL1A2 gene

In the proα2 encoding gene there are eight regular clustered regions that show an elevated number of lethal glycine substitutions. These sequences correspond to helix positions

1. 320 → 365
2. 450→ 500
3. 545 → 580
4. 615 → 635
5. 690 → 710
6. 755 → 810
7. 855 → 915
8. 935 → 1000

These eight regions overlay with binding sites for glycosaminoglycans and the protein phosphophoryn (**Figure 22**)[20]. As in COL1A1, these regions have a key role in matrix interactions and mutations at these sites can lead to the abolishment of ligand binding.

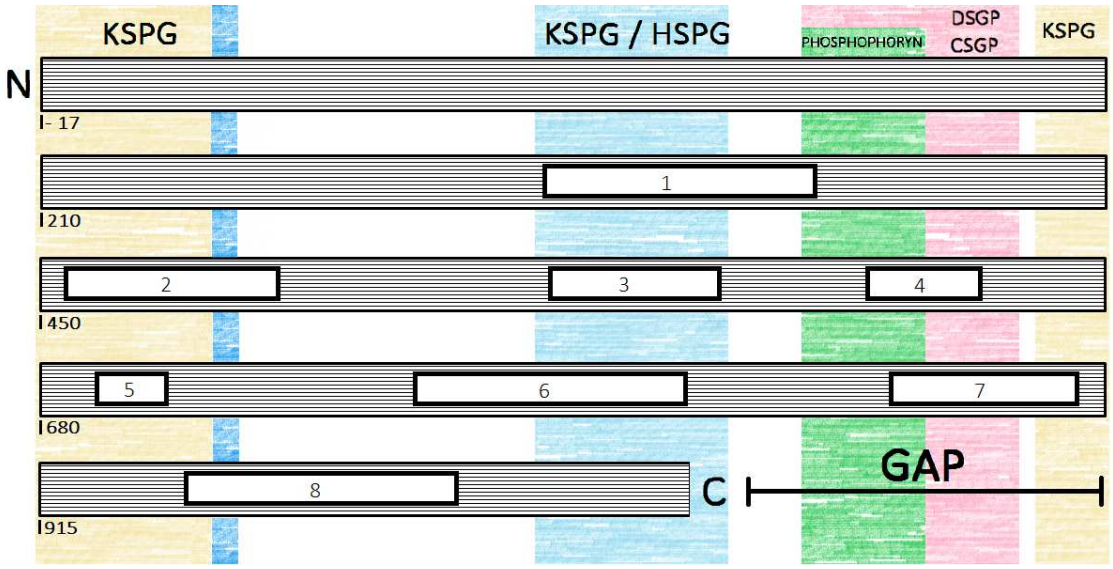


Figure 22 | Representation of the lethal clusters in COL1A2 on the D-period. The colour-shadowed regions are ligand binding sites. Keratan sulphate (KSPG), heparan sulphate (HSPG), dermatan sulphate (DSPG) and chondroitin sulphate (CDPG) are glycosaminoglycans that bind to collagen.

The regularly distributed lethal cluster is highly suggestive for a regional model related with ligand binding sites. The abolishment of the LBS probably worsens the OI phenotype. Evidence for this is that alterations in the processing of glycosaminoglycans result in various diseases (mucopolysaccharidosis), all of them related to skeletal dysplasia. This fact supports that glycosaminoglycans have a key role in bone and their mismanagement can lead to additional alterations.

3. 6. SPLICE SITE MUTATIONS

Splice site mutation is a type of mutation that affects the sequence around the specific sites where splicing of introns takes place. Splicing, as seen in chapter 3.3., is a maturation process of the mRNA that cuts out the non-coding regions known as introns.

To carry out the splicing process certain molecules recognise specific sequences where they have to cut and glue the mRNA sequence. Analysing the sequences where splicing occurs in various species, a consensus splicing sequence has been set (**Figure 23**) [45].

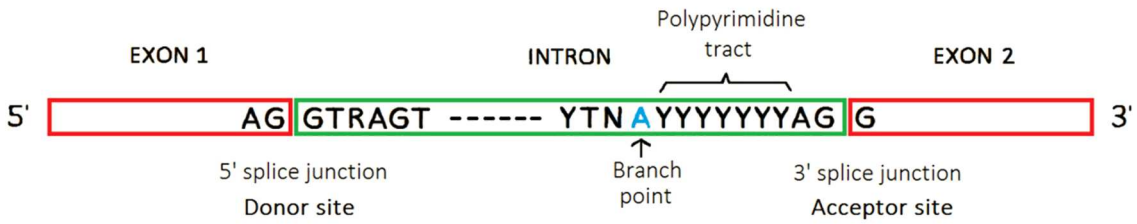


Figure 23 | Splice site consensus sequence. The nucleotides forming part of the splice site sequence are represented (R, purine; Y, pyrimidine; N, any nucleotide).

Mutations in these highly conserved regions can abolish the splicing process of introns having different outcomes. The consequence of these mutations cannot be unified because different elements influence the way in which the mutation is expressed. Factors that can influence splicing and thus have an effect on mutation expression are:

1. **Extension of the mutation**. Some substitutions do not have any effect on splicing while others can abolish the mechanism through the disruption of the consensus sequence.

2. **Cryptic splice site donors and acceptors**. If a wild type splice site is mutated it is possible that cryptic splice sites are activated and used. Cryptic splice sites are intronic or exonic sequences that are similar to the wild type splice site. They can be used as donor or acceptor sites instead of the wild type altering the mature mRNA.

3. **Exonic and intronic splicing silencers/enhancers**. These sequences regulate the proper mRNA splicing process and changes in these regions can lead to altered splicing of introns.

The two consequences of splice site mutations are exon skipping or intronic inclusion, meaning that exons are eliminated from the mRNA or introns are conserved and consequently transcribed into protein (**Figures 24** and **25**) [45].

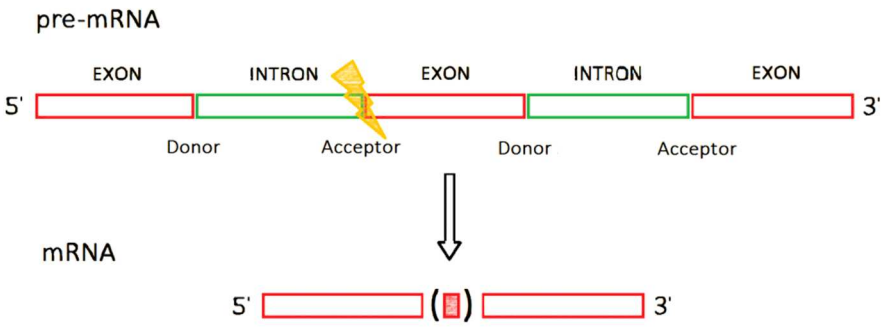


Figure 24 | Typical situation in a case of splice site mutation and exon skipping. The intron-exon acceptor site is mutated (yellow mark) and the splicing mechanisms do not recognize the site. This leads to the recognition of the following exon as non-coding DNA and thus is eliminated from the mature mRNA. The mechanisms remove all the sequence until they identify the following acceptor site [45]. This situation is supposed that no cryptic splice site is used.

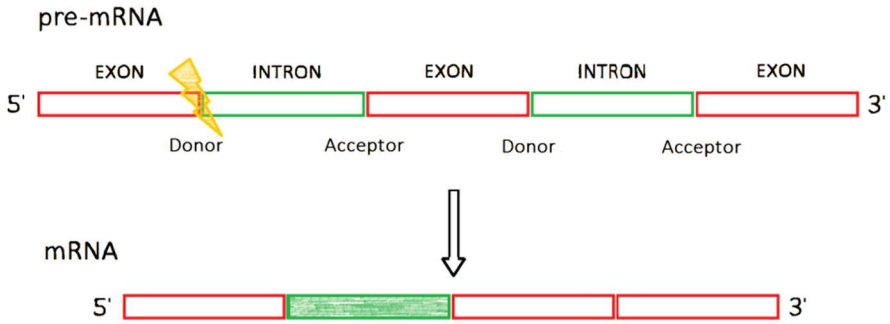


Figure 25 | Typical situation in a case of splice site mutation and intron retention. The exon-intron donor site is mutated (yellow mark) and the splicing mechanisms do not recognize the site. This leads to the recognition of the following intron as coding DNA and thus is incorporated into the mature mRNA. The mechanisms retain all the sequence until they identify the next donor site [45]. This situation is supposed that no cryptic splice site is used.

In collagen type I genes whole exon skippings are always in-frame because all the exons begin with a glycine codon and end with a Yaa codon. This means that when exons are eliminated the mRNA will be shorter but never will trigger NMD, and therefore the shortened protein will be incorporated into tropocollagen molecule.

On the contrary, intron inclusions or partial exon skippings can be out-of-frame or in-frame depending on the number of incorporated or eliminated nucleotides. The use of cryptic splice sites or the intron length determines the reading frame. Splice site mutations that result in frameshifts or premature stop codons can activate NMD and result in a matrix insufficiency rather than a qualitative defect. If the mutation is in-frame the mRNA will be shorter or longer than normal and incorporated into tropocollagen molecule.

The consequence of in-frame splice site mutations that do not trigger NMD is the destabilisation of the whole tropocollagen molecule, having a dominant-negative effect. Another consequence of exon skippings or intron inclusions are the abolishment of important ligand binding sites or cleavage sites. This leads to additional disturbances and increases the consequences and severity of the mutation.

Splice site mutations training set

This study includes a total of 200 splice site mutations, 139 in COL1A1 and 61 in COL1A2 (**Table 8**). To assess the consequences of the splice site mutations the Human Splicing Finder is used in order to predict the outcome of splicing defects.

	COL1A1			COL1A2		
	Nº mutations	Lethal	% Lethal	Nº mutations	Lethal	% Lethal
Donor sites						
+1	35	3	8.57 %	16	4	25.0 %
+2	13	3	23.08 %	6	0	0.00 %
+3	2	0	0.00 %	1	0	0.00 %
+4	2	0	0.00%	1	0	0.00 %
+5	12	3	25.0 %	9	4	44.44 %
Indels	9	1	11.1 %	9	2	22.2 %
Others	2	0	0.00 %	3	1	33.3 %
Total	75	10	13.3 %	45	11	24.44 %
Acceptor sites						
-1	27	0	0.00 %	4	0	0.00 %
-2	25	3	12.0 %	8	2	25.0 %
-3	4	0	0.00 %	2	1	50.0 %
-4	0	0	0.00 %	0	0	0.00 %
-5	0	0	0.00 %	0	0	0.00 %
Indels	5	0	0.00 %	1	0	0.00 %
Others	3	0	0.00 %	1	0	0.00 %
Total	64	3	4.70 %	16	3	18.75 %

Table 8 | The splice site mutations from the training set are listed. They are classified according to whether they affect donor or acceptor sites. In each category, the mutations are broken down according to the position of the splicing sequence that they affect. Indels are reported separately and duplications or mutations affecting further positions are recorded in the category 'others'.

Splice site mutations in COL1A2 are in 21.6 % of the cases lethal whereas in COL1A1 only 9.0 % are lethal. This fact differs from glycine substitutions which showed a higher lethality rate in COL1A1. The majority of mutations are located at splice donor sites and in case of COL1A2 there is three times more donor than acceptor site mutations.

In splice acceptor sites all the mutations are located at the -1, -2 and -3 positions. They result from the substitution of the obligate A^{-2} and G^{-1} nucleotides for other ones. Mutations at -3 position resulted to be substitutions of different nucleotides but in all cases for guanine, suggesting that this nucleotide at this specific position could interfere with splicing. The non-detection of mutations in positions -4 and -5 suggests that these positions are not crucial for splicing.

In splice donor sites the mutations are distributed more heterogeneous between the +1 and +5 positions. Positions +1 and +2 account more than half of thee mutations. Surprisingly, position +5 has several mutations and one third of them have lethal outcomes. This fact also suggests that G^{+5} has a critical role in splicing, in contrast to +3 and +4 that only have a few detected mutations.

Indels are located among the positions +1 → +5 and -5 → -1. Almost all of them delete one or more nucleotides abolishing the splicing site. In some cases a random nucleotide is inserted with no further consequence for splicing, but it is possible that it creates a frameshift and a premature stop codon downstream.

In the category of other mutations, duplications and further intronic substitutions are included. Only one of them resulted in a lethal outcome. The mutation results in the substitution of C^{+11} for thymine (intron 25), and its analysis predicted a creation of an ESE site and potential alteration of splicing.

An overall correlation has been noted concordant with the model of the severity gradient towards C-terminus. In COL1A1, in the first 14 introns no lethal mutation has been found, and in COL1A2 the first 25 introns are free of lethal mutations. This fact is concordant, like in glycine substitutions, that if a larger region can normal assemble, lesser is the disruption of the triple helix and therefore the outcome is milder.

No correlation has been found with the identity of the substituting nucleotide.

Consequences for mRNA and protein

The analysis of the single mutations in the training set showed a wide heterogeneity in terms of phenotypic outcome. In many mutations the structure of the mature mRNA cannot be predicted exactly since there are various possibilities depending on the use of cryptic splice sites and the creation of ESEs and ESSs.

However, some clues of how splice site mutations are expressed could be obtained. The following exposed situations were obtained from the splicing prediction tool and since the mutant mRNA could not be studied directly, they could not be 100 % accurate.

1. **Mild forms**. OI type I and type I/IV overlapping result from mRNA that is substrate for NMD and exon skippings near N-terminus.

   To produce a premature termination codon and therefore trigger NMD, either a frameshift or inclusion of an intronic stop codon must occur. Mutations at donor or acceptor sites that produce retention of introns are candidates to produce frameshifts resulting in degradation producing a practical null allele. The use of cryptic splice sites is also highly probable to cause a frameshift and therefore end in NMD. To produce an ECM insufficiency and not a qualitative defect, all the products of the mutant allele must be degraded.
   As explained above, exon skippings far downstream i.e. near the N-terminus do not disturb greatly helix folding and have a relatively large well-assembled region, so that the outcome is not so severe (**Figure 25 A**).

2. **Moderately severe forms**. OI types IV and III are caused by different spliced mRNAs, meaning that there are multiple alternatively spliced forms and therefore diverse protein forms are synthesised.

   Simple exon skippings account the minority of splicing products while the other forms are PTCs or cryptic splice sites. The mixture of both products produces a quantitative defect combined with qualitatively altered proteins that are incorporated into the ECM. This situation is produced by both in-frame and out-of-frame mRNA variants of the same allele, making that part of the transcripts are subjected to NMD while the other part is translated into defective proteins. The consequence is an ECM insufficiency with partially unfolded collagen molecules (**Figure 25 B**).

3. **Severe and lethal forms**. OI type II and type II/III overlapping are caused by simple in-frame exon skippings and in-frame retentions. No premature stop codons are present, and only cryptic splice sites that maintain reading frame are used. The result is a qualitative altered matrix comparable with collagen molecules containing aberrant and destabilising glycine substitutions (**Figure 25 C**)

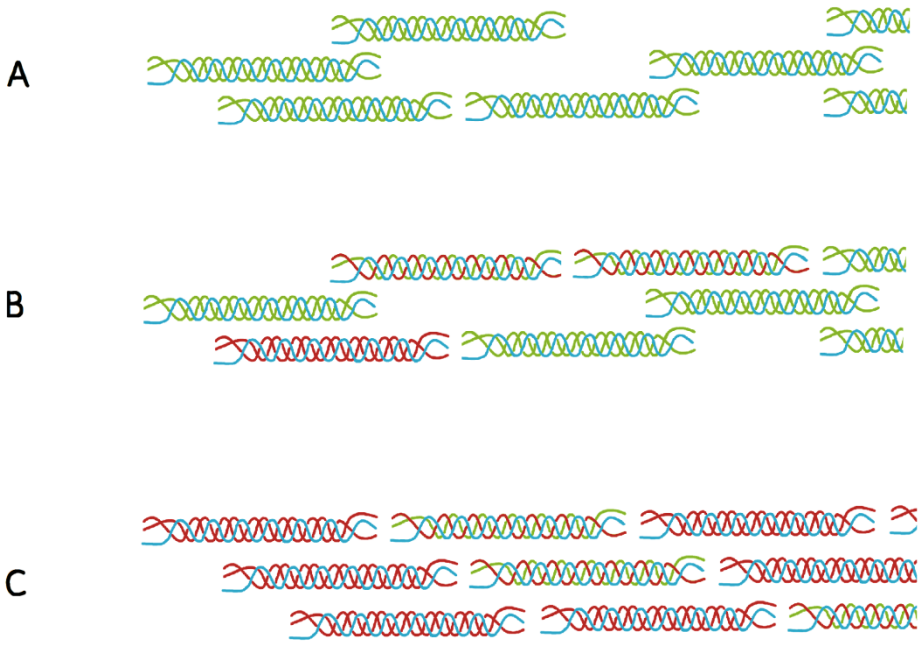


Figure 25 | A) Matrix insufficiency resulting from a quantitative defect. No aberrant proteins are present. This situation results in a mild form of OI, presumably type I or I/IV. **B)** Matrix insufficiency combined with altered proteins (collagen molecules with a red α-chain). This situation results in the moderate and severe forms of OI (types IV and III). **C)** Exclusive qualitative defect with no matrix insufficiency. This collagen defect results in OI type II/III or the perinatal lethal (OI-II) form.

The role of exonic splicing enhancers and silencers could not be estimated precisely. Deep intronic mutations could result in the creation or elimination of these sequences, although in this training set only one of these mutations is reported. The outcome of these alterations would vary widely depending on the effect they have on splicing.

Another possible alteration is the punctual mutation resulting in the substitution of the obligate adenine nucleotide in the branch point. This could lead to a splicing defect since the lariat-shaped molecule could not form. The outcome of this change is not clear, although it could lead to partial or total skipping of the following exon. There are only a few described branch-point mutations; one of them is reported in a collagen-encoding gene (COL5A1) and causes a hereditary connective tissue disorder, Ehlers-Danlos type II [8]. This supports the fact that they are also possible in COL1A1 and COL12 and therefore candidates to cause OI.

3. 7. INTRINSIC AND EXTRINSIC FACTORS

At this point of the research project it has been noted that finding a clear genotype-phenotype correlation will be difficult and even challenging for professional researchers. Nonetheless, in this following part, some factors will be discussed that can influence the pathological outcome of mutations and thus make this study more difficult. None of these aspects can be neither measured in numbers nor be exactly quantified, but they have a key role in modulating the phenotype. They are based on relationships from observable manifestations; however, they can have different influence depending on the persons.

At first the environmental factors were ignored when studying this condition. However, over recent years it has been noted that they can influence indirectly the disorder phenotype. Availability of treatment, diet, physical activity and other factors can contribute to the severity of the disease. Although they do not change the underlying genetic defect, they can impact greatly on the quality of life of the patient and either improve or worsen the syndrome.

The role of non-collagenous proteins (NCPs) in osteogenesis imperfecta is also still not well understood. These molecules have indispensable roles in posttranslational modifications or matrix mineralisation and thus are essential for the stability of bones and dentin [30]. The correlation between NCPs and the underlying collagen defect is studied through the observation of the non-classical osteogenesis imperfecta types. These types of brittle bone disease show the same features as OI types I → IV, usually resulting in a severe phenotype. Recently, the genes related to these uncommon types of OI have been identified and most of them affect collagen ligands, hydroxylases and chaperones [40][61].

So far, we know that collagen is a structural protein found in extracellular matrix and in most tissues it is found as a flexible molecule forming part of soft organs. The posterior mineralisation of collagen networks gives the rigidity to bone and dentin.

While doing this research project, two questions arose:

- Why is bone more affected by collagen defects than other tissues such as tendon and skin, if those are also made up of collagen?
- Which factors contribute to the different expression of a single type of mutation?

3. 7. 1. Unfolded protein response

Unfolded protein response (UPR) is a cytoprotective surveillance system triggered by ER stress in order to maintain cell homeostasis. If a cell produces a lot of misfolded or unfolded proteins and they accumulate in the lumen of the ER, a series of mechanisms are activated to reduce and reestablish normal cell function [36][56][125].

The activation of UPR is a protection against the production of aberrant proteins. In many cases cell homeostasis can be achieved through this mechanism. A number of proteins[18] are released in a cascade of intracellular signalling events having different effects on the cell.
UPR is often activated in tumour cells and can be induced through nutrient deprivation, hypoxia and other metabolic disturbances [125].

This surveillance mechanism also has crucial role in neurodegenerative diseases such as Alzheimer's disease, Creutzfeldt–Jakob disease and other prionic and amyloid disorders [120]. Most of these illnesses are proteinopathies, mainly caused by misfolded proteins that accumulate in brain cells. This pathologic accumulation of misfolded proteins induces UPR and then cell death giving rise to the neurodegeneration seen in these diseases.

As can be observed, this surveillance system can have positive or harmful consequences on the organism depending on the situation in which it is activated.

UPR in osteogenesis imperfecta

In many cases, mutant collagen chains trigger UPR. This has a harmful effect on the person harbouring the mutation since the cell will never produce normal proteins and thus the adaptive objective of this mechanism is not achieved. Normally, varying conditions can activate UPR when misfolded proteins are detected but it is fully reversible when ER stress lessens.

Mutant collagen chains containing either glycine substitutions or splice site defects accumulate in the ER lumen. The defective alpha chains slow down the nucleation process, delaying helix folding and exposing the N-terminal region to modifying enzymes. The everlasting UPR is not deactivated since the cause of the UPR (mutation) will never 'disappear' and the whole collagen synthesis process is hindered.

[18] The proteins released during UPR are IRE1, ATF6 and PERK (ordered by activation). This protein complex determines the duration of the UPR and if they are released during a long time, they can induce apoptosis (specially PERK) [125].

The abnormal forming chains accumulate and aggregate in the ER. At this point, UPR is triggered and a succession of events occurs in order to normalise the situation (**Figure 26**) [125].

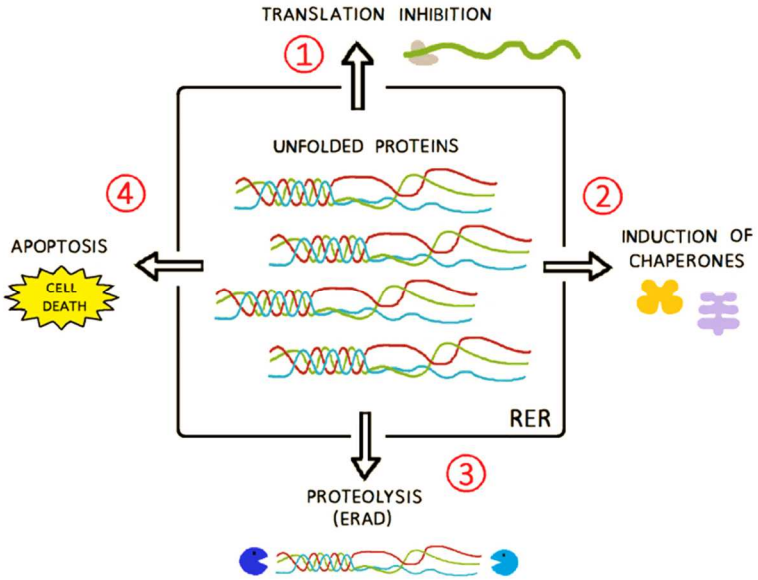


Figure 26 | Representative illustration of the cascade-like events that occur when UPR is activated.

1. First of all, molecules that inhibit translation are released. The number of synthesised proteins is lower giving more time to fold proteins and relieve the cell from ER stress.

2. If this anterior step does not lessen UPR, factors are released that induce the production of chaperones. Chaperones are proteins that assist non-covalent folding of proteins. A higher number of ER resident chaperones can help to fold the misfolded or unfolded proteins.

3. At this point the unfolded proteins are degraded by a co-mechanism known as endoplasmic-reticulum-associated protein degradation (ERAD). This mechanism tags the protein as aberrant and delivers it to the protein-degrading complex called proteasome.

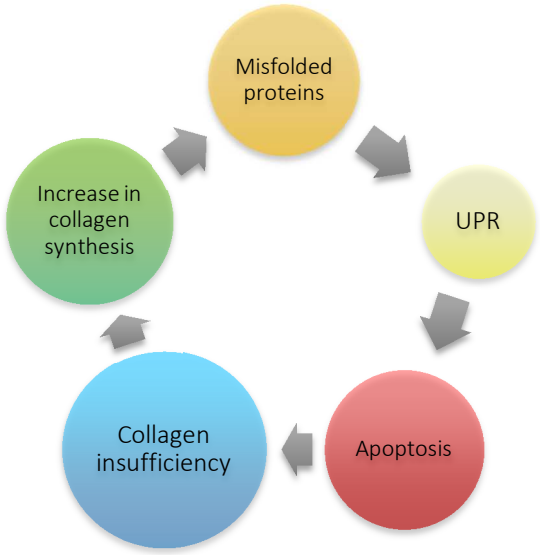
4. The last step of UPR is programmed cell death (PCD). If the anterior steps could not relieve the ER stress, apoptosis is initiated.

These four steps compose the UPR process [125]. In osteogenesis imperfecta, the first two do not work properly because the unfolded proteins contain mutations that prevent the correct folding. However, translation inhibition stops the exponential accumulation of misfolded collagen. Proteolysis is then induced degrading the misfolded collagen molecules. This has a temporarily mitigation effect on ER stress, but the ever new coming defective alpha chains maintain indefinitely the UPR. After a determined time lapse, the collagen-synthesising cells end in apoptosis.

Consequences of UPR for the OI phenotype

The consequences of apoptosis are evident; there is an exponential decrease in the number of osteoblasts. Osteoblasts are the cells responsible for the synthesis of collagen in bone.

UPR is presumably implicated in osteogenesis imperfecta types II, III and IV. In quantitative defects (OI-I) the proteins are normal, there is no misfolding and therefore no apoptosis.
The fact that the number of collagen-synthesising cells decreases suggests that UPR worsens the OI severity. Apart from the underlying structural defect of collagen, the decrease in collagen-producing cells drops the overall collagen availability [110].
In **Graph 6** a possible long-term consequence of UPR in OI is showed. The reduction of osteoblasts probably makes that the other cells produce more collagen in order to compensate the loss. This fact gives rise to a feedback loop between higher collagen demand and higher apoptosis rate.



Graph 6 | Representation of the hypothetical vicious circle that produces UPR in osteogenesis imperfecta.

Although the pathophysiology of UPR in osteogenesis imperfecta is not well known yet, it is clear that it has a key role in modulating the phenotypic outcome of glycine substitutions and splice site mutations. The programmed cell death of osteoblasts gives rise to an overall loss of bone matrix subsequently leading to bone fragility and fracture propensity. The apoptosis process is presumably faster in the more severe types of OI whereas in milder forms the UPR rarely ends in PCD and mainly inhibits translation and/or degrades misfolded collagen.

3. 7. 2. Matrix mineralisation

The fact that collagen type I defects almost only affect bone and dentin must lie, among other reasons, in the difference that they undergo a mineralisation process. The addition of hydroxyapatite to the collagen network gives to bone and dentin its rigidity needed to support the forces they undergo continuously [50][52][66][120].

There is evidence that with age, crosslinks in skin and tendon collagen increase and thus loose elasticity. This means that to conserve flexibility of connective tissue a limited number of crosslinks are necessary, but not too much [120]. When collagen is defective, the lack of stability is traduced into an increased elasticity of tissues often seen in OI. Nevertheless, these features are mild and are not as severe as in other connective tissue disorders like various Ehlers-Danlos syndromes, in which skin and tendons appear very brittle. This gives rise to the fact that soft tissues are not very sensible to collagen type I defects, unlike dentin and bone.

In osteogenesis imperfecta there is a mineralisation defect often seen as a lack of mineral deposition giving rise to the brittleness of bones and teeth [66]. Hydroxyapatite, the main mineral found in bone and dentin, is incorporated into the tiny spaces (gaps) between collagen molecules [120]. However, this phenomenon is not uniform since there are also OI patients that show regions with high bone matrix mineralisation. Nonetheless, in all cases the mineralisation is irregular and not heterogeneous with decreased crystal size. This defect is possibly the uniform pathogenic consequence of collagen alterations[66]. The irregular extracellular network of collagen forces the mineral hydroxyapatite to deposit itself in an irregular pattern. The instability of the collagen web causes breakage of the mineral matrix, causing the bone fragility in OI.

The fact that some patients with OI show areas with overmineralisation does not indicate greater bone stability. Regions with a superior mineral deposit still have an irregular arrangement of hydroxyapatite, and therefore are fragile against tensile and flexural strength. A staggered arrangement of low and high bone density can also increase fracture risk, making that low-density regions found between high-density regions are susceptible to fracture.

The fact that overmineralised bone is also fragile can be supported when observing another condition called osteopetrosis (**Figure 26**). Osteopetrosis, also known as Albers-Schönberg syndrome or marble bone disease, is a heritable disorder characterised by osteosclerosis, an abnormal high bone density. The pathogenesis is a decreased activity of osteoclasts which participate in bone resorption and remodelling [82].

In osteopetrosis too much bone is formed and the old bone tissue is not destroyed. The main symptoms are bone pain, increased likelihood of fractures and haematopoietic alterations [82].

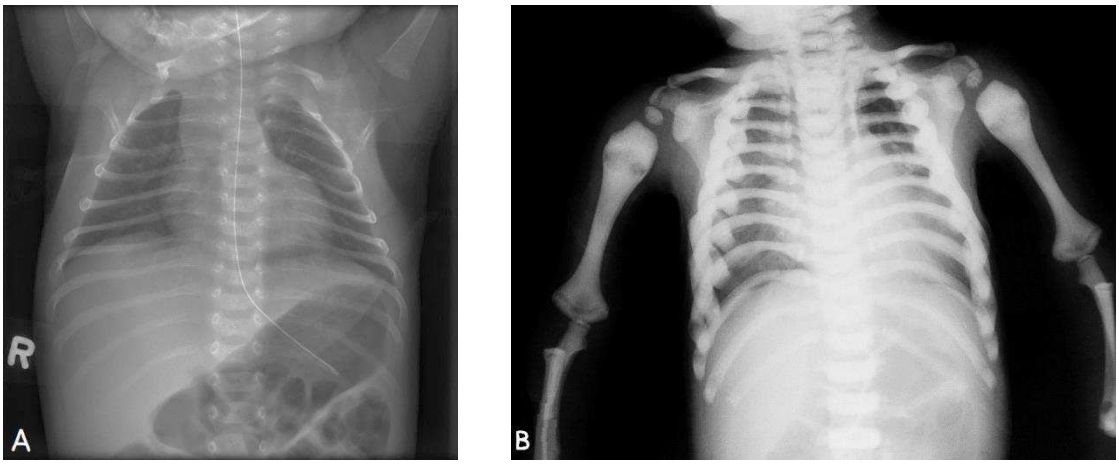


Figure 26 | a) Thorax X-ray from a patient with OI. The ribs appear thin, deformed and translucent, indicating low density of bone and decreased mineral deposit. **b)** Thorax x-ray from a patient with osteopetrosis. Ribs, arms and vertebrae appear radiopaque and thickened due to high bone mass. Both conditions present with bone fragility but contrary underlying pathogenesis. Compared with a control (normal) chest x-ray. [21]

Although both high and low bone mass are pathologic, bisphosphonates therapy in OI have shown to reduce fracture likelihood and bone pain. This group of drugs act as inhibitors of osteoclast activity and thus decrease bone remodelling and resorption. Treatment with these medications is very helpful during growth in children since the higher demand of bone tissue can be achieved reducing bone turnover, therefore increasing overall bone mass. Bisphosphonates do not improve or increase collagen synthesis or mineralisation, they only inhibit the down-breakage of bone giving more time to form new bone tissue [110][120].

Bone formation physiologically decreases after skeletal maturation since there is not the demand to form new bone tissue, only to maintain and replace the existing one. This is the reason why most OI patients improve after puberty having more bone mass and less fractures.

There is evidence that equilibrium must exist between new bone formation by osteoblasts and bone resorption by osteoclasts. When this equilibrium is interrupted, alterations occur usually seen as bone fragility. From this fact we can deduce that bone matrix destruction is a physiological process indispensable for the maintenance of bones.

3. 7. 3. Thermal (in)stability of collagen

Collagen is thermally unstable at body temperature, thus meaning that his melting point is below the normal temperature of a living being [34]. Why does Nature create such a protein?

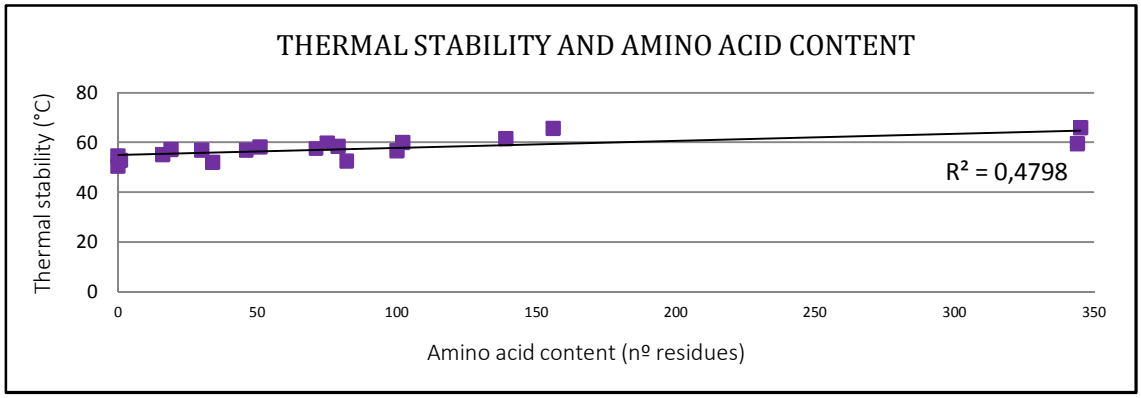
To answer this question, rather than ask us why it is not stable, we must ask us why it should be stable. There must be a reason why collagen melts at body temperature.

Studies have shown that collagen denatures completely within a couple of days at a temperature of 36 °C. Above this temperature the denaturation rate increases exponentially and melting occurs within hours [34]. When analysing the amino acid content and distribution of residues within the collagen molecule, it seems that the construction itself of collagen is rather impeccable.

Of the 20 proteinogenic amino acids that exist, 17 can be found at the Xaa and Yaa positions. Two residues that are not present in the triple helix, tryptophan and cysteine, have shown to be the less thermal stable amino acids in host-guest training sets, and specially cysteine interferes with assembly through spontaneous formation of disulphide bonds.

When analysing the other 17 amino acids it results that almost all of them are found in their most optimal position. Through a host-guest peptide training set [$(GPO)_5$ - GXaaYaa - $(GPO)_5$], thermal stability of each amino acid was evaluated in both positions, Xaa and Yaa. The frequency of each amino acid in both positions was also calculated. It resulted that 81 % of the residues that are found in their most common position in the Gly – Xaa – Yaa structure, also had the highest melting temperature at this position thus meaning that they have a rather stabilising than destabilising effect.

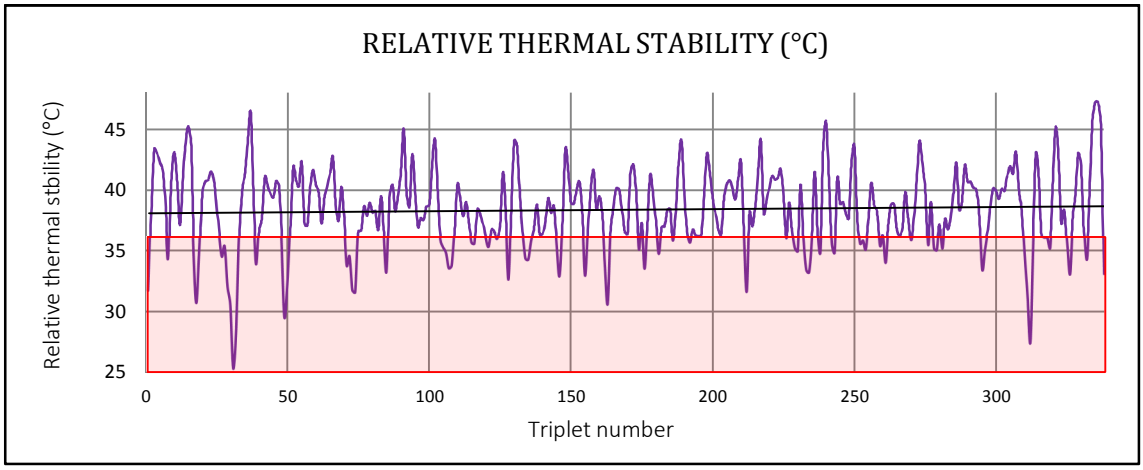
Using the data from the anterior training set another correlation was assessed. Using the melting temperature of each amino acid at their most common position, a relation was found with the amino acid content in the tropocollagen molecule. Residues that showed to have higher melting temperatures at host-guest peptides also appeared more often in the collagen molecule. As exposed in **Graph 7**, the correlation coefficient between the higher melting temperatures and the residue content resulted to be 0,7 , indicating that there is a relative strong relationship, although correlation does not always involve causation.



Graph 7 | Scatterplot showing the correlation between thermal stability of residues in host-guest training sets and its respective amount (nº of residues) in tropocollagen molecule.

The conclusion is that there is a strong positive correlation between the thermal stability of residues and their respective position or amount in the collagen molecule. Considering also the high amount of hydroxyproline with a stabilising effect, we can deduce that collagen should be a thermally stable protein. But it is not.

When observing the stability profile of the triple helix (**Graph 8**), there appear regions that are almost 10° C under normal body temperature [49].



Graph 8 | Line chart showing the relative thermal stability of each triplet (Gly – Xaa – Yaa) in the tropocollagen molecule. The regions found in the red box are the ones that are below body temperature.

This fact indicates that collagen is thermally unstable at body temperature with regions that are susceptible to unfold due to a low stability. So far, we can deduce that there must be a reason why collagen has a low melting temperature. As explained in the section above, the destruction of bone matrix is an indispensable process and this could be the cause of why collagen is intrinsically unstable.

The role of proteases and other factors contributing to collagen instability

During synthesis collagen is bound to different molecules that avoid its unfolding and melting until it is incorporated into ECM [120]. When found in the extracellular space, collagen is exposed to different enzymes and the heat from the physiological body temperature. This means that after secretion, collagen begins to unfold until it is incorporated into fibrils where single tropocollagen molecules become more stable [34].

The collagen melting temperature is so adjusted that initial folding and fibrillogenesis are possible, but when found in ECM collagen slowly begins to unfold in the regions with lesser stability until it adopts a conformation of a random coil rather than a triple helix. The average unfolding time at body temperature is several days, and a decrease of about 1°C in melting temperature reduces this time to several hours [34].

One of the reasons why collagen has to be unstable is that it is a substrate for different proteases. Proteases are enzymes that break down proteins, and as seen in the anterior chapter, the proteolysis of collagen is necessary for bone resorption. When collagen is found in fibres, proteases (collagenase, cathepsin K) are unable to destroy collagen molecules. However, when there are unfolded regions the enzymes can reach these sites and break down fibres from the ECM.

Supporting evidence for this postulate can be taken from alterations in collagen-specific peptidases. Cathepsin K is a cysteine proteinase that plays an important role in osteoclastic activity participating in the degradation of bone and cartilage [61]. Mutations that suppress cathepsin K activity result in a disorder known as pycnodysostosis (aka Toulouse-Lautrec syndrome), an osteochondrodysplasia characterized by a generalised osteosclerosis, short stature and delayed closure of skull sutures [24][42][104]. The lack of degradation of the ECM and the high bone density were directly correlated with the deficiency of cathepsin K, giving rise to the evidence that degrading molecules must be present for bone remodelling.

Another reason for the local thermal instability of collagen is bone flexibility. The organic part of bone makes the bone flexible giving it more strength against bending. The microunfolding of the instable regions probably confers this elasticity. If bone is too rigid with a high mineral content it would break more easily. On the contrary, if bone has moderate deformation ability, it will support better bending forces. This is the reason why children often suffer greenstick or torus fractures rather than complete fractures.

Implications for osteogenesis imperfecta

Mutant collagen chains with glycine substitutions or truncated molecules have a decreased thermal stability. This situation leads to an increased melting rate of collagen, reducing its lifetime to several hours. Apart from this, the misfolding of collagen molecules can expose critical regions to digestion by matrix proteinases like cathepsin K. The sum of the thermal instability plus the propensity of degradation by enzymes reduces considerably the organic part of bone matrix.

The higher degradation rates put the cell under pressure to increase protein synthesis. The result is a fast protein fabrication, probably leading to more ER stress and failing to create new collagen. This worsens the UPR, eventually leading to apoptosis. Osteoblasts secrete unfolded proteins into bone matrix giving rise to an irregular arrangement of molecules and thus leading to a defective mineral deposit.

Osteoblast differentiation alterations are also seen. In OI type I, the quantitative defective ECM fails to give normal feedback signals that promote cell differentiation [56]. The thermally unstable collagen has a faster turnover, giving rise to an alteration in the ECM ratio of organic and inorganic components. This alters considerably the matrix-osteoblast feedback, making that lesser immature cells differentiate into collagen-synthesising osteoblasts. This decrease in osteoblasts summed to the numerous cells that undergo apoptosis drops the overall content of collagen in the bone matrix.

3. 8. THE ROLE OF NON-COLLAGENOUS PROTEINS AND ENZYMES

Non-collagenous proteins (NCPs) are a group of organic molecules found in bone and dentin matrix that promote, control and regulate fibrillogenesis, crystal growth and mineralization during osteogenesis and dentinogenesis [30][33][52][61].

This group of proteins have a key role in the synthesis and maintenance of the ECM. Therefore, it was assumed that they have a considerable influence in the outcoming phenotype of osteogenesis imperfecta. The main NCPs are listed in **Table 9**. In recent years, it has been observed that alterations that abolish the function of these proteins result in phenotypes extremely similar to OI [109][120]. This fact supports the great influence they could have in the classical collagen-related types of OI.

Table 9 | The different non-collagenous proteins are listed and classified according to their corresponding superfamily.

Protein name	Gene	Related disorder	OMIM Entry
Chaperones			
Heat shock protein 47 [106]	SERPINH1	OI - X	*600943
Prolyl 3-hydroxylation			
Cartilage-associated protein [84]	CRTAP	OI - VII	*605497
Leprecan [88]	LEPRE1	OI - VIII	*610339
Peptidyl-prolyl isomerase [103]	PPIB	OI - IX	*123841
Prolyl 4-hydroxylation [28]			
Lysyl hydroxylation [118]			
Lysyl hydroxylase-1	PLOD1	EDS - VI	*153454
Lysyl hydroxylase-2	PLOD2	Bruck syndrome	*601865
Lysyl hydroxylase-3	PLOD3	CTD	*603066
Other NCP			
Interferon-induced transmembrane protein 5 [87]	IFITM5	OI - V	*614757
Serpin peptidase inhibitor, Clade F, Member 1 [105]	SERPINF1	OI - VI	*172860
FK506-binding protein 10 [86]	FKBP10	OI - XI	*607063
Dentin matrix protein 4	DMP4	Raine syndrome	*611061
SIBLING proteins [30][52][124]			
Osteopontin	SPP1	-	*166490
Bone sialoprotein 2	IBSP	-	*147563
Dentin matrix protein 1	DMP1	AR Rickets	*600980
Dentin sialophosphoprotein	DSPP	Deafness + DI	*125485
Matrix extracellular phosphoglycoprotein	MEPE	Osteomalacia	*605912

3. 8. 1. Collagen chaperoning: heat shock protein 47

Heat shock protein 47 (HSP47) is an ER-resident chaperone that assists triple helix folding and is induced by heat shock [70]. Unlike other chaperones, HSP47 binds to the fully folded collagen and therefore it is thought it has an unusual function since most of the chaperones bind to unfolded proteins [65].

HSP47 binds to all three collagen α-chains recognising the Gly – Pro – Arg sequence. The arginine residue at Yaa position is very important; without it the chaperone does not bind to collagen. The catalytic domain of HSP47 is composed by highly conserved aspartic acid residues that form salt bridges[19] and some hydrophobic interactions (**Figure 27**) [70].

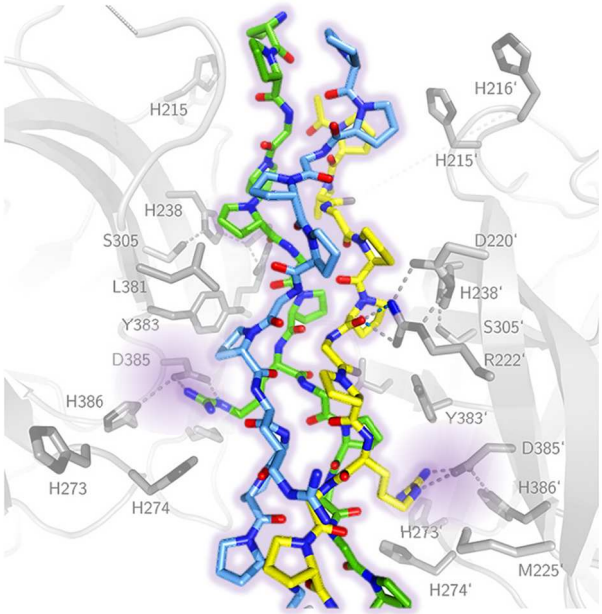


Figure 27 | Close-up view showing the catalytic domain of HSP47 (greyish colours) bound to the collagen triple helix (multicolour). The purple-shadowed areas point out the binding sites.

When collagen has finished folding the chaperone binds to the tropocollagen helix. Then, the complex HSP47-collagen travels to the Golgi apparatus where it is released [65][106]. This suggests that HSP47 has thermal-protective function avoiding the unfolding of the triple helix. However, when looking at the whole complex, another function can be attributed to the chaperone. The double-binding to one tropocollagen molecule has a protective effect against lateral aggregation with other collagen molecules.

[19] A salt bridge is a non-covalent bond, actually being a combination of electrostatic interactions (positive-negative charge) and a hydrogen bond.

The big globular shape of HSP47 avoids that the newly formed collagen triple helices come in contact and thus form a mesh-like aggregation (**Figure 28**).

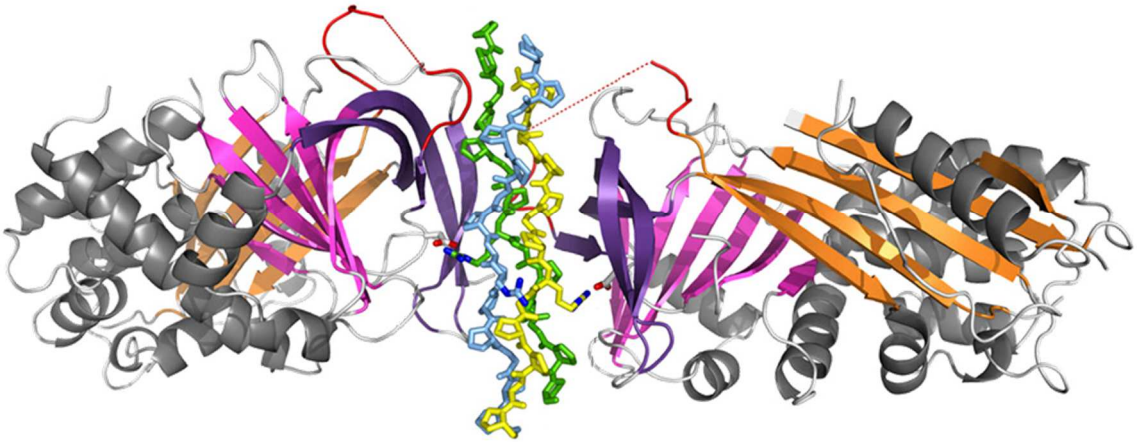


Figure 28 | Zoom out representation of the chaperone-substrate complex. The HSP47 protein is shown as cartoon and the collagen molecules as sticks. The proteins are artificial coloured. As can be observed, the globular shape of the chaperone has a protective effect against spontaneous lateral aggregation with other collagen molecules. Picture taken with PyMol (PDB file 4AU3).

The fact that the complex is released in the vesicular–tubular cluster (ER – Golgi intermediate compartment) is probably a secondary function of the chaperone related with the trafficking control at that boundary [70][106]. The release most likely occurs through a denaturation–renaturation phenomenon of HSP47 as a response against the pH change (the Golgi apparatus is slightly acidic while the ER has a neutral pH value).

The main function of HSP47, apart from avoiding clump formation, is to maintain collagen thermally stable until it is secreted [65][70]. The chaperone is induced by heath shock, and since collagen is thermally unstable at body temperature, HSP47 is permanently found in the ER lumen. However, a third function can be attributed to the chaperone. When looking at **Figure 29** it can be seen that thanks HSP47, collagen is soluble and can be transported freely through the cell. Without it collagen would be unstable since its hydrophobic structure is exposed to the environment.

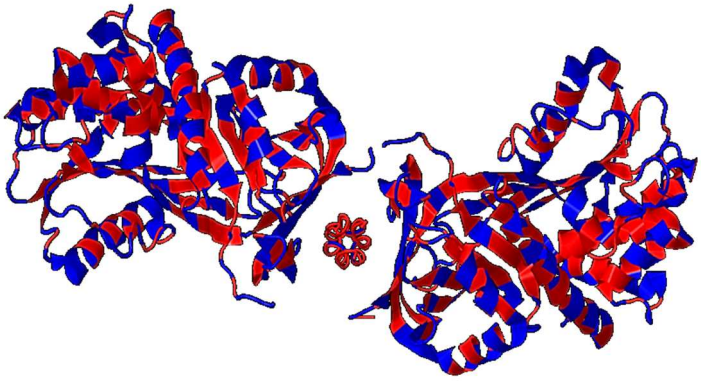


Figure 29 | Zoom out representation of the chaperone-substrate complex. Both proteins are shown as cartoon. Colour blue corresponds to hydrophilic regions and colour red to hydrophobic regions. HSP47 forms a hydrophilic complex with collagen, making it soluble and easy to transport through the cell. Picture taken with PyMol (PDB file 4AU3).

The protective function of this molecule is crucial. This fact can be supported by knockout mouse models for HSP47. These mice are homozygous for a knockout mutation and do not synthesise functional HSP47. This leads to a lethal outcome during embryonic development with very fragile bones and mineralisation defects [98][106].

Mutations in the homologous gene in humans lead to osteogenesis imperfecta type X with a severity comparable to OI-III in the Sillence classification. Bone histology in these patients shows an irregular arrangement of collagen matrix. As can be deduced, no overmodification is present and the amount of collagen molecules is nearly normal. However, the triple helices exhibit random coil-like conformations concordant with the lack of HSP47. The collagen is protease sensitive and the transit between ER and Golgi was accelerated, although overall transportation was delayed [15][98].

HSP47 can also influence the outcome of the classical OI types. The obligate Gly – Pro – Arg sequence, if altered, disturbs the binding of the chaperone. Consequently, collagen lacks its protective function and is exposed to thermal instability and lateral aggregation of molecules.

When looking at the amino acid sequence of both proα1 and proα2 there appear 12 putative binding sites for HSP47.

1. If the conformation is α1 · α2 · α1 and HSP47 binds to both α1 chains, 9 possible binding sites result[20].
2. If the conformation is α1 · α1 · α2 and HSP47 binds to one α1 and one α2 chain, there are only 3 possible overlapping binding sites[21] (1 N-terminal and 2 C-terminal).

Surprisingly, of the nine possible binding sites in the first conformation only three have reported glycine substitutions in the Gly – Pro – Arg sequence. One of them is an OI-IV phenotype caused by a glycine substitution (Gly^{40}) very N-terminal. The other two correspond to lethal glycine substitutions for cysteine and serine (Gly^{913} and Gly^{988}), two residues that are generally non-lethal. This gives rise to the idea that glycine substitutions in the putative HSP47 binding sites are almost always lethal during prenatal period, resulting in this lack of reported mutations.

In the second conformation the situation is similar. Of three possible binding sites, only two have reported mutations. These two mutations are glycine substitutions for serine resulting in the very severe OI II/III type.

[20] Glycines at helix positions 7, 40, 124, 181, 517, 730, 913, 961 and 988.
[21] Glycines at helix positions 7, 730 and 961.

In aberrant collagen HSP47 is induced constantly since mutant chains have a lower thermal instability than normal ones. The delayed helix folding avoids the binding of HSP47, leading to a delayed secretion and accumulation in the ER since the chaperone regulates the transit from the ER to the Golgi apparatus. Moreover, if the chaperone cannot bind due to an interruption in the triple helix, the outcome is deleterious (**Figure 30**). The collagen chains, apart from harbouring a disruption in the helix, they also lack the protective effect of HSP47. This leads to a total conformational alteration of the helix with its corresponding consequences.

There is another possible situation in which the chaperone could not bind to collagen. As seen in anterior chapters, there are glycine substitutions in which renucleation can occur and other ones that the helix remains open (**Figure 30**). In the latter, if binding sites for HSP47 are located in the open region, the chaperone will not be able to bind. This will also lead to a lack of the protective effect of HSP47. The result will be more harmful than if the chain can renucleate.

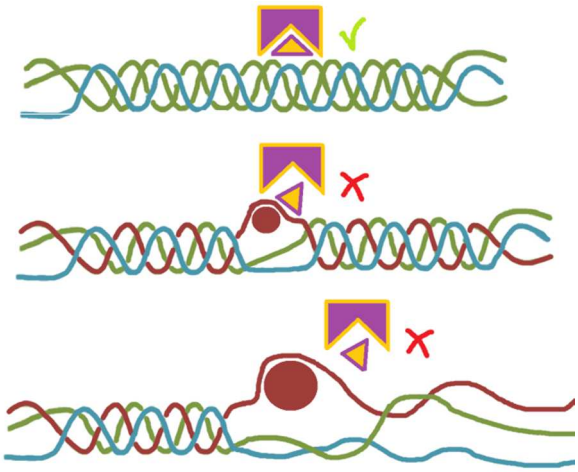


Figure 30 | Representative picture of the different situations in which the HSP47 chaperone cannot bind to the collagen triple helix.

It is clear that HSP47 has a crucial role maintaining the integrity of the triple helix until it is released into the Golgi apparatus. However, it remains unclear in which conformation HSP47 binds to the newly synthesised collagen chains. The chaperone also worsens the UPR making that collagen aggregates accumulate in the ER. There are a lot of variables that must be studied but in the initial stadiums of collagen synthesis, HSP47 has a key role.

All this is supported by the fact that OI type X is caused by homozygous or compound heterozygous mutation in the HSP47 gene [15]. This very severe deforming type of osteogenesis imperfecta was first described in 2010 in a child from a consanguineous Saudi Arabian couple. This gave rise to the autosomal recessive inheritance of this disorder [15][98].

3. 8. 2. Prolyl hydroxylation

The high amount of hydroxyproline in collagen is thought to be due to a stabilizing effect of this non-proteinogenic amino acid [60]. Peptides containing hydroxyproline at different triplet positions are shown in **Table 10**. This host-guest training set simulates the melting temperature (T_m) of peptides containing eleven Gly – Xaa – Yaa triplets.

Table 10 | Training set with different host-guest peptides. The hydroxyproline residue is placed in different positions in the Gly – Xaa – Yaa structure in order to assess its stabilizing effect through the resulting melting temperature. As guest peptides, triplets with amino acids have been taken that show a des/stabilizing effect on triple helix.

Peptide	Predicted melting temperature (T_m)
$(Gly – Pro – Hyp)_{11}$	65,8 °C
$(Gly – Hyp – Hyp)_{11}$	65,0 °C [60]
$(Gly – Hyp – Pro)_{11}$	No helix [60]
$(Gly – Pro – Pro)_{11}$	49,6 °C
$(Gly – Pro – \mathbf{Hyp})_5 – Gly– Glu – Hyp – (Gly – Pro – \mathbf{Hyp})_5$	61,4 °C
$(Gly – Pro – \mathbf{Pro})_5 – Gly– Glu – Hyp – (Gly – Pro – \mathbf{Pro})_5$	45,2 °C
$(Gly – Pro – \mathbf{Hyp})_5 – Gly– Ser – Hyp – (Gly – Pro – \mathbf{Hyp})_5$	56.5 °C
$(Gly – Pro – \mathbf{Pro})_5 – Gly– Ser – Hyp – (Gly – Pro – \mathbf{Pro})_5$	40.3 °C
$(Gly – Pro – \mathbf{Hyp})_5 – Gly–Trp – Hyp – (Gly – Pro – \mathbf{Hyp})_5$	50.4 °C
$(Gly – Pro – \mathbf{Pro})_5 – Gly– Trp – Hyp – (Gly – Pro – \mathbf{Pro})_5$	34.2 °C

The peptide with a GPO sequence has resulted to be the most thermal stable structure. The hydroxyproline residue at the Yaa positon has a stabilising effect whereas the Hyp residue at Xaa position does not form helix at body temperature. When placed at both positions it even decreases thermal stability. These results support the fact that the Hyp residues only have a stabilising effect when found at Xaa position.

If we look at the host-guest peptides, we can observe that the stabilising effect of hydroxyproline is not dependent on the neighbouring residues. When an amino acid residue is introduced that has shown to stabilise triple helix (Glu), the melting temperature increases in 16,2 °C, the same rise as with a destabilising amino acid (Trp). However, this fact can vary since in the tropocollagen molecule there are a lot of different adjacent residues that can have an influence on the thermal stability.

The presence of the stabilising $(PGPO)_n$ sequence is also important for the renucleation after a folding interruption. Glycine substitutions can delay chain assembling making that the triple helix does not form correctly and has a bunch. It has been observed that renucleation depends on a high stabilising sequence N – terminal to the mutation site [31].

Prolyl 4-hydroxylation

Prolyl 4-hydroxylation (P4H) is a major posttranslational modification in collagen. In this process an enzyme called prolyl 4-hydroxylase adds a –OH group onto the fourth carbon of prolines in the Yaa position (**Figure 31**) [119]. As seen above, this raises the thermal stability of collagen and is indispensable for the correct synthesis of this structural protein.

Prolyl 4-hydroxylase is composed of four subunits: 2 alpha subunits and 2 two identical beta subunits[22]. This enzyme complex catalyses the formation of hydroxyl groups and needs as cofactor ascorbic acid (vitamin C) [116].

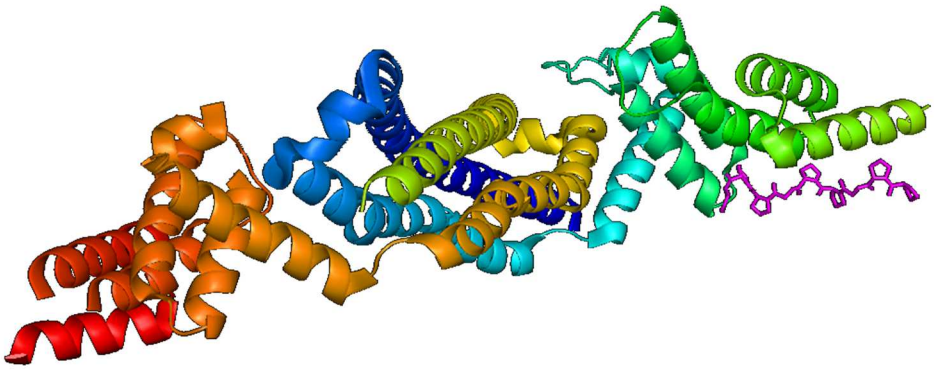


Figure 31 | Crystal structure of the catalytic domain of P4H. The enzyme is artificial coloured and shown as cartoon. The collagen chain (magenta) is shown as sticks. Only the alpha subunits of P4H are represented. Picture taken with PyMol (PDB file 4BTA).

Unexpectedly, the lack of hydroxylation does not affect bone tissue [28]. Absence of hydroxylation is seen in ascorbate-deficient patients and causes a disease called scurvy. The main symptoms of scurvy are lethargy, poor wound healing and tendency to haemorrhages as a consequence of tissue fragility [116]. This is supported by the P4H1 knockout mouse model. Null mice harbouring homozygous mutations for this gene showed abnormal assembly of collagen type IV, but fibril-forming collagens type I and III do not showed significant alterations and they were secreted efficiently [28].

[22] The corresponding encoding genes are P4HA1 and P4HA2 for the two alpha subunits and P4HB for the beta subunit.

Prolyl 3-hydroxylation

Prolyl 3-hydroxylation (P3H) is a minor and highly conserved post-translational modification in collagen type I. This enzyme, called prolyl 3-hydroxylase, catalyses the formation of one –OH group on the third carbon of Pro986 (α1) [29][39][1119].

Prolyl 3-hydroxylase is a heterotrimer composed by three subunits: cartilage-associated protein (CRTAP)[84], leucine- and proline-enriched proteoglycan (leprecan; LEPRE1)[88] and peptidyl-prolyl isomerase B (cyclophilin B; PPIB)[103][39].

In contrast to prolyl 4-hydroxylase, mutations in prolyl 3-hydroxylase encoding genes have devastating effects on bone. Homozygous or compound heterozygous mutations in each of the subunits lead to severe OI phenotypes with extreme bone fragility [11].

1. *CRTAP*$^{-/-}$ mutations: OI type VII (moderate)[95]
2. *Leprecan*$^{-/-}$ mutations: OI type VIII (severe/lethal)[96]
3. *Cyclophilin B*$^{-/-}$ mutations: OI type IX (severe/lethal)[97]

Knockout mouse models are also available for these OI types. Null mice for each of the three subunits lead to osteochondrodysplasia combined with low bone mass. Protein studies reveal decrease or absence of prolyl 3- hydroxylation [11][39].

It is not clear how the lack of hydroxylation of a unique residue can lead to such consequences. There are some hypotheses that Pro985 could help chain recognition, chain alignment or be a binding site for a specific enzyme or chaperone (Silverman, personal communication). It is known that the catalytic site resides in leprecan and thus its absence results in a more severe phenotype than the absence of the other subunits.

There is no glycine substitution reported in the Gly–Xaa–Yaa sequence containing Pro986. However, glycine substitutions one triplet N-terminal and one triplet C-terminal show lethal outcomes, giving rise to the assumption that they could have an influence on P3H. Further investigation must be done in order to assess more specifically the function of prolyl 3-hydroxylase.

3. 8. 3. Lysyl hydroxylation

Collagen lysyl hydroxylation is a major posttranslational modification consisting in the attachment of hydroxyl groups onto selected[23] lysines. This reaction is catalysed by an enzyme called lysyl hydroxylase and is composed by three subunits[24]. The hydroxylation of some lysines is essential for the posterior intra- and interchain crosslinking since the hydroxyl groups are needed to attach galactose and glucose/galactose saccharides [116][118].

Mutations in the training set were evaluated in order to assess the role of lysyl hydroxylation sites. One glycine substitution downstream and one substitution upstream for a lysine hydroxylation site were studied.
There was no case in which a special outcome has been observed. In all glycine substitutions mild to moderate outcomes are obtained. This fact does not support the role of lysyl hydroxylase in the phenotype modulation in OI. The reason is probably that glycine substitutions do not interfere in hydroxylation.

This unexpected result can be interpreted in a different way: the presence of a near crosslink stabilises the chain and probably lessens the severity of the glycine substitution. This can be supported with the fact that near the crosslink sites, high-destabilising glycine substitutions (for valine or aspartic acid) result in mild phenotypes.
The fact that LH is not exclusively a collagen enzyme also gives rise to the idea that lysyl hydroxylation is not essential for collagen therefore explaining the mixed phenotypes.

On the contrary, mutations that abolish the catalytic activity of lysyl hydroxylase result in a variety of different connective tissue disorders. Null mutations in LH1 (PLOD1) result in Ehlers-Danlos syndrome type VI (minimal bone involvement). Homozygous mutations of LH2 (PLOD2) result in Bruck syndrome, an OI-like disorder characterised by extreme bone fragility and congenital joint contractures. Lastly, there has been reported one mutation in LH3 (PLOD3). This caused a mixed CTD with moderate bone involvement.

[23] Lysines with triple helical positions K^{87} and K^{174} in alpha-1 and positions K^{87}, K^{174} and K^{219} in alpha-2 are hydroxylated.

[24] The subunits forming lysyl hydroxylase are
procollagen-lysine, 2-oxoglutarate 5-dyoxigenase 1 (PLOD1, OMIM entry *153454),
procollagen-lysine, 2-oxoglutarate 5-dyoxigenase 2 (PLOD2, OMIM entry *601865),
procollagen-lysine, 2-oxoglutarate 5-dyoxigenase 3 (PLOD3, OMIM entry *603066).

3. 8. 4. SIBLING proteins

SIBLING proteins (small integrin-binding ligand, N-linked glycoprotein) are a family of non-collagenous proteins forming part of the extracellular matrix of bone and dentin [52]. This group of proteins have a key role in matrix mineralisation and they are all located on human chromosome 4q21-23. The proteins forming part of the SIBLING family are shown in **Table 11**.

Table 11 | Members of the SIBLING protein family. All these proteins share a similar structure and undergo the same post-translational modifications.

SIBLING proteins	
Osteopontin	SPP1
Bone sialoprotein 2	IBSP
Dentin matrix protein 1	DMP1
Dentin sialophosphoprotein	DSPP
Matrix extracellular phosphoglycoprotein	MEPE

The exact role of the SIBLINGs in bone and dentin mineralisation is not yet well understood [65]. Herein, the implications of these proteins in OI and dentinogenesis imperfecta will be discussed. The evidence of their effect on these two conditions can be supported by the following facts:

1. They are only expressed in bone and dentin, main structures affected in OI [22][30][52][61].
2. Knockout mouse models for some of these proteins show alterations in the mineralisation process of bone and dentin, giving rise to the idea that they could influence the resulting OI phenotype.

These proteins regulate the crystal deposit, arrangement and size during mineralisation [61]. If the substrate, the collagen network, is defective as in osteogenesis imperfecta, most probably there will be an alteration in mineralisation. The SIBLING proteins will have troubles with carrying out their job properly affecting the overall bone strength. The impossibility of the proteins to carry out its function will be compared to the absence of the protein (knockout) since the resulting effect is almost identic.

Three of the SIBLING proteins (osteopontin, bone sialoprotein 2 and matrix extracellular phosphoglycoprotein) do not show significant impairment of osteogenesis and dentinogenesis in knockout mouse models. The lack of both alleles results in a slight defective mineralisation showing an increased mineral content and high trabecular (spongy) bone mass [22]. This suggests that they do not have a significant role, if at all, on the severity of OI.

The other two SIBLING proteins (dentin matrix protein 1 and dentin sialophosphoprotein) do have related disorders in total knockout models [13][18].

Dentin matrix protein 1

The first one, dentin matrix protein 1, causes autosomal recessive hypophosphataemic rickets when absent. The murine model *Dmp1*$^{-/-}$ shows osteomalacia and a mineralisation defect [18]. DMP1 most probably participates in phosphate homeostasis [48] making that sufficient levels of phosphate are present to form apatite crystals. The binding site of DMP1 on the collagen molecule mapped to the N-propeptide (very N-terminal) region. There is also evidence that DMP1 accelerates fibril formation *in vitro* [16][27].

When looking at the mutations training set it appears that no mutations in the N-propeptide are reported. At both sites of the amino acids corresponding to the N-propeptide of both alpha chains there is a large gap comprising the upstream and downstream amino acids where there are no mutations. This fact can be due to either an early lethal outcome or the no-pathogenicity of these mutations. Since DMP1 binds to this region, it is probably that these mutations are lethal during embryonic development although there is not enough evidence to support this proposal.

Dentin sialophosphoprotein

Dentin sialophosphoprotein (DSPP) is only expressed by odontoblasts. This means that they participate only in the mineralisation process of teeth and not in bone [52]. Mutations in this gene in humans, as well as knockout models, show a dentinogenesis imperfecta phenotype (**Figure 32**) associated in some cases with deafness [71]. This raises the assumption that DSPP has a role in osteogenesis imperfecta since the latter is usually accompanied with DI.

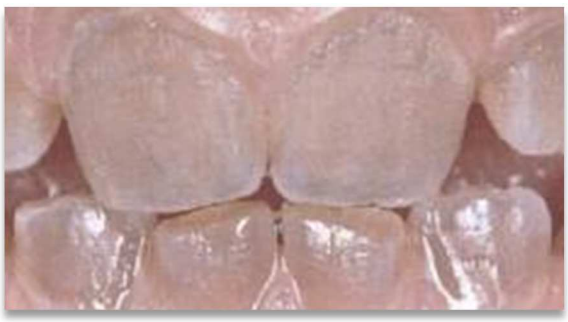


Figure 32 | Photography showing a patient affected by dentinogenesis imperfecta. The teeth appear opalescent due to dentin dysplasia. DI causes the teeth to break very easily leading to a fast wearing and premature tooth loss.

Dentin matrix protein is cleaved into three protein forms: the N-terminal for dentin sialoprotein, the middle for dentin glycoprotein and the C-terminal for phosphophoryn. The latter has a key role in dentin mineralisation. The ligand binding site for phosphophoryn has already been determined (for more details see **Figure 22,** page 46). To assess if ligand binding sites for phosphophoryn have a relationship with the comorbidity of dentinogenesis imperfecta in OI, data was taken from Rauch et al. (2010) [53]. They reported a set of patients with mutations in the triple helical domain describing also the presence or absence of DI (**Figure 33**).

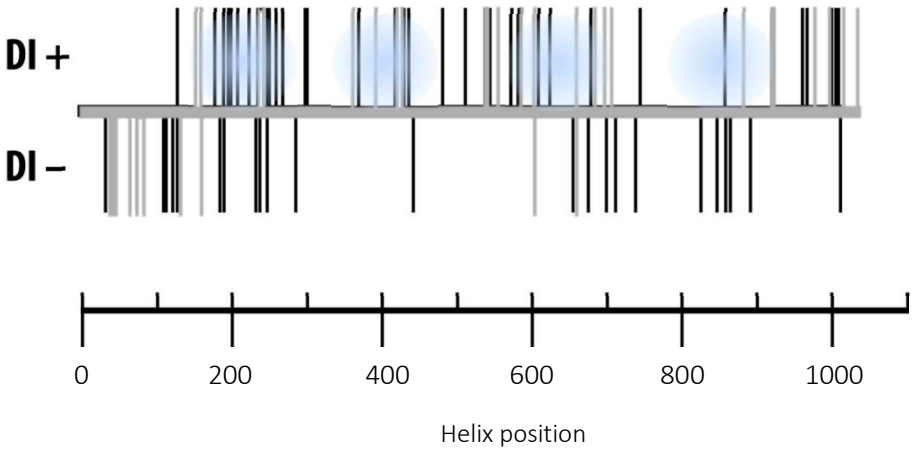


Figure 33 | Schematic representation of the mutations associated with DI along the collagen chain. Presence of dentinogenesis imperfecta (DI +) and absence (DI -). The grey lines represent the alpha-1 chain and the black lines represent the alpha-2 chain. The blue-shadowed circles denote the ligand binding sites for phosphophoryn (see **Figure 22** in page 46 for more information).

As can be observed in the figure above, in three of four binding sites for phosphophoryn there accumulate mutations associated with DI. The fourth binding site region, about residue 850, does not show a clear relationship; it is indeed the contrary, since we can see that there are more mutations without DI. This gives rise to the idea that there can be a relationship but additional investigation must be done, optimally with a larger number of mutations.

However, there could not be found a relationship with the presence or absence of deafness. The lack of data and also the difficulty of this task makes impossible to find a relationship. A revision of the literature also resulted in a negative outcome when searching a correlation between hearing loss and the genotype [63]. The reason for this inconsistency could be the combination of conductive and sensorineural hearing loss seen in these patients [56], without following any regular pattern or concrete severity expression.

3. 9. INTRODUCTION INTO THE THERAPY OPTIONS FOR OI

Understanding the genetic background of osteogenesis imperfecta is crucial if we want to develop gene or cell therapy in a relative near future.

The nature of OI makes that develop a definite gene therapy is very difficult. The main reasons for this difficulty are:

1. **Osteogenesis imperfecta is a multigenic disease**. There are more than twelve genes involved in the different types of OI making that for each subtype a different therapy must be designed.

2. **Different inheritance patterns**. Some subtypes of OI are dominantly inherited while others are recessively inherited. This implies that different therapies have to be considered since the expression pathways are different in dominant or recessive disorders.

3. **Heterogeneity in mutations types**. There are a lot of different mutations that can lead to OI. These mutations account compound heterozygous, homozygous, haploinsufficiency and dominant negative mutations. This makes more difficult to find a unifying therapy since each of them has a different underlying pathogenesis.

4. **Dominant negative effect**. Most of the mutations in collagen-related OI types are dominant negative, meaning that the aberrant product interferes with the product of the normal allele. In these cases it is not enough with supplying the functional allele since the aberrant one must also be silenced.

All these factors contribute to the great difficulty to propose an effective cell or gene therapy. Nonetheless, some effort has been made with a variety of results. The following sections are a little introduction into the actual research stand of OI therapy and which possibilities are in a potential proximity for successful cell and gene therapy.

Besides, it is worthwhile mentioning that due to all these anteriorly cited factors OI is a good model for other connective tissue disorders. Ehlers-Danlos syndrome, which is also a multigenic and greatly heterogeneous group of disorders, is a good candidate that can take advantage and benefit of gene therapy models for OI. The underlying pathogenesis is practically identical to that causing OI and if we manage to treat successfully patients with OI, it will be a great progress also for EDS patients.

3. 10. MESENCHYMAL STEM CELL TRANSPLANTATION

To perform bone marrow mesenchymal stem cell transplantation there exist two different possibilities: allotransplantation[25] and autotransplantation combined with gene therapy [41][44]. The first will be discussed herein while the second in the following section.

Osteoblasts, the collagen-synthesising cells, are differentiation products of mesenchymal stem cells (MSC). MSCs are found, among other places, in bone marrow and thus are candidates to be transplanted in OI patients since they can differentiate into collagen-synthesising cells [19][35].

The first condition to conduct an allogeneic transplantation is to find a compatible donor. Reasonably, the donor can neither harbour any mutation for OI nor any genetic mutation that could have any harm on the receptor. If the HLA-compatibility is assured, the obtained MSC are introduced into blood stream. Then, theoretically, the MSC are incorporated into bone and differentiate into osteoblasts which synthesise and deposit collagen into ECM[35].

Support for this theory can be obtained from trials on murine models. In an attempt to verify the effectiveness and potential of MSC transplantation in OI, osteogenic cells marked with green fluorescent protein (GFP) were transplanted into developing OI mouse models. Histological analysis revealed that the transplanted cells were incorporated into bone and secreted efficiently collagen type I. This study supports the fact that MSCs can engraft into the targeted tissue and are capable of forming bone *in vivo,* as represented in **Figure 34**[35].

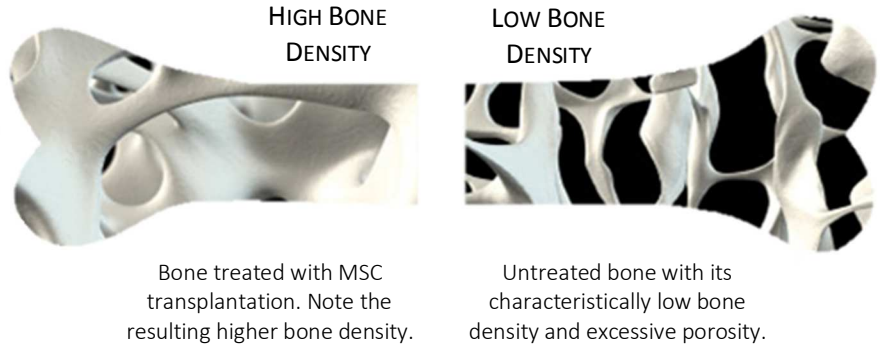


Figure 34 | Representative image of a MSC transplantation in OI. As can be seen on the left side of the figure, the transplanted osteoblasts have been incorporated into bone and secrete efficiently normal collagen type I thus increasing the bone density (less black holes). On the right side, we can observe the low bone density caused by a defective formation of collagen type I produced by osteoblasts harbouring a mutation leading to OI (more black holes). The lower bone density due to defective collagen weakens the whole bone scaffold leading to higher fracture susceptibility.

[25] Allotransplantation is the transplantation of cells, tissues or organs to a recipient from a genetically non-identical donor of the same species.

Although MSC transplantation can be a viable treatment option in OI, trials on humans have shown mixed outcomes. Fracture rate was decreased and total bone mass was increased but long-term effects were not persistently observable [69].
To improve these discouraging results, MSC treatment should be applied on developing foetus since the incorporation of the transplanted cells into bone marrow is easier in these stages.

Bone formation begins with intramembranous ossification[26] of skull and vertebrae. Then, long bones begin to form by endochondral ossification[27] approximately at 10 – 12 weeks gestational age [121]. At this point the skeleton begins to form, meaning that it is the most adequate period to perform MSC transplantation on a developing foetus with OI.

Nonetheless, this implies prenatal diagnosis of OI which is usually only possible in the most severe types (OI-II and OI-III) and from second trimester (approx. 14 weeks gestational age).
If either family history of OI is present or previous OI pregnancies, gene testing can be done through amniocentesis or chorionic villus sampling (the latter permits earlier testing option).

Furthermore, there is another disadvantage related to MSC transplantation during development. This procedure is complicated to perform since it must be conducted *in utero*. Although nowadays *in utero* procedures are becoming more common there are still considerable risks associated to these procedures.

Even though MSC transplantation can be an option for treating OI, there are some difficulties like the obligate prenatal diagnosis or the risk associated to the procedure. Nevertheless, in OI type II this treatment could eradicate the perinatal lethality of this subtype and contribute to a better prognosis for these patients. There is still the option to undergo MSC transplantation in early childhood but poorer results must be expected [19][35].

[26] Intramembranous ossification is the direct laying down of bone into the primitive connective tissue, the mesenchyme. It occurs on skull, clavicles, mandible and vertebrae (axial skeleton)

[27] Endochondral ossification involves a hyaline cartilage template for bone formation. This type of ossification occurs on long bones of extremities (distal bones) and lasts until skeletal maturation on adult individuals.

3. 11. GENE THERAPY

The ultimate treatment for inherited disorders is gene therapy. The aim of gene therapy is to correct the underlying defect causing the disease: the mutation.

In recent years, numerous trials for gene therapy have been conducted. This procedure involves the transference of a nucleotide sequence (e.g. a gene) into a targeted cell [45][73].

Gene therapy can be done in two different ways:

1. *Ex vivo*: cells are obtained from the patient and the vector delivers the gene direct on these selected cells, but outside the body [45].
2. *In vivo*: the vector is introduced into the body through different administration ways (intramuscular, intravenous) and delivers the gene into different cells [45].

The efficacy of gene therapy relies on the type of vector used and the response of the organism against the introduced gene. In the following sections it will be discussed which are the best strategies and tools to use for gene therapy for OI, considering all the factors mentioned at the beginning of this section.

3. 11. 1. Basic milestones of gene therapy in OI

1 Depending on the resulting effect produced onto the cell we can divide gene therapy into three groups. In osteogenesis imperfecta all three types are viable. Nonetheless, the decision whether using one or the other depends on the underlying genetic defect. The different combinations are shown in **Table 12**.

Table 12 | The different combinations of gene therapy according to the OI type are exposed.

	MUTATION CORRECTION Genetic defect is directly corrected.	GENE SUPPLY Healthy gene copy is supplied.	EXPRESSION INHIBITION Aberrant product is inhibited and silenced.
HAPLOINSUFFICIENCY (OI TYPE I)	The defect is fully corrected, normal status is achieved.	Null allele is neutralised; normal levels of collagen are synthesised.	Not applicable; would cause homozygous null allele and worsen the situation.
NEGATIVE DOMINANCY (OI TYPES II, III AND IV)	The defect is fully corrected, normal status is achieved.	Could compensate the qualitative defect, but ineffective if not combined with expression inhibition.	If conducted alone, converts the disease into OI type I. If combined with gene supply, normal status is achieved.

In OI type I the underlying defect is a null allele. In this case, it is possible to conduct a mutation correction or gene supply. In the other OI types (II, III and IV) it is not as simple; it must be decided if a fully correction is conducted or only a conversion to OI type I.

The decision whether an *in vivo* or an *ex vivo* gene therapy is done depends on the therapeutic effect that want to be achieved.

2 *Ex vivo* gene therapy requires the extraction of the target cells and its cultivation *in vitro*. The problem relies on the fact that the target cells, osteoblasts, do not proliferate since they do not undergo mitosis. Moreover, most vectors only can inject the gene into dividing cells [73].

The outstanding possibility is using induced pluripotent stem cells (iPSC) (**Figure 35**). Somatic cells are extracted from the patient and induced to pluripotency. Gene therapy is done and the cells proliferate. When enough cells have divided, then differentiation into osteogenic cells (osteoblast precursors) is induced. Finally, the obtained cells are transplanted into the patient[19]. This procedure has the advantage of being completely histocompatible with the patient in addition to the collection of a large number of cells. Nonetheless, there are some disadvantages already seen in the anterior chapter.

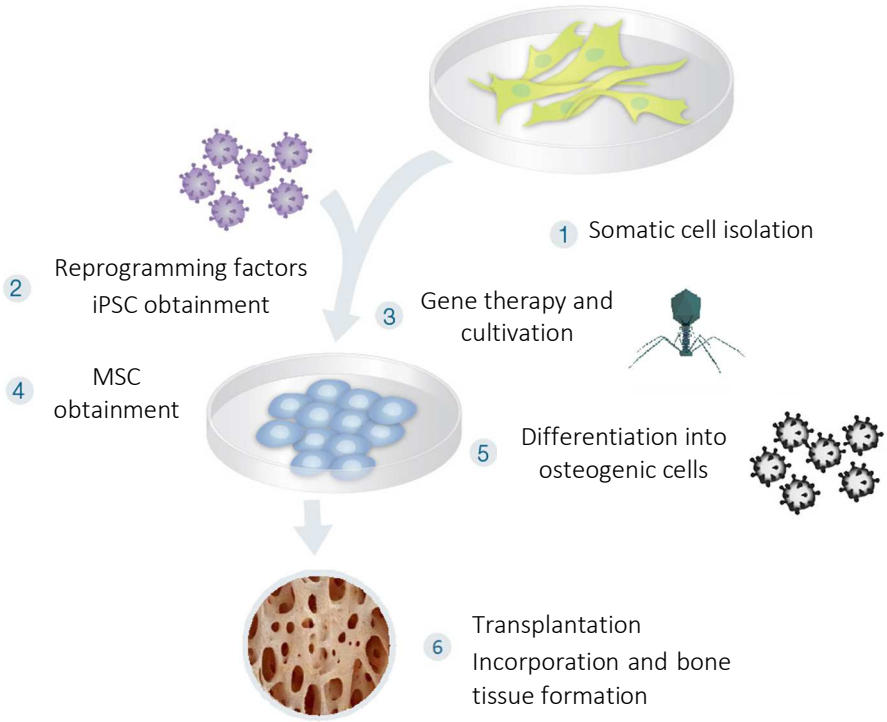


Figure 35 | Schematic representation of the procedure using induced pluripotent stem cells combined with gene therapy.

3 *In vivo* gene therapy requires a vector that actuates on the target cells inside the organism. Many different vectors have been used for gene therapy but all of them have pros and cons. Several difficulties have arisen from vectors used in this procedure and some of them can have harmful effects. A perfect vector is the one that delivers the gene in the correct place, does not cause immune reaction and does not induce oncogenesis (cancer) through insertional mutagenesis (for example, introducing the gene into a tumour suppressor gene)[45][75].

4 Many different vectors have been used for gene therapy but all of them have advantages and disadvantages. The following exposed viruses are possible candidates to be used in OI:

1. **Retroviruses**. This family of viruses have the ability to integrate their genes into the genome of the target cell. They possess the enzyme reverse transcriptase, able to synthesise a double-stranded DNA sequence from a single-stranded RNA template. The subfamily of lentivirus (for example, HIV) is able to insert the gene into non-dividing cells and thus a candidate as vector for gene therapy in OI (**Figure 36**)[45][75].

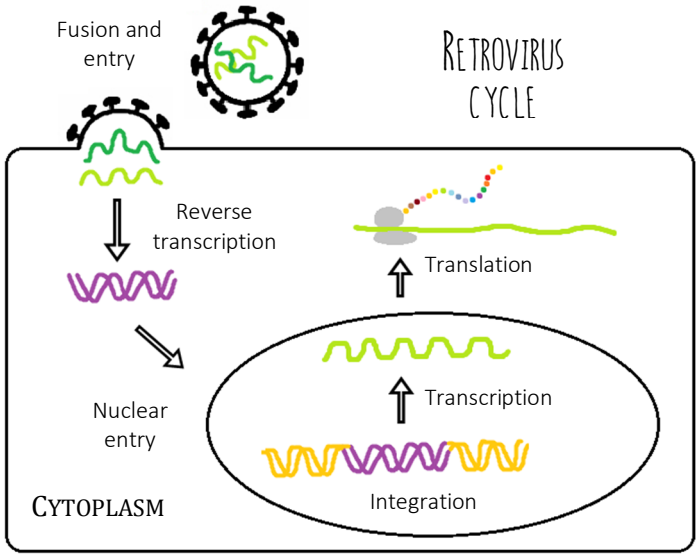


Figure 36 | Schematic representation of the replicative cycle of a retrovirus (e.g. lentivirus). The virus inserts its RNA strands into the target cell and reverse transcription occurs in the cytoplasm. Then, the double-stranded DNA sequences are transported into the cell nucleus where the genes incorporate into the genome. The inserted genes are transcribed and translated using the cell machinery.

2. **Adenoviruses**. This family of viruses have double-stranded DNA genes but do not incorporate their genome into the cell, thus having limited long-term effects on cells. The last generation of replication-defective adenoviruses evolved to the *gutless* adenoviruses, in which the whole viral genome has been deleted and the virus only contains the therapeutic gene in order to avoid the conversion to replication-competent viruses [45][75].

3. **Adeno-associated viruses**: adeno-associated viruses (AAV) belong to the dependovirus genus of the parvovirus family. In recent years this family of viruses have gained much popularity in gene therapy strategies.

   AAV are non-pathogenic viruses that infect human cells. Its genome consists of a single-stranded DNA chain. These viruses, to complete its replicative cycle (**Figure 37**), need certain *helper* functions which are provided by another virus, usually an adenovirus that co-infects the cell (from that comes the term "adeno-associated") [45][75].

   Replication-competent AAVs are capable of integrate their genes into a specific locus in the host cell (19q13). Nonetheless, this ability is lost in recombinant and replication-defective AVVs making that the integration occurs in a random locus as in retroviruses [45].

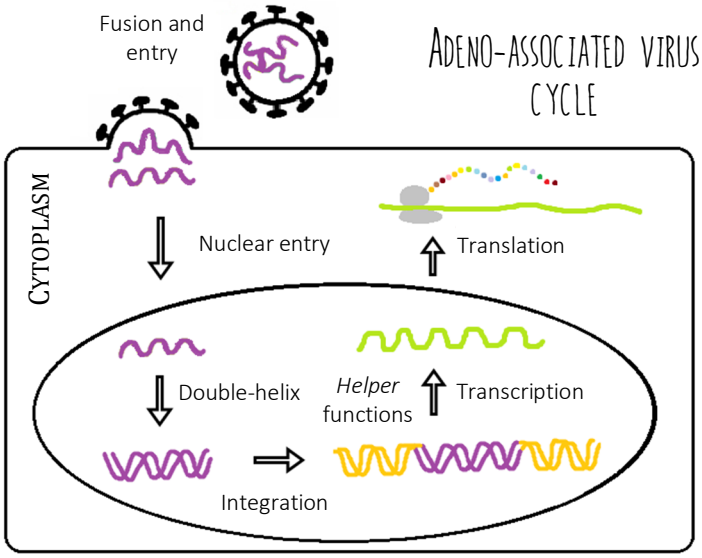


Figure 37 | Schematic representation of the replicative cycle of adeno-associated virus. The virus inserts its single-stranded DNA into the cell. Then, the DNA fragment enters the nucleus and the complementary strand is synthesised. Replicative-competent viruses integrate the gene into a specific locus (19q13), while replicative-defective viruses make it in a random locus.

The integrated virus genes remain in a latent status until the *helper* functions are provided by another virus (e. g. adenovirus). If this occurs, the inserted genes are transcribed and translated using the cell machinery.

3. 11. 2. Suitable therapies for osteogenesis imperfecta

The following sections expose appropriate therapies that can be functional in OI according to the underlying genetic defect. Although they are thought to be used *in vivo* so that a definite cure can be achieved they can also be applied when correcting the gene defect *ex vivo* in autotransplantation of iPSCs.

Osteogenesis imperfecta type I (premature stop codons and frameshifts)

OI type I is probably the easiest type to apply gene therapy. OI-I is caused by a null allele of either COL1A1 or COL1A2, meaning that only half gene product is synthesised. Therefore, it would be enough with supplying the cells with a normal copy of the missing gene.

A transgenic, replication-defective virus must be constructed that has incorporated the selected transgene. For *in vivo* therapy the most effective and safe option is the use of an adeno-associated virus since a long-term effect want to be achieved (retroviruses are preferred to be used in *ex vivo* therapies).
These AAVs will deliver the gene into osteoblasts providing them a normal copy of the gene (COL1A1 or COL1A2). Osteoblasts will produce the normal amount of collagen, thus improving or fully correcting the matrix insufficiency (**Figure 38**).

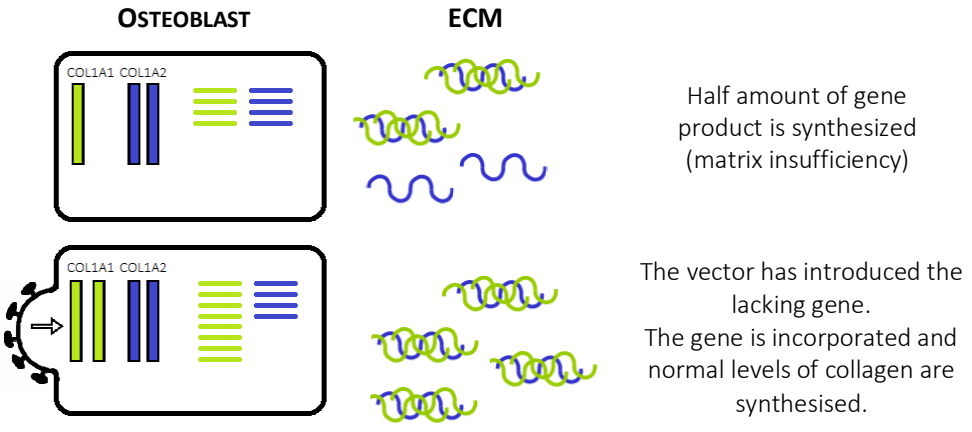


Figure 38 | Schematic representation of gene therapy in cases of null allele. The vector introduces the lacking gene into the cell, which will incorporate the gene into its genome. Using the cell machinery, the transgene will be expressed and therefore normal protein levels will be achieved.

Nevertheless, this technique could also be used in the non-classical OI types caused by homozygous null alleles. If the disease-causing mutation is identified, the correct gene is incorporated into the virus and delivered into the cells through the same mechanism as described above.

Osteogenesis imperfecta types I, II, III and IV (glycine substitutions and splice site mutations)

Glycine substitutions are caused by single nonsynonymous nucleotide substitutions in the glycine-encoding codons. These types of mutations are dominant-negative natured, meaning that the α-chain containing the glycine substitution has a harmful effect onto the normal synthesised α-chains.

To conduct gene therapy for these OI types there are two options: direct correction of the mutation or silencing the aberrant product and convert the disease into an OI-I phenotype.

The first option, targeted gene correction, is less developed and shows very poor results. This method uses chimeraplasts, hybrid sequences of DNA flanked at both terminal ends with ribonucleotides. The chimeraplast is complementary to the sequence that contains the mutation except at the middle of the short DNA sequence. Then, repair enzymes replace the cell's DNA sequence with that of the chimeraplast. This leaves the chimeraplast's correct sequence in the cell's DNA resulting in the correction of the mutation[45].

While this method sounds very ingenious and promising, it has become controversial and very poor results have been obtained. Therefore, it is not a viable option to use in OI unless it shows better results in the future.

The second option is based on the silencing of the aberrant product using antisense oligonucleotides. There are different methods that can be used with antisense nucleotides:

1. **Direct binding to the DNA**. The antisense oligonucleotides can inhibit the gene expression by direct binding to the DNA double helix through Hoogsteen base pairing, giving rise to a DNA triple helix or quadruple helix (**Figure 39**) that cannot be transcribed and thus is not expressed[45].

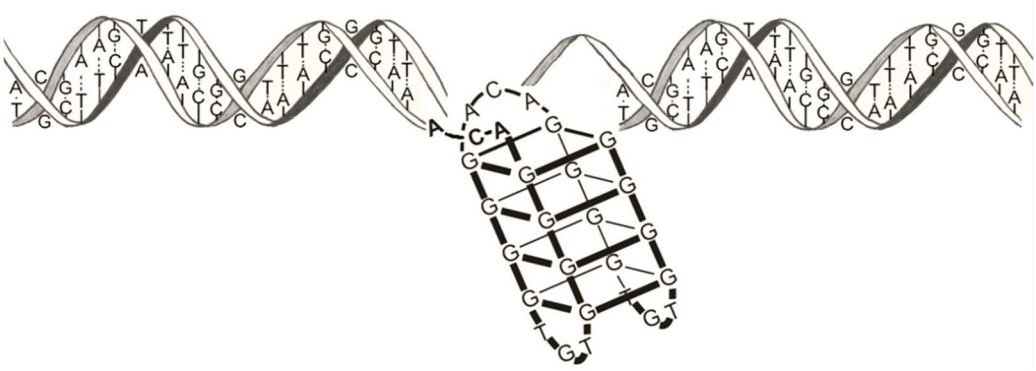


Figure 39 | Representation of the formation of a quadruple DNA helix through Hoogsteen base pairing. The poly-guanine DNA quadruplex is attached to one DNA strand of the gene that wants to be silenced. This DNA conformation inhibits the transcription by the RNA polymerase.
This four-stranded DNA structure naturally occurs in cell's telomeres.

2. **Binding to the mRNA**. There exist three types of special RNAs called microRNA (miRNA), small interfering RNA (siRNA) and piwi-interacting RNA (piRNA). These three types of RNA are small sequences of RNA that can bind to the mRNA sequences inhibiting translation [59].

   The miRNA, siRNA and piRNA can be part of a multiprotein complex known as RISC (RNA induced silencing complex). RISC uses these three RNAs types as templates to identify complementary mRNA strands. If this occurs, then RISC activates certain proteins from the Argonaute family which are the main catalytic components of the complex and degrade or cleave the bonded mRNA [26].

   However, only miRNA and siRNA act on somatic cells with an antisense function while piRNA is active in germline cells. These two molecules can be used in OI to silence the aberrant mRNA produced by the mutant allele. A schematic representation of this is showed in **Figure 40** [43][44][72].

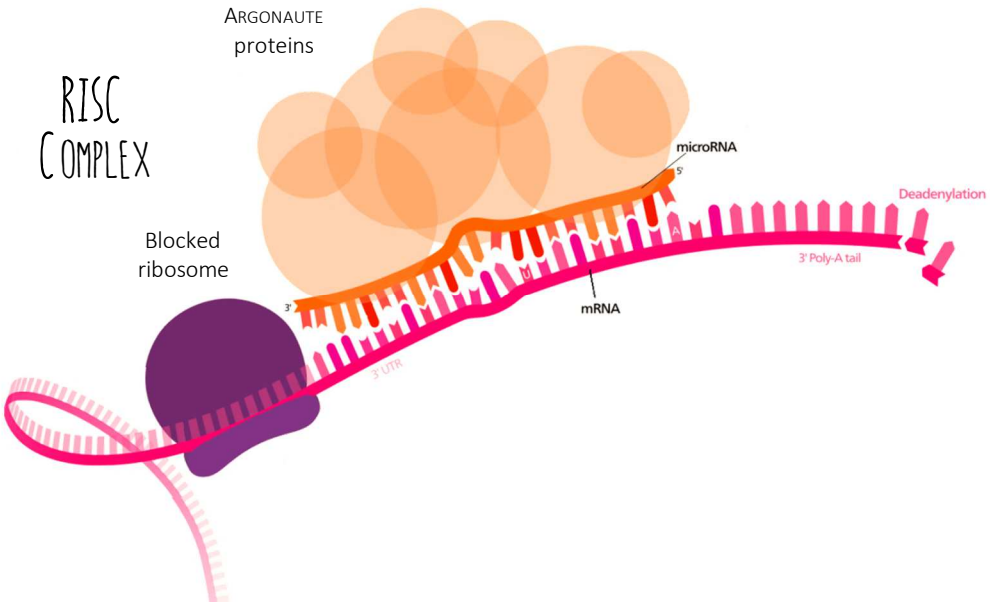


Figure 40 | Schematic representation of the RISC complex bound to a targeted mRNA. As can be seen in the picture, the miRNA is complementary to the targeted mRNA. Both mRNAs bind together to form a double-stranded RNA molecule which cannot be translated by the ribosomes. Consequently, this leads to the degradation of the formed RNA dimer, thus inhibiting the translation of the mRNA.

This antisense method can be used in OI in order to target and degrade the mRNAs coming from the mutant allele with either a glycine substitution or a splice site mutation. However, this method has an inconvenient: specificity of miRNAs are relatively low, making that also normal mRNA is degraded.

In glycine substitutions the aberrant mRNA only differs in one base pair from the normal mRNA. This makes difficult that RISC can correctly recognise and differentiate the abnormal from the correct mRNA. This could lead to the degradation of both types of mRNAs leading to a harmful side effect of this antisense therapy (**Figure 41**).

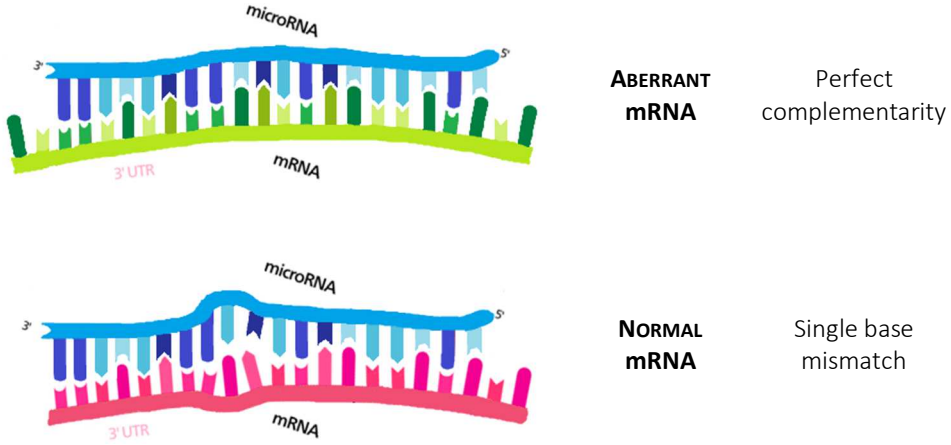


Figure 41 | Schematic representation of the binding of the miRNA to mRNA. In the picture above it can be observed that the miRNA perfectly binds to the targeted aberrant mRNA.
In the picture below, the miRNA recognises and targets a normal mRNA as an aberrant one, since in glycine substitutions there is only a single base pair that is different between both mRNAs. This leads to the degradation of normal mRNAs. In this case, there is a single base mismatch but this does not prevent the miRNA from binding and degrading the targeted mRNA.

In case of splice site mutations, the antisense therapy can have better results than in glycine substitutions. After the mutation identification, the mRNA sequence is determined. Then, an individualized miRNA is synthesised that recognises the mRNA containing the complete or partial exon skippings and introns retentions. Since in splice site mutations the difference between the normal and aberrant mRNAs is more notable, there are not as many recognition problems as in glycine substitutions (**Figure 42**).

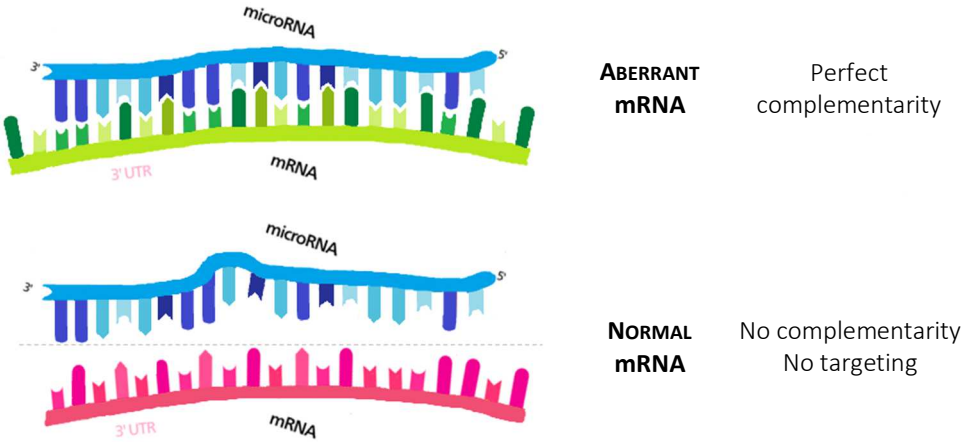


Figure 42 | Schematic representation of the binding of the miRNA to mRNA. In the picture above, it can be observed that the miRNA perfectly binds to the targeted aberrant mRNA.
In the picture below, the miRNA does not target the normal mRNA since the sequence differs greatly from that containing the splice site mutation.

To apply antisense oligonucleotides for gene therapy an individualized nucleotide sequence (transgene) for each patient must be developed according to the mutation they harbour.

In humans, miRNAs and siRNAs are encoded by genes that are transcribed[28] independently or together sharing the same promoter as their host genes. If we want to introduce a gene encoding a miRNA or siRNA that targets the mutant mRNA, theoretically it can be done through both ways [26][59].

However, if we create a replication-defective AAV (safest and most effective option) that only carries the synthesised transgene, we must also provide a strong promoter and gene expression-regulating factors that assure the transcription of this gene and avoid its translation into protein (**Figure 43a**).

Another option is to combine this transgene with an additional copy of the affected gene (either COL1A1 or COL1A2), both of them controlled by a single promoter and integrated in an autonomous operon (**Figure 43b**). In this case, we compensate the collagen insufficiency produced by the mRNA degradation with an additional normal gene copy.

A third option would be to integrate the transgene directly into the same operon as the mutant allele, thereby regulating both genes by the same promoter getting a simultaneous transcription. With this method, we produce equal amount of mRNA and miRNA/siRNA, ensuring that all the aberrant mRNA can be destroyed (**Figure 43c**). This is not yet possible, but in a future we will be able to integrate transgenes into a specific locus of the host cell.

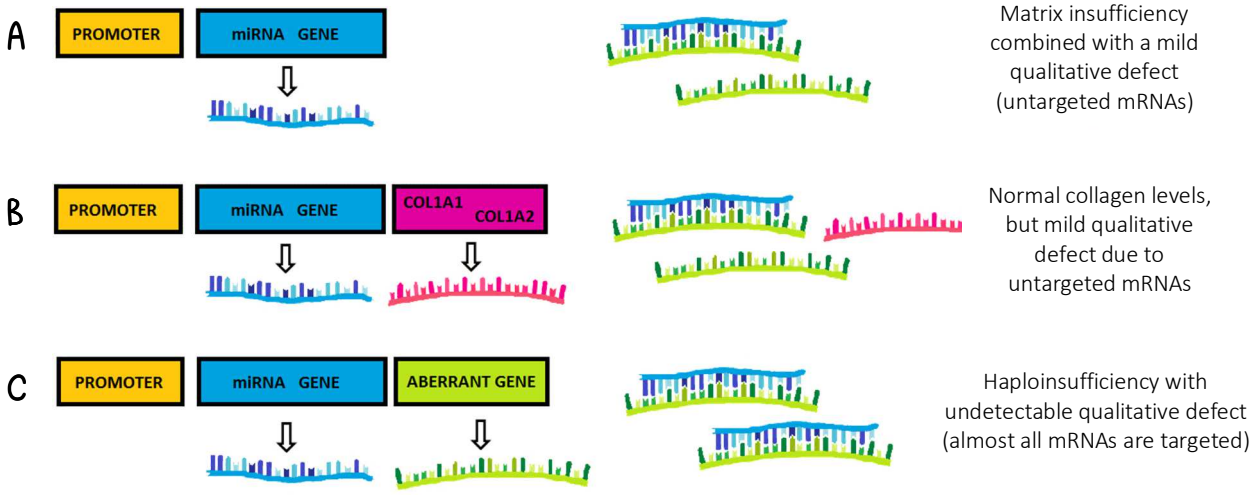


Figure 43 | Different antisense therapy methods are represented.

[28] Both miRNAs and siRNAs are encoded by genes that are transcribed by RNA polymerase II. This fact is important since this give us the option to transcribe this type of genes together with protein-encoding genes (also transcribed by RNA polymerase II). It must be also considered to provide the transgene with the corresponding consensus sequences so that the cell machinery recognises the sequence.

3. 12. PERSPECTIVES FOR THE FUTURE

Nowadays, the field of genetics is still a great enigma. This research project was an attempt to answer and get a bit closer to solve a little mystery I posed at the beginning of this assay.

In my view, the objectives of this research project have been achieved in great measure. Although the initial approach was different than the one finally accomplished, the obtained results justify this slight deviation.

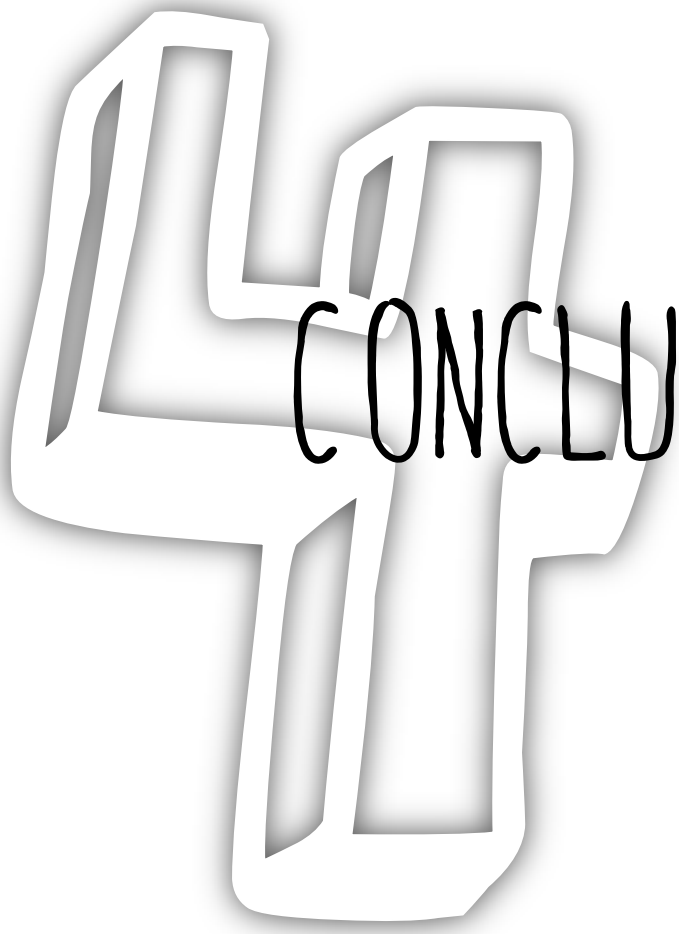
The preliminary idea was to create a system to predict the phenotypic outcome of mutations causing Ehlers-Danlos syndrome. After this, the obtained results would have been integrated into an application compatible with a molecular visualization software. This latter step would have been the research project of my friend Eric, and would have been the practical application of the theoretical work.

Nonetheless, at initial stages of conceptual inception Eric sadly passed away so that the idea of this team work suddenly vanished. At this point I felt that it would be very uncomfortable for me to carry out this research project by myself. This was the reason that I decided to study our sister-disease, osteogenesis imperfecta.

Now, after finishing this research project, I think that all this has been a great personal progress and gaining clarity on my own circumstances. However, I dare to say that it has also been a little progress in understanding such complicated and heterogeneous diseases like hereditary connective tissue disorders.

For the future, I hope that the research of gene therapy progresses and allows us to save hundreds of lives, or at least, improve the quality of life of many people affected by a genetic disorder. We do not realise it, but genetic diseases are present in our daily lives. Although there are more common ones, osteogenesis imperfecta is a good example for that. Almost everyone has heard something about “a child that has bones which break as glass”.

If we manage to understand OI it will be a great progress for many other diseases, including Ehlers-Danlos syndrome. I wish that in a near future we will be able to say that no one must die anymore from a single base substitution.



CONCLUSIONS

1. **Cell surveillance pathways.** Eukaryotic cells have developed different surveillance systems. Among these, nonsense-mediated mRNA decay and unfolded protein response have been studied in this research project. The first one is crucial for the recognition of aberrant mRNA. When this mechanism is involved, a quantitative defect results giving rise to the mild OI type I. Therefore, it can be deduced that indels involving a large number of bp most probably end in NMD and are not extremely severe. Nonetheless, as supposed at the beginning, premature stop codons very downstream are not recognised and result in various OI types.

2. **Glycine substitutions.** Glycine substitutions are the main mutations seen in osteogenesis imperfecta. Its genotype-phenotype relationship is subjected to different mechanisms and correlations. As supposed initially, they are not targeted as aberrant and are introduced into ECM. The obligate glycine residue in every third position has resulted to be crucial for the correct functioning of collagen. This assumption could be proven when looking at mutations corresponding to glycine substitutions.

3. **Regional models.** For long time, there was the strong belief that there are strict and clear regional models in the collagen molecule that could explain the different OI types. The initial assumption of the role of the hydroxylation and electrical charge of neighbouring residues has been disproven, since there could not be found any relationship.

   Now, after doing this research project, we know that there are regional models but they follow rather an irregular fashion than a systematic distribution. It is true that mutations C-terminal are more aberrant than the ones N-terminal since there is a greater disruption of the helix folding. Nonetheless, it has been seen that binding sites for chaperones (HSP47), other proteins (glycosaminoglycans, phosphophoryn) and enzymes (lysyl and prolyl hydroxylase) play a major role in determining the severity of the disease.

4. **Collagen alpha subunits.** At the beginning, it has been supposed that mutations in the alpha-2 chain are more aberrant. It is indeed the contrary; lethal mutations are more often seen in alpha-1 since a mutant chain can be introduced twice in a single tropocollagen molecule and thus have a double-destabilising effect.

5 **Symptom heterogeneity**. There is indeed a spatial relationship that could explain some symptoms that accompany OI. The presence of dentinogenesis imperfecta and/or hearing loss most probably correlates with binding sites for dentin sialophosphoprotein. The presence of blue sclera and wormian bones is more difficult to determinate; there is possibly no clear relationship that can be postulated since they are quantitatively determined (light or dark blue sclera) and not qualitatively (presence/absence of DI or deafness).

6 **Non-glycine substitutions**. Substitutions for non-glycine residues have different consequences. Residues that are important for chain assembly, calcium ion binding or ligand recognition have mixed outcomes depending on the abolished function. Non-glycine substitutions are also seen in splice site mutations, having outcomes ranging from mild to lethal. Therefore, the initial assumption of the harmlessness of these mutations has been invalidated.

7 **The imperfect perfections of the collagen molecule**. In this research project it has been seen that the collagen molecule is perfectly adapted to its environment. Crosslinks and hydroxylation determine the stability of the tropocollagen molecule. Nevertheless, minimal alterations and variations can lead to a disastrous disruption of normal protein architecture. It has been seen, on the contrary than initially supposed, that collagen is thermally instable at body temperature. There are also regions in the tropocollagen molecule that show a low melting temperature. Therefore, minimal variations in its architecture have atrocious consequences.

8 **Other collagen-related diseases.** Osteogenesis imperfecta is a relative common disease among the connective tissue disorders. If osteogenesis imperfecta is well studied, this will be a great improvement for other collagen related disorders. There are many other connective tissue disorders: Bethlem and Ullrich congenital muscular dystrophies, Marfan syndrome, various osteochondrodysplasias, Ehlers-Danlos syndrome, Larsen syndrome and Loeys-Dietz syndrome, to name a few. This is a very complex group of diseases with different pathogenesis. However, if we manage to understand some of them, it will be a great progress for other related diseases that only affect a few people around the world.

9 **Non-collagen related OI types.** The recently discovery of OI types related to enzymes, NCPs and chaperones has improved exponentially the understanding of the classical OI types. Now, we know that the severity of OI is influenced in a multifactorial manner, giving rise to a different comprehension of the disease. In parallel, people affected by OI that underwent genetic testing without finding a causative mutation now have the chance to look for other OI subtypes that could explain their disease. Maybe in a near future, there will also be available a gene therapy for these OI types.

10 **The importance of understanding the genetic background of OI.** Osteogenesis imperfecta is a very heterogeneous and complex disease. This research project is an attempt to find an unifying mechanism that can be used to assess the severity of the disease.

Patients with osteogenesis imperfecta are usually clinically diagnosed. This leads to an unawareness of the condition and ambiguity when it comes to typify the disease. If genetic testing is done, the mutation can be identified and the clinical outcome and prognosis can be better evaluated. Consequently, the quality of life of the patient can be improved. There will be methods available to prevent further health worsening and deterioration, applying preventive measures and more individualised treatments.

Genetic testing has another important application. If the genetic and molecular background of osteogenesis imperfecta is understood, there will be better chances to find a gene therapy. A curative treatment for osteogenesis imperfecta will be very challenging due to its dominant negative nature. If gene therapy is developed, it is not enough with supplying the healthy gene; it is also necessary to silence the aberrant allele and induce upregulation and overexpression of the healthy one. All this research will be faster and more accurate if we understand completely all the mechanisms influencing the disease.

11 **Improve the quality of life of millions of woman.** Osteoporosis is a bone-mass-decreasing disease affecting about a third of the population, mostly postmenopausal woman above age 50. There are a lot of environmental factors involved in the disease, but there are also genetic factors that put at risk of developing this progressive condition. If we study osteogenesis imperfecta and its genetics, it will be possible to better understand the role of different genes in osteoporosis. This, consequently, will improve the health of thousands of men and women around the world.

– ABBREVIATIONS –

bp: base pairs

CTD: connective tissue disorder

DI: dentinogenesis imperfecta

DMP1: dentin matrix protein 1

DSPP: dentin sialophosphoprotein

ECM: extracellular matrix

EDS: Ehlers-Danlos syndrome

EJC: exon-junction complex

ER: endoplasmic reticulum

ERAD: endoplasmic-reticulum-associated protein degradation

ESE: exonic splicing enhancer

ESS: exonic splicing silencer

GFP: green fluorescent protein

HSP47: heat shock protein 47

Hyp (O): hydroxyproline

IBSP: bone sialoprotein 1

indels: mutations involving both insertions and deletions at the same time.

LBS: ligand binding sites

LH: lysyl hydroxylase

miRNA: micro RNA

MLBR: major ligand binding regions

MSC: mesenchymal stem cells

NCP: non-collagenous proteins

NMD: nonsense–mediated mRNA decay

OI: osteogenesis imperfecta

- **OI – I:** mild
- **OI – IV:** moderate
- **OI – III:** severe
- **OI – II:** lethal

OMIM: online Mendelian inheritance in man

PCD: programmed cell death (apoptosis)

PH3: prolyl 3-hydroxylase

PH4: prolyl 4-hydroxylase

piRNA: piwi-interacting RNA

PTC: premature termination (stop) codon

PTM: post-translational modifications

RISC: RNA induced silencing complex

SIBLING proteins: small integrin-binding ligand, N-linked glycoproteins

siRNA: small interfering RNA

SPP1: osteopontin

UPR: unfolded protein response

Bibliography.

1. AGRAWAL N, DASARADHI PV, MOHMMED A, MALHOTRA P, BHATNAGAR RK, MUKHERJEE SK. *RNA interference: biology, mechanism, and applications*. Microbiol Mol Biol Rev. 2003;67:657-85.

2. ASH JE, BUDAVARI S, O'NEILL M ET AL. *The Merck Index*. Chapman & Hall; 1996.

3. BAKER KE, PARKER R. *Nonsense-mediated mRNA decay: terminating erroneous gene expression*. Curr Opin Cell Biol. 2004;16(3):293-299.

4. BECK K, CHAN VC, SHENOY N, KIRKPATRICK A, RAMSHAW JA, BRODSKY B. *Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine*. Proc Natl Acad Sci USA. 2000;97(8):4273-4278.

5. BODIAN DL, CHAN TF, POON A, ET AL. *Mutation and polymorphism spectrum in osteogenesis imperfecta type II: implications for genotype-phenotype relationships*. Hum Mol Genet. 2009;18(3):463-471.

6. BODIAN DL, MADHAN B, BRODSKY B, KLEIN TE. *Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations*. Biochemistry. 2008;47(19):5424-5432.

7. BRINCKMANN J, NOTBOHM H, MÜLLER P. *Collagen, Primer in Structure, Processing and Assembly*. Springer Science & Business Media; 2005.

8. BURROWS NP, NICHOLLS AC, RICHARDS AJ, ET AL. *A point mutation in an intronic branch site results in aberrant splicing of COL5A1 and in Ehlers-Danlos syndrome type II in two British families. Am J Hum Genet*. 1998;63(2):390-8.

9. BYERS PH. *Folding defects in fibrillar collagens*. Philos Trans R Soc Lond, B, Biol Sci. 2001;356(1406):151-157.

10. BYERS PH, WALLIS GA, WILLING MC. *Osteogenesis imperfecta: translation of mutation to phenotype*. J Med Genet. 1991;28(7):433-442.

11. CABRAL WA, CHANG W, BARNES AM, ET AL. *Prolyl 3-hydroxylase 1 deficiency causes a recessive metabolic bone disorder resembling lethal/severe osteogenesis imperfecta*. Nat Genet. 2007;39(3):359-365.

12. CHAN VC, RAMSHAW JA, KIRKPATRICK A, BECK K, BRODSKY B. *Positional preferences of ionizable residues in Gly-X-Y triplets of the collagen triple-helix*. J Biol Chem. 1997;272(50):31441-31446.

13. CHEN Y, BAL BS, GORSKI JP. *Calcium and collagen binding properties of osteopontin, bone sialoprotein, and bone acidic glycoprotein-75 from bone*. J Biol Chem. 1992;267(34):24871-24878.

14. CHENG H, RASHID S, YU Z, YOSHIZUMI A, HWANG E, BRODSKY B. *Location of glycine mutations within a bacterial collagen protein affects degree of disruption of triple-helix folding and conformation.* J Biol Chem. 2011;286(3):2041-2046.

15. CHRISTIANSEN HE, SCHWARZE U, PYOTT SM, ET AL. *Homozygosity for a missense mutation in SERPINH1, which encodes the collagen chaperone protein HSP47, results in severe recessive osteogenesis imperfecta*. Am J Hum Genet. 2010;86(3):389-398.

16. DAHL T, SABSAY B, VEIS A. *Type I collagen-phosphophoryn interactions: specificity of the monomer-monomer binding*. J Struct Biol. 1998;123(2):162-8.

17. DALGLEISH R. *The human type I collagen mutation database*. Nucleic Acids Res. 1997;25(1):181-187.

18. DESHPANDE AS, FANG PA, ZHANG X, JAYARAMAN T, SFEIR C, BENIASH E. *Primary structure and phosphorylation of dentin matrix protein 1 (DMP1) and dentin phosphophoryn (DPP) uniquely determine their role in biomineralization*. Biomacromolecules. 2011;12(8):2933-2945.

19. DEYLE DR, KHAN IF, REN G, ET AL. *Normal collagen and bone production by gene-targeted human osteogenesis imperfecta iPSCs*. Mol Ther. 2012;20(1):204-13.

20. DI LULLO GA, SWEENEY SM, KORKKO J, ALA-KOKKO L, SAN ANTONIO JD. *Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen*. J Biol Chem. 2002;277(6):4223-4231.

21. DRAKE R, VOGL AW, MITCHELL AW. *Gray's Anatomy for Students*. Elsevier Health Sciences; 2009.

22. FISHER LW, TORCHIA DA, FOHR B, YOUNG MF, FEDARKO NS. *Flexible structures of SIBLING proteins, bone sialoprotein, and osteopontin*. Biochem Biophys Res Commun. 2001;280(2):460-465.

23. GAJKO-GALICKA A. *Mutations in type I collagen genes resulting in osteogenesis imperfecta in humans*. Acta Biochim Pol. 2002;49(2):433-441.

24. GELB BD, SHI GP, CHAPMAN HA, DESNICK RJ. *Pycnodysostosis, a lysosomal disease caused by cathepsin K deficiency*. Science. 1996;273(5279):1236-1238. [Abstract].

25. GRAW J. *Lehrbuch der Molekularen Zellbiologie*. John Wiley & Sons. 2012.

26. GREGORY RI, CHENDRIMADA TP, COOCH N, SHIEKHATTAR R. *Human RISC couples microRNA biogenesis and posttranscriptional gene silencing*. Cell. 2005;123(4):631-40.

27. HE G, GEORGE A. *Dentin matrix protein 1 immobilized on type I collagen fibrils facilitates apatite deposition in vitro*. J Biol Chem. 2004;279(12):11649-56.

28. HOLSTER T, PAKKANEN O, SOININEN R, ET AL. *Loss of assembly of the main basement membrane collagen, type IV, but not fibril-forming collagens and embryonic death in collagen prolyl 4-hydroxylase I null mice*. J Biol Chem. 2007;282(4):2512-2519.

29. HOMAN EP, LIETMAN C, GRAFE I, ET AL. *Differential effects of collagen prolyl 3-hydroxylation on skeletal tissues*. PLoS Genet. 2014;10(1):e1004121.

30. HUANG B, SUN Y, MACIEJEWSKA I, ET AL. *Distribution of SIBLING proteins in the organic and inorganic phases of rat dentin and bone.* Eur J Oral Sci. 2008;116(2):104-112.

31. HYDE TJ, BRYAN MA, BRODSKY B, BAUM J. *Sequence dependence of renucleation after a Gly mutation in model collagen peptides.* J Biol Chem. 2006;281(48):36937-36943.

32. KHOSHNOODI J, CARTAILLER JP, ALVARES K, VEIS A, HUDSON BG. *Molecular recognition in the assembly of collagens: terminal noncollagenous domains are key recognition modules in the formation of triple helical protomers*. J Biol Chem. 2006;281(50):38117-38121.

33. LAMANDÉ SR, BATEMAN JF. *Procollagen folding and assembly: the role of endoplasmic reticulum enzymes and molecular chaperones*. Semin Cell Dev Biol. 1999;10(5):455-464.

34. LEIKINA E, MERTTS MV, KUZNETSOVA N, LEIKIN S. *Type I collagen is thermally unstable at body temperature*. Proc Natl Acad Sci USA. 2002;99(3):1314-1318.

35. LI F, WANG X, NIYIBIZI C. *Distribution of single-cell expanded marrow derived progenitors in a developing mouse model of osteogenesis imperfecta following systemic transplantation*. Stem Cells. 2007;25(12):3183-93.

36. LISSE TS, THIELE F, FUCHS H, ET AL. *ER stress-mediated apoptosis in a new mouse model of osteogenesis imperfecta*. PLoS Genet. 2008;4(2):e7.

37. MAKAREEVA E, CABRAL WA, MARINI JC, LEIKIN S. *Molecular mechanism of alpha 1(I)-osteogenesis imperfecta/Ehlers-Danlos syndrome: unfolding of an N-anchor domain at the N-terminal end of the type I collagen triple helix.* J Biol Chem. 2006;281(10):6463-6470.

38. MAKAREEVA E, MERTZ EL, KUZNETSOVA NV, ET AL. *Structural heterogeneity of type I collagen triple helix and its role in osteogenesis imperfecta*. J Biol Chem. 2008;283(8):4787-4798.

39. MARINI JC, CABRAL WA, BARNES AM, CHANG W. *Components of the collagen prolyl 3-hydroxylation complex are crucial for normal bone development*. Cell Cycle. 2007;6(14):1675-1681.

40. MARINI JC, FORLINO A, CABRAL WA, et al. *Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans*. Hum Mutat. 2007;28(3):209-221.

41. MILLINGTON-WARD S, MCMAHON HP, FARRAR GJ. *Emerging therapeutic approaches for osteogenesis imperfecta*. Trends Mol Med. 2005;11(6):299-305.

42. MOTYCKOVA G, FISHER DE. *Pycnodysostosis: role and regulation of cathepsin K in osteoclast function and human disease*. Curr Mol Med. 2002;2(5):407-421. [Abstract]

43. NIYIBIZI C, SMITH P, MI Z, ROBBINS P, EVANS C. *Potential of gene therapy for treating osteogenesis imperfecta*. Clin Orthop Relat Res. 2000;(379 Suppl):S126-33.

44. NIYIBIZI C, WANG S, MI Z, ROBBINS PD. *Gene therapy approaches for osteogenesis imperfecta*. Gene Ther. 2004;11(4):408-16.

45. NOVO VILLAVERDE, F J. *Genética humana: conceptos, mecanismos y aplicaciones de la genética en el campo de la medicina*. Pearson/Prentice-Hall. 2007.

46. PACE JM, KUSLICH CD, WILLING MC, BYERS PH. *Disruption of one intra-chain disulphide bond in the carboxyl-terminal propeptide of the proalpha1(I) chain of type I procollagen permits slow assembly and secretion of overmodified, but stable procollagen trimers and results in mild osteogenesis imperfecta. J* Med Genet. 2001;38(7):443-449.

47. PACE JM, WIESE M, DRENGUIS AS, ET AL. Defective *C-propeptides of the proalpha2(I) chain of type I procollagen impede molecular assembly and result in osteogenesis imperfecta*. J Biol Chem. 2008;283(23):16061-16067.

48. PENIDO MG, ALON US. *Phosphate homeostasis and its role in bone health*. Pediatr Nephrol. 2012;27(11):2039-48.

49. PERSIKOV AV, RAMSHAW JA, BRODSKY B. *Prediction of collagen stability from amino acid sequence*. J Biol Chem. 2005;280(19):19343-19349.

50. POCHAMPALLY RR, HORWITZ EM, DIGIROLAMO CM, STOKES DS, PROCKOP DJ. *Correction of a mineralization defect by overexpression of a wild-type cDNA for COL1A1 in marrow stromal cells (MSCs) from a patient with osteogenesis imperfecta: a strategy for rescuing mutations that produce dominant-negative protein defects*. Gene Ther. 2005;12(14):1119-25.

51. PROCKOP DJ, COLIGE A, HELMINEN H, KHILLAN JS, PEREIRA R, VANDENBERG P. *Mutations in type 1 procollagen that cause osteogenesis imperfecta: effects of the mutations on the assembly of collagen into fibrils, the basis of phenotypic variations, and potential antisense therapies*. J Bone Miner Res. 1993;8 Suppl 2:S489-92.

52. QIN C, BABA O, BUTLER WT. *Post-translational modifications of sibling proteins and their roles in osteogenesis and dentinogenesis*. Crit Rev Oral Biol Med. 2004;15(3):126-136.

53. RAUCH F, LALIC L, ROUGHLEY P, GLORIEUX FH. *Genotype-phenotype correlations in nonlethal osteogenesis imperfecta caused by mutations in the helical domain of collagen type I*. Eur J Hum Genet. 2010;18(6):642-647.

54. RICARD-BLUM S. *The collagen family*. Cold Spring Harb Perspect Biol. 2011;3(1):a004978.

55. ROSSERT J, TERRAZ C, DUPONT S. *Regulation of type I collagen genes expression*. Nephrol Dial Transplant. 2000;15 Suppl 6:66-8.

56. SHAPIRO JR, BYERS PH, GLORIEUX FH ET AL. *Osteogenesis Imperfecta: A Translational Approach to Brittle Bone Disease*. Academic Press; 2013.

57. SILVERMAN SL. *Bisphosphonate use in conditions other than osteoporosis*. Ann N Y Acad Sci. 2011;1218:33-7.

58. SILVERMAN SL, EVANGELISTA L. *Osteoporosis*. Rheum Dis Clin North Am. 2011;37(3):ix-x.

59. SIOMI H, SIOMI MC. *On the road to reading the RNA-interference code*. Nature. 2009;457(7228):396-404.

60. SHOULDERS MD, RAINES RT. *Collagen structure and stability*. Annu Rev Biochem. 2009;78:929-958.

61. STAINES KA, MACRAE VE, FARQUHARSON C. *The importance of the SIBLING family of proteins on skeletal mineralisation and bone remodelling*. J Endocrinol. 2012;214(3):241-255.

62. SWEENEY SM, ORGEL JP, FERTALA A, ET AL. *Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates*. J Biol Chem. 2008;283(30):21187-97.

63. SWINNEN FK, COUCKE PJ, DE PAEPE AM, ET AL. *Osteogenesis Imperfecta: the audiological phenotype lacks correlation with the genotype*. Orphanet J Rare Dis. 2011;6:88.

64. SYMOENS S, HULMES DJ, BOURHIS JM, COUCKE PJ, PAEPE AD, MALFAIT F. *Type I Procollagen C-Propeptide Defects: Study of Genotype-Phenotype Correlation and Predictive Role of Crystal Structure*. Hum Mutat. 2014.

65. TASAB M, BATTEN MR, BULLEID NJ. *HSP47: A molecular chaperone that interacts with and stabilizes correctly-folded procollagen*. EMBO J. 2000;19(10):2204-2211.

66. VEIS A, DORVEE JR. *Biomineralization mechanisms: a new paradigm for crystal nucleation in organic matrices*. Calcif Tissue Int. 2013;93(4):307-315.

67. WADE LG. *Organic Chemistry*. 7th Editon, Pearson/Prentice-Hall. 2012.

68. VAN DIJK FS, HUIZER M, KARIMINEJAD A, ET AL. *Complete COL1A1 allele deletions in osteogenesis imperfecta*. Genet Med. 2010;12(11):736-741.

69. VANLEENE M, SALDANHA Z, CLOYD KL, ET AL. *Transplantation of human fetal blood stem cells in the osteogenesis imperfecta mouse leads to improvement in multiscale tissue properties*. Blood. 2011;117(3):1053-60.

70. WIDMER C, GEBAUER JM, BRUNSTEIN E, ET AL. *Molecular basis for the action of the collagen-specific chaperone HSP47/SERPINH1 and its structure-specific client recognition*. Proc Natl Acad Sci USA. 2012;109(33):13243-7.

71. XIAO S, YU C, CHOU X, ET AL. *Dentinogenesis imperfecta 1 with or without progressive hearing loss is associated with distinct mutations in DSPP*. Nat Genet. 2001;27(2):201-4.

72. WANG Q, MARINI JC. *Antisense oligodeoxynucleotides selectively suppress expression of the mutant alpha 2(I) collagen allele in type IV osteogenesis imperfecta fibroblasts*. *A molecular approach to therapeutics of dominant negative disorders*. J Clin Invest. 1996;97(2):448-54.

73. WATSON, JAMES D. *Molecular Biology of the Gene*. Academic Internet Publishers; 2014.

74. WILLING MC, DESCHENES SP, SLAYTON RL, ROBERTS EJ. *Premature chain termination is a unifying mechanism for COL1A1 null alleles in osteogenesis imperfecta type I cell strains*. Am J Hum Genet. 1996;59(4):799-809.

75. ZHANG X, GODBEY WT. *Viral vectors for gene delivery in tissue engineering*. Adv Drug Deliv Rev. 2006;58(4):515-34.

Online bibliographic references.

76. GENETICS HOME REFERENCE. *Cathepsin K.* [Last consulted on August 2014]
< http://ghr.nlm.nih.gov/gene/CTSK >

77. GENETICS HOME REFERENCE. *COL1A1.* [Last consulted on September 2014]
< http://ghr.nlm.nih.gov/gene/COL1A1 >

78. GENETICS HOME REFERENCE. *COL1A2*. [Last consulted on September 2014]
< http://ghr.nlm.nih.gov/gene/COL1A2 >

79. GENETICS HOME REFERENCE. *Duchenne and Becker muscular dystrophy.*
[Last consulted on May 2014]
< http://ghr.nlm.nih.gov/condition/duchenne-and-becker-muscular-dystrophy >

80. GENETICS HOME REFERENCE. *Dystrophin.* [Last consulted on May 2014]
< http://ghr.nlm.nih.gov/gene/DMD >

81. GENETICS HOME REFERENCE. *Osteogenesis imperfecta*. [Last consulted on November 2014]
< http://ghr.nlm.nih.gov/condition/osteogenesis-imperfecta >

82. GENETICS HOME REFERENCE. *Osteopetrosis*. [Last consulted on July 2014]
< http://ghr.nlm.nih.gov/condition/osteopetrosis >

83. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Bone morphogenetic protein 1; BMP1.* MIM Number: {166220 }

84. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Cartilage-associated protein; CRTAP*. MIM Number: {605497}

85. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Ehlers-Danlos syndrome, type VII [artrochalasia].* MIM Number: {130060}

86. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *FK506-binding protein 10; FKBP10.* MIM Number: {607063}

87. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Interferon-induced transmembrane protein 5; IFITM5.* MIM Number: {614757}

88. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Leucine- and proline-enriched proteoglycan 1; LEPRE1*. MIM Number: {610339}

89. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type I*. MIM Number: {166200}

90. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type II*. MIM Number: {166210}

91. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type III*. MIM Number: {259420}

92. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type IV.* MIM Number: {166220 }

93. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type V.* MIM Number: {610967}

94. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type VI*. MIM Number: {613982}

95. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type VII*. MIM Number: {610682}

96. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type VIII.* MIM Number: {610915}

97. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type IX.* MIM Number: {259440}

98. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type X*. MIM Number: {613848}

99. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type XI.* MIM Number: {610968}

100. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type XII.* MIM Number: {613849}

101. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type XIII.* MIM Number: {614856}

102. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Osteogenesis imperfecta, type XIV.* MIM Number: {615066}

103. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Peptidyl-prolyl isomerase B; PPIB.* MIM Number: {123841}

104. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Pycnodysostosis.* MIM Number: {265800}

105. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Serpin peptidase inhibitor, clade F, member 1; SERPINF1.* MIM Number: {172860}

106. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Serpin peptidase inhibitor, clade H, member 1; SERPINH1.* MIM Number: {600943}

107. ONLINE MENDELIAN INHERITANCE IN MAN, OMIM®. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MD. *Transmembrane protein 38B; TMEM38B*. MIM Number: {611236 }

Webgraphy.

108. OSTEOGENESIS IMPERFECTA FOUNDATION. *Facts about Osteogenesis Imperfecta.* (2008).
< http://www.oif.org/site/PageServer?pagename=AOI_Facts > [Last consulted on August 2014]

109. OSTEOGENESIS IMPERFECTA FOUNDATION. *Osteogenesis imperfecta types. (2008).*
< http://www.oif.org/site/PageServer?pagename=AOI_Types > [Last consulted on July 2014]

110. OSTEOGENESIS IMPERFECTA FOUNDATION. *Understanding the structure of bones.* (2002).
< http://www.oif.org/site/PageServer?pagename=BoneStruct > [Last consulted on August 2014]

111. PEARSON EDUCATION INC. *Amino acids, peptides and proteins*. (2014)
[Last consulted on April 2014]

112. RADIOPAEDIA. *Alternating radiolucent and radiodense metaphyseal lines (zebra stripe sign).*
[Last consulted on May 2014]
< http://radiopaedia.org/articles/alternating-radiolucent-and-radiodense-metaphyseal-lines >

113. RADIOPAEDIA. *Osteopetrosis.* [Last consulted on August 2014]
< http://radiopaedia.org/articles/osteopetrosis >

114. UNIPROT. *Collagen alpha-1(I) chain.* [Last consulted on October 2014]
< http://www.uniprot.org/uniprot/P02452 >

115. UNIPROT. *Collagen alpha-1(II) chain.* [Last consulted on October 2014]
< http://www.uniprot.org/uniprot/P08123 >

116. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Collagen*. [Last consulted on October 2014]
< http://en.wikipedia.org/wiki/Collagen >

117. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Glycine*. [Last consulted on May 2014]
< http://en.wikipedia.org/wiki/Glycine >

118. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Hydroxylysine.* [Last consulted on May 2014]
< http://en.wikipedia.org/wiki/Hydroxylysine >

119. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Hydroxyproline.* [Last consulted on May 2014]
< http://en.wikipedia.org/wiki/Hydroxyproline >

120. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Lysine*. [Last consulted on September 2014]
< http://en.wikipedia.org/wiki/Lysine >

121. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Ossification*. [Last consulted on November 2014]
< http://en.wikipedia.org/wiki/Ossification >

122. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Osteogenesis imperfecta.* [Last consulted on August 2014]
< http://en.wikipedia.org/wiki/Osteogenesis_imperfecta >

123. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Proline*. [Last consulted on May 2014]
< http://en.wikipedia.org/wiki/Proline >

124. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *SIBLING proteins*. [Last consulted on September 2014]
< http://en.wikipedia.org/wiki/SIBLING_proteins >

125. WIKIPEDIA, THE FREE ENCYCLOPAEDIA. *Unfolded protein response*. [Last consulted on August 2014]
< http://en.wikipedia.org/wiki/Unfolded_protein_response >

- FIGURES REFERENCES –

- **Title photo**. Anlene World Osteoporosis Day.
- **Figure 1.** Wikimedia Commons.
  < http://commons.wikimedia.org/wiki/File:Fibers_of_Collagen_Type_I_-_TEM.jpg >
- **Figure 4a**. Wikimedia Commons.
  < http://commons.wikimedia.org/wiki//File:Collagen_axis.png >
- **Figure 6**. KADLER KE, HOLMES DF, TROTTER JA, CHAPMAN JA. *Collagen fibril formation*. Biochem J. 1996;316 (Pt 1):1-11.
- **Figure 8**. Wikimedia Commons.
  < http://commons.wikimedia.org/wiki//File:OI_full_body_x_rays_types.png >
- **Figure 12b**. Wikimedia Commons.
  < http://commons.wikimedia.org/wiki//File:OI_type_V_forearm.jpg >
- **Figure 35**. Wikimedia Commons.
  < http://commons.wikimedia.org/wiki//File:Gene_therapy_steps_iPSC.png >
- **Figure 39**. Wikimedia Commons.
  < http://commons.wikimedia.org/wiki//File:MiRNA.svg >

Figures that are not cited above are own creations. The radiographies and medical photographs contained in this research project are from personal resources. In molecular representations, the file source and name are already cited.

This research project and the information in it contain confidential medical information. Any distribution or use for different purposes without consent of the author is unauthorised. For further information, contact with the author.