DEVELOPING NEW SILK BASED COMPOSITE MATERIALS FOR DENTAL AND BONE REPAIR

Alberto Giménez Gómez[‡], Carole C. Perry[†]

[‡] Universitat Rovira i Virgili, Av. Països Catalans Núm. 26, Tarragona, Espanya.

[†] Biomolecular and Materials Interface Research Group, Interdisciplinary Biomedical Research Centre, School of Science and Technology, Nottingham Trent University, Clifton Lane, UK NG11 8NS.

Chemistry Bachelor

2017/2018

Table of Contents

RESUM/ABSTRACT

CATALÀ 

En aquest projecte s'ha treballat amb una de les cadenes que conformen la fibroïna de la seda natural, anomenada cadena lleugera. S'ha fet recerca bibliogràfica sobre com és la millor manera de funcionalitzar aquesta i així millorar les seves propietats com a material per a aplicacions biomèdiques. Tot i que no s'ha aconseguit realitzar cap procediment per aconseguir-ho, s'ha continuat amb la segona part del treball. En aquesta s'ha estudiat a fons la cadena lleugera en termes d'estructura secundària i altres factors que afecten la seva estabilitat. S'ha aconseguit trobar una seqüència (pèptid) en aquesta cadena suficientment curta (37 aminoàcids) com per a ser sintetitzada via innovadores tècniques de síntesi de pèptids (Síntesi de pèptids en fase sòlida, SPPS) i que és estable per si mateixa, aïllada de la resta de la proteïna. S'ha pogut purificar amb tècniques cromatogràfiques (Cromatografia líquida d'alta resolució, HPLC) i confirmar la seva presència amb tècniques espectroscòpiques (Ionització làser assistida per matriu analitzada per temps de vol, MALDI-TOF). Futures investigacions es duran a terme per a continuar aquest projecte, com caracteritzar completament el pèptid via tècniques espectroscòpiques (dicroisme circular, CD-UV/ ressonància magnètica nuclear, RMN/ infraroig, IR) i comprovar la seva reactivitat i estabilitat en medis *in vivo*.

ENGLISH 

In this project the main issue is the study of one of the chains that form the natural silk fibroin, which is named light chain. Bibliographical research about which is the best way to functionalise it has been done, in order to improve its properties as a biomaterial for medical purposes. Although any procedure to do this functionalization has been achieved the project has continued to the second part of the objective. In this part the light chain has been studied thoroughly in terms of secondary structure and stabilization factors. A short enough (37mer) sequence has been found, so it has been feasible to synthesize it with state of the art techniques (SPPS). This peptide has been purified (HPLC) and its presence has been confirmed (MALDI-TOF). Further investigations will be done in this field, like complete characterization (CD-UV, NMR, IR), and stability in *in vivo* medium.

AKNOWLEDGMENTS

I would like to say thank you to all the research fellows that have helped me during these 3 months abroad. Without you this project would have been more than taught. Unluckily I do not have enough space to write down every word that I should. To name all of you, Carole Perry, Daniel Oliver, Aaron Salinger, Andrew Organisciak, Sa Dave Belton, Graham Hickman, Victor Volkov, Monika Michaelis, Aneeqa Fayyaz.

1. AIM OF THIS PROJECT

The current study was focused on the functionalization of silks fibroin (SF) protein and silk fibroin peptide fragment with a particular poly (ethylene glycol) (PEG) in order to found modified, possibly improved properties. Enhanced toughness, tissue stability, *in vivo* environment resistance might be some of them. All of these properties would upgrade the possibility of this material as tissue engineering construct. When functionalization of the silk has been assessed the next step was to found a Light Chain-Silk Fibroin (LC-SF) fragment that must be stable enough by itself. Nevertheless, due to a lack of results in the functionalization part, the project went more to the computational side.

Once this fragment was studied and we ensured that was feasible to synthesize it, we made it and purified it. After that, the main objective was to functionalize this fragment following the previous procedure and compare the properties of both functionalized SF and SF-LC fragment, it could not be done.

2. INTRODUCTION

The use of foreign materials in the human body has increased dramatically over the past 50 years although the origin of their use dates at least to the time of the pharaohs who used natural silk for sutures and more recently 150 years ago when amalgam dental fillings were first used[1]. Over the years these materials have become more sophisticated with biocompatibility and bioacceptance being important. Silk is a valuable natural source of biomaterials that can be fabricated into differing physical forms[2] and can be treated from either aqueous solution or organic solvent. It can be processed as fibers, hydrogel, sponge scaffolds, microspheres or films.

Due to their low immunogenicity *in* vivo, their ease to be chemically modified[3] and their paper as pro-regenerative scaffolds, silk and its derivatives are unique for being employed as a composite materials for physiological purposes[4]. Recently studies about controlled drug release[5] or tissue engineering constructs[6] enables novel applications of silk to be developed for a range of real world issues from biomedical prosthesis[7] to clean fuel generation[8].

Silk fibroin (SF) is a natural polymer produced by a variety of insects and spiders. Therefore, it is subject to a wide diversity in structure, sequence and properties. The best characterized silks[9] are the cocoon silk from the domesticated silkworm *Bombyx Mori* and the dragline silk from the spider *Nephila clavipes*, although there are industrially –viable fermentation processes [10,11] for obtaining engineered silk protein analogues . Silkworm silk has been studied in more depth because silkworms are easier to domesticate and the silk can be obtained in an easier way comparing to spider silk. Therefore, the focus of this project will be on the use of SF obtained from silkworms and short peptide synthetic fragments obtained from silkworms' SF sequence.

2.1 Silk properties

Silk from *B. Mori* is composed of silk fibroin protein coated with sericin proteins. Sericin is a glue-like protein that accounts for 25-30% of the total silkworm cocoon weight [2]. Sericins have been associated with immune response, but due to their higher hydrophilicity as compared to fibroin they can be easily removed by boiling silk in alkaline solutions.[2] The silk fibroin consists of a light chain (LC) (M_w~25kDa) and a heavy chain (HC) (M_w~325kDa)[5] linked by a disulphide bond.

The heavy chain of silk fibroin is a block copolymer rich in hydrophobic β –sheet-forming blocks linked by small hydrophilic linker segments or spacers. The crystalline regions are primarily composed of glycine-X

repeats, where X is alanine, serine, threonine or valine. Within these domains lie subdomains that are rich in glycine, alanine, serine and tyrosine. The result is a hydrophobic protein that self-assembles to form strong and resilient materials. The hydrophilic part of the core is non-repetitive and very short compared to the size of the hydrophobic repeats.

The dominance of the β-sheet-forming regimes within the fibroin structure imparts the protein-based materials with high mechanical strength and toughness. The toughness of silk fibers is greater than the best synthetic materials, including *Kevlar*[2]. In terms of strength, silk-worm silk is superior to commonly used polymeric degradable biomaterials such as collagen and poly (L-Lactic acid) (PLA). The ultimate tensile strength of *B. mori* silk fibers is 740 MPa. In contrast, collagen has an ultimate tensile strength of 0.9–7.4 MPa and PLA 28–50 MPa[2]. Hence, silk fibroin is an excellent candidate polymer for biomedical applications.

2.2 Silk PEGylation

Poly (ethylene glycol) has been used for many years as a modification and conjugation reagent for biological molecules. Part of the rationale for using PEG polymers in bioconjugation applications includes a dramatic increase in the water solubility of modified molecules, a decrease in immunogenicity due to the shielding of modified molecules from the immune system, protection of modified protein and peptides from digestion by proteolytic enzymes, and effectively increasing their hydrodynamic volume and decreasing clearance rates by renal filtration, all of which results in an increase in the serum half-life of modified molecules *in vivo*. The majority of PEG applications involve the use of long-chain linear or branched PEG polymers having molecular masses of at least 2.5 kDa to over 50 kDa. [12]

PEG is typically made from ethylene oxide by an anionic ring-opening reaction, which results in long polymer molecules, consisting of the general structure $HO(CH_2CH_2O)_nH$. The number of repeat units (n) in standard commercial PEG polymers created through this process can typically be anything from less than 50 to over 1100.[12] PEG, being a synthetic polymer, is polydispersed and, even in the best of cases, a polydispersivity value (M_w/M_n) ranging approximately from 1.01 for low molecular weight oligomer (3-5 kDa), to 1.2 for high molecular weight (20kDa) may be expected. This polydispersivity is a negative property since it is reflected in polydispersivity of the conjugates. Thus the conjugates have slightly different physicochemical properties. Nowadays polydispersivity is unavoidable, so the only thing that can be done is to measure it and take it into account when we are using PEGylated silk. At present, the best way for testing the M_w/M_n ratio is Matrix Assisted Laser Desorption/Ionization (MALDI) mass spectroscopy, but Gel Permeation Chromatography (GPC) with a refractive index detector is also useful.

Monomethoxylated form of PEG is generally used in protein conjugation, since its monofunctionality yields cleaner chemistry. However, a certain amount of PEG diol is always present, in the range of 1-10%, depending upon the molecular weight (lower for low-mass PEG).[13] High diol concentration will yield unwanted cross-linked conjugates. Since most applications of PEG conjugation involve labile molecules, the coupling reaction generally requires mild chemical conditions. In case of polypeptides, the most common reactive groups involved in coupling are nucleophiles with the following decreasing rank order of reactivity: thiol, alpha amino group, epsilon amino group, carboxylate, hydroxylate. However, this order is not absolute, since it depends also on the reaction pH, furthermore other residues may react in special conditions, as the imidazole group of histidine. The thiol group is rarely present in proteins, furthermore it is often involved in active sites. The carboxylic groups cannot be easily activated without having reaction with the protein amino groups, to yield intra or inter molecular cross linking. Therefore, amino groups,

namely the alpha amino or the epsilon amino of lysine (*Fig.2*), are the usual sites of PEG linking. Furthermore, the other basic amino acids, arginine (*Fig.3)* and histidine (*Fig.1*), can be also interesting in terms of PEG-linking. As a consequence, the problem is to transform the hydroxyl terminal group of PEG to an activated one.

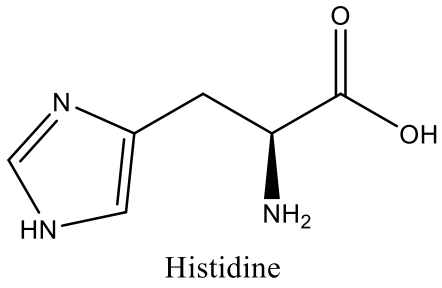


Figure 1 Histidine (His or H) structure.

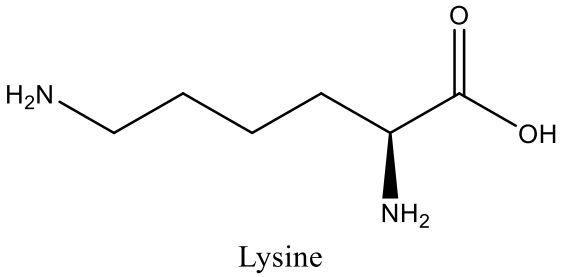


Figure 2 Lysine (Lys or K) structure.

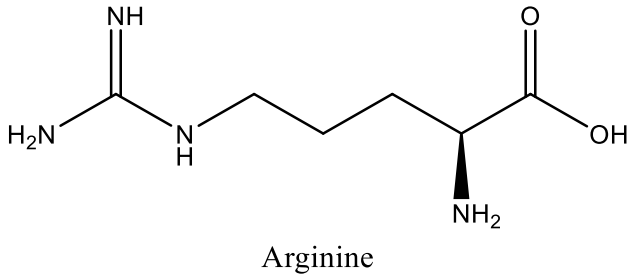


Figure 3 Arginine (Arg or R) structure.

There are plenty of techniques to make PEG-silk bioconjugates using basic amino acids as linking points in the silk. It may be divided into two important classes depending on the type of activated PEG that is used.

2.2.1 Alkylating PEGs

These products do not modify the charge of amino residues. Among these stands *PEG-aldehyde (Fig.4)* that gives a permanent linkage after Shiff base formation followed by *cyanoborohydride* reduction. Best reaction takes place with α- and ε-amino groups in protein at neutral or mild alkaline pHs, while good selectivity for the α-amino terminal amino acid at pH, 5/6, was found[14]. This is a convenient way for conjugation when the amino positive charge is critical for the retention of biological activity. However, the reaction rate for the Shiff base formation is relatively low, sometimes up to a day is necessary to reach the completion, with consequent inactivation of labile molecules. The reaction pH is critical for selectivity, α-amino terminal modification can be achieved around pH5.[13]

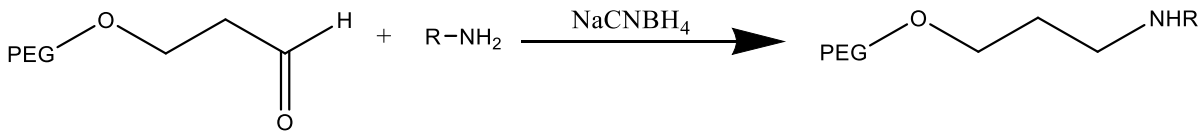


Figure 4 PEG-aldehyde reactivity.

2.2.2 Acylating PEGs

Most of these are *hydroxysuccinimidyl esters* (-OSu) of carboxylated PEGs *(Fig.5)*. These compounds are highly reactive towards amino groups, but the reaction rate may change significantly from the X structure. The distance between the active ester (-COOSu) and the last PEG ether can vary in different available products, by up to four methylene units, and this has profound influence on the reaction towards amino groups as well towards water. As an example, the $t^{1/2}$ hydrolysis rate of PEG-O-CH_2- CH_2- CH_2-COOSu is 23h, while that of PEG-O- CH_2-COOSu is 0.75h[13].

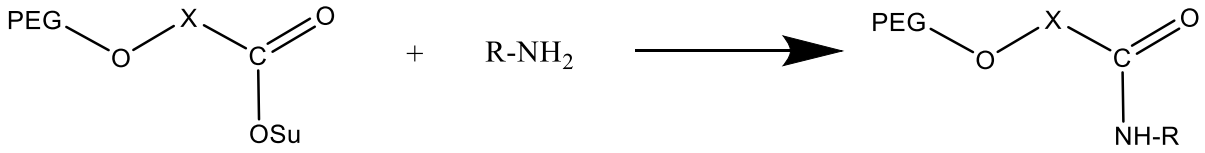


Figure 5 Hydroxysuccinimidyl carboxylated PEG´s reactivity.

Although all of these facts, due to the availability of reagents in the research group that were used in prior investigations, the PEG that have been spent in this project is *methoxypolyethylene glycol acetic acid (M_w= 5000) (Fig.6)*.

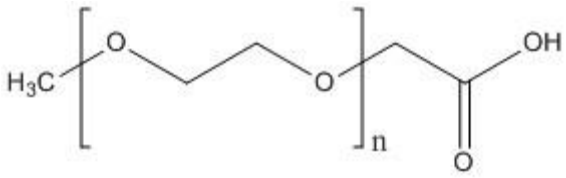


Figure 6 Methoxypolyethylene glycol acetic acid.

It is worth to say that the choice of this PEG have not been random, it was completely meditated. As it is said before, methoxy group provided a cleaner chemistry because it gives one sided functionalized PEGs, and an acid group with just one methylene give us the perfect reaction point to graft the PEG to silk. The only step that is needed to form and amide linkage made by the condensation of silk's primary amine with this carboxylic acid is an NHS ester activation using a carbodiimide as a crosslinker.

2.2.3 Zero length crosslinkers

The smallest available reagent systems for bioconjugation are the so-called zero-length crosslinkers. These compounds mediate the conjugation of two molecules by forming a bond containing no additional atoms. Thus, one atom of a molecule is covalently attached to an atom of a second molecule with no intervening linker or spacer.

Carbodiimides (RN=C=NR') are unsaturated compounds with an allene structure and are the most popular type of zero-length crosslinker in use. They are used to mediate the formation of amide linkages between carboxylates and amines or phosphoroamidate linkages between phosphates and amines[16]. Moreover, they are efficient in forming conjugates between two protein molecules, between a peptide and a protein, between an oligonucleotide and a protein, between a biomolecule and a surface or particle, or any combination of these with small molecules.

There are two basic types of carbodiimides: water soluble and water insoluble. The water soluble ones are the most common choice for biochemical conjugations, because most macromolecules of biological origin are soluble in aqueous buffer solutions. Not only is the carbodiimide itself able to dissolve in the reaction medium, but the by-product of the reaction, an isourea, is also water soluble, facilitating easy purification. Water insoluble carbodiimides, by contrast, are used frequently in peptide synthesis and other conjugations involving molecules soluble only in organic solvents. Both the organic soluble carbodiimides and their isourea by-products are insoluble in water.

2.2.4 NHS Ester Activation and Coupling

1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) (*Fig.7)* is the most popular carbodiimide used for conjugating biological substances containing carboxylates and amines. In fact, it may also be the most frequently used crosslinking agent of all [16]. Its application in particle and surface conjugation procedures along with *N-hydroxysuccinimide* (NHS) or sulfo-NHS is nearly universal[17] and this fact makes it the most common bioconjugation reagent in use today. EDC is water soluble, which allows for its direct addition to a reaction without prior organic solvent dissolution. Both the reagent itself and the *isourea* formed as the by-product of the crosslinking reaction are water soluble and may be removed easily by dialysis or gel filtration. However, the reagent is labile in the presence of water, especially in acidic solutions. In alkaline solutions the reagent is more stable, but it reacts much more slowly.

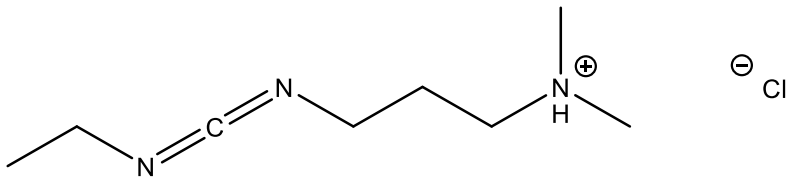


Figure 7 Structure of the EDC.

Mechanism (*Fig.8*)

N-substituted carbodiimides, such as EDC, can react form an amide bond in aqueous solution, but this involves a number of potential side reactions that can occur in addition to the desired conjugation product. The preferred reaction route is facilitated first by protonation of one of the nitrogen on the imide group of EDC, which results in the formation of an intermediate carbocation on the central carbon atom **(1)**. At this point, the modified carbodiimide can itself hydrolyse to form an inactive *isourea* that no longer can participate in the reaction process **(2)**. It can also react with an available ionized carboxylate group to create the desired *O-acylisourea* reactive ester intermediate **(3)**.

This ester again can accept another proton to form a second carbocation on the central carbon atom, and it is this form of the reactive ester **(4)** that can go on to react with an amine to create an amide bond **(5)**. However, this intermediate also can undergo a hydrolysis event to yield the same undesired *isourea* derivative **(2)**, hydrolysis by water is the major competing reaction, cleaving off the activated ester intermediate, again making it unreactive[18]. Therefore, there are two stages in the reaction sequence in which hydrolysis can occur and inhibit the desired product formation. If hydrolysis does not occur, there are at least three subsequent reactions that can happen, including the desired amide bond formation **(5)**.

If a neighbouring carboxylate group is in close proximity to the O-*acylisourea* ester, it may react with it and form an anhydride intermediate **(6)**. This especially can occur in polymers containing repeating carboxylate groups, such as in *polymethacrylate*, where the primary intermediate reactive group formed from EDC may be anhydrides.

Fortunately, an anhydride is also reactive with amine groups, so the desired amide bond formation can still occur with at least one of the two carboxylates making up the anhydride. Indeed, anhydride formation may result in higher yields of amide bond formation in certain instances. In addition, if EDC is in large excess over the amount of carboxylates present, then the intermediate ester may exist for a longer period and potentially it can rearrange by reacting with the neighbouring secondary amines in the carbodiimide and thus form an N-acylisourea derivative **(7)**, which is inactive and permanently attaches the EDC derivative to the carboxylate compound. Finally, *O-acylisourea* can also become cyclic producing an *oxazolone* **(8)**, which can also react with an amine group to yield the desired amide bond. *Oxazolones* can lead to the epimerization of the amino acid depending on which face of the cycle the amine attacks. The propensity for EDC to undergo side reactions may be a reason that high variability has been reported using the carbodiimide for particular conjugation reactions.

Reaction conditions

EDC amide bond formation reactions results [18,19]indicate that carboxylate activation occurs most effectively with EDC at pH 3.5-4.5, while amide bond formation occurs with highest yield in the range of pH 4.7 to 6. However, EDC hydrolysis occurs maximally at acidic pH values with increasing stability of the carbodiimide in solution at or above pH 6.5. When working with proteins and peptides, experience indicates that EDC-mediated amide bond formation effectively occurs between pH 4.5 and 7.5.

However, conjugations performed under mildly alkaline pH conditions (e.g., pH 8.5) can also be carried out to limit the polymerization of proteins, while still facilitating the coupling of a carboxylate-containing molecule at a lower substitution level per protein even though the coupling reaction occurs more slowly with lower yields.

Some procedures recommend the use of water as the solvent in an EDC reaction, while the pH is maintained constant by the addition of HCl. Buffered solutions are more convenient, because the pH does not have to be monitored during the course of the reaction. For acidic pH conjugations, MES *[2-(N-morpholino) ethane sulfonic acid]* buffer at 0.1M works well. When carrying out neutral pH reactions, a phosphate buffer at 0.1M is appropriate. [20].

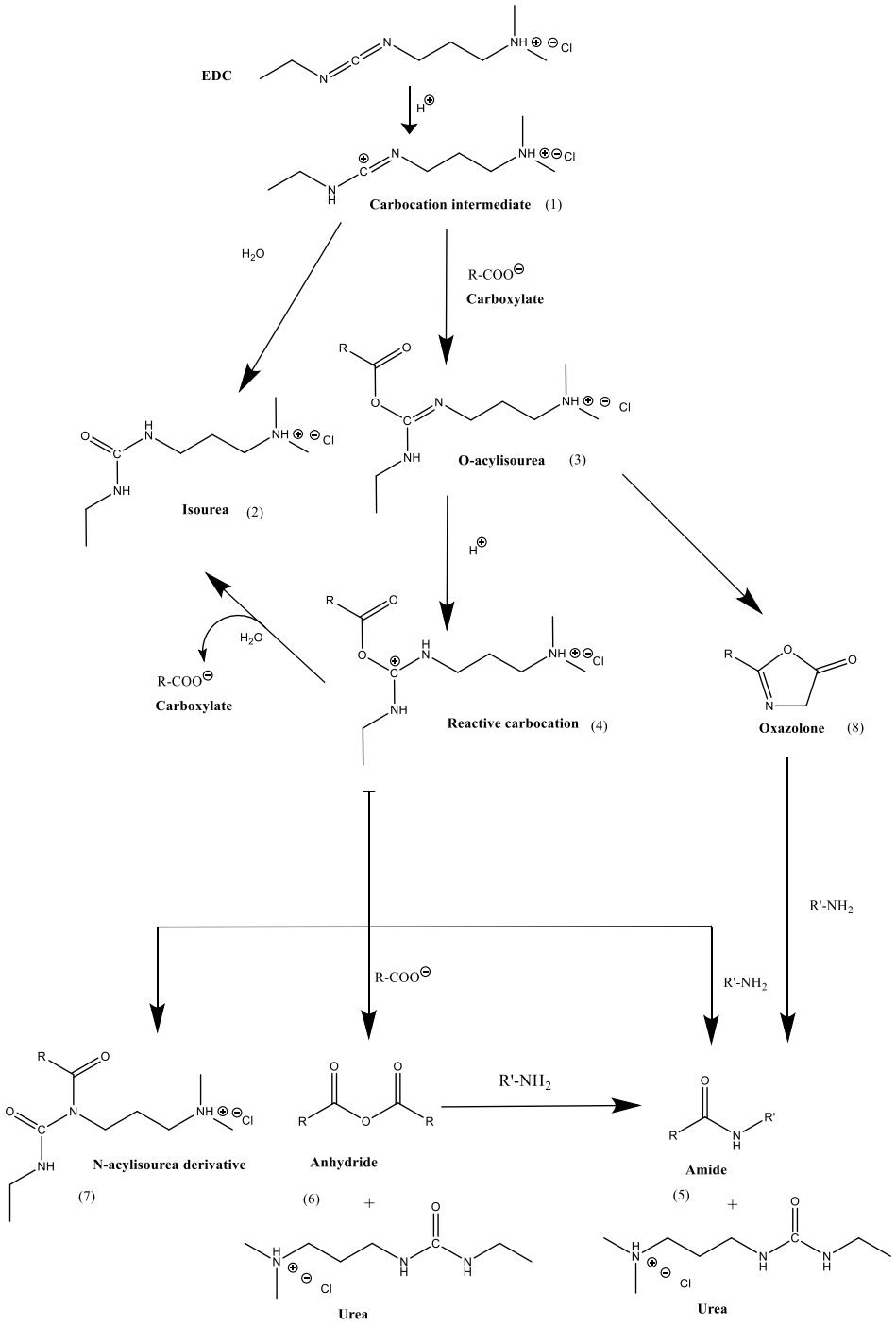


Figure 8 Amide bond formation mechanism via EDC coupling.

2.3 I-TASSER

Once silk had been functionalised with the mPEG acetic acid by EDC/NHS coupling chemistry next project step was to functionalise a small fragment of the LC fibroin, such as a small peptide, and checked if the properties were preserved. The decision of which peptide was suitable for being synthesized and modified afterwards requires an in deep knowledge of protein folding. To help this purpose a previously obtained amino acid sequence of both LC-SF and HC-SF, acquired by the research fellows using amino acids analysis technique, was used. LC-SF contains 262 amino acids which means that there are 261 amide bonds with 3 angles, phi(ϕ), psi(ψ) and omega(ω) that describes them. 783 different numbers that just give information about the orientation of the main carbon chain, without taking into account hydrogen bonds and steric hindrance from residue side chains… All of this data are beyond human power calculation by itself. A system that helps to plot this amino acid sequence in a three-dimensional (3D) structure is needed. There are some services that can provide secondary structure with an amino acid sequence given. The best of them is Iterative Threading Assembly Refinement (ITASSER), number 1 in different community-wide blind experiments of protein prediction such as Critical Assessment of Techniques for Protein Structure Prediction (CASP) in 2016.

Functional characterization of a protein is often facilitated by its 3D structure. However, the fraction of experimentally known 3D models is so tiny due to the inherently time-consuming and complicated nature of structure determination techniques. One way to narrow the gap between proteins with known sequences and proteins with experimentally characterized structure and function is by developing advanced computational approaches for modelling structure and function from sequences.

I-TASSER is a composite of model quality assessment programs that try to approach tertiary and secondary protein structures as well as the main function of them to non-expert scientific community public in order to ease their studies. I-TASSER request just the amino acid sequence to give a five-model results of how your protein look like, the structure, the accessibility of the different residues among other different parameters that assess the quality of the prediction.

Is based in force fields calculation methods and search algorithms that compare extensive data banks of well-known structures like Protein Data Bank (PDB). This features combined bring us to template-based protein structure modelling techniques[22].

Template-based protein structure modelling techniques rely on the study of principles that dictate the 3D structure of natural proteins from the theory of evolution viewpoint. Template-based approaches to structure prediction have their advantages and limitations. Comparative protein structure modelling usually provides high-quality models that are comparable with low-resolution X-ray crystallography or medium-resolution NMR solution structures. However, the applicability of these approaches is limited to those sequences that can be confidently mapped to known structures.

All current comparative modelling methods consist of five sequential steps:

1. To search for proteins with known 3D structures that are related to the target sequence.
2. To pick those structures that will be used as templates
3. To align their sequences with the target sequence of amino acids.
4. To build the model for the target sequence given its alignment with the template structures
5. To evaluate the model using a variety of criteria.

However, I-TASSER is not just one technique, as it is said before is a composite of them. I-TASSER process, as a template-based technique, consists in five general steps: threading template identification, iterative structure assembly simulation, model selection and refinement, and structure-based function annotation and estimation of prediction accuracy (*Fig.9*).

2.3.1 Threading template identification

First step of the procedure consists in identify similar protein sequences or similar structural motif that will work as a template to our problem sequence (query). This is performed by a basic local alignment search tool (BLAST)[23] that uses PDB as a library of sequences. More specifically it uses position-specific iterated BLAST (PSI-BLAST). It is a BLAST with "gaps" that allows a percentage of variation in the amino acid sequence. From the sequences and the amino acid frequencies obtained from this threading the program makes another homolog alignment. The process is repeated until the program does not find new homolog alignments. After that, the query sequence is threaded through PDB using a local metha-threading server (LOMETS)[24], then the templates are ranked by a variety of sequence-based and structural-based scores. The top template hits form each threading program are selected for further consideration. The quality of the template alignments is judged based on the statistical significance of the best threading alignment, i.e., the Z-score, which is defined as the energy score in standard deviation units relative to the statistical mean of all alignments.

2.3.2 Structural assembly

In the second stage, continuous fragments in the threading alignments are excised from the template structures. These are used to assemble structural conformations of the sections that aligned well, with the unaligned regions (mainly loops/tails) built by *ab initio* modelling[25,26]. To improve the efficiency of conformational search, I-TASSER adopts a reduced model to represent the protein chain, with each residue described by its Cα atom and side-chain centre of mass. Then a new threading process starts. During this one, the fragment assembly is performed using a modified replica-exchange Monte Carlo simulation technique[27], which implements several replica simulations in parallel at different temperatures, with the temperatures periodically exchanged between the replicas. The overall simulation is guided by a composite knowledge-based force field that includes: (1) general statistical terms derived from the PDB (Cα/side-chain correlations, H-bonds and hydrophobicity); spatial restraints from threading templates; and sequence-based contact predictions from an intra-residue contact prediction program of the same server (SVMSEQ)[28]. Conformations generated in the low-temperature replicas are clustered in order to find low energy states.

2.3.3 Fragment assembly simulation

In the third step, the fragment assembly simulation is performed again starting from selected cluster centroids. External constraints are grouped from the LOMETS threading alignments and the PDB structures that are structurally closest to the previous cluster centroids, parameters like the nuclear Overhauser effect (NOE) are taken into account. This second iteration is made to remove steric clashes and refine the global topology of the centroids. Final structural models are chosen after an optimization of hydrogen bonding networks.

2.3.4 Structure-based functional annotation

Finally, structure is inferred by structurally matching the predicted models against well-known structures. For this purpose, three protein structure/function libraries have been constructed independently and biweekly updated. Although the global structural similarity search is used for recognizing proteins with similar global fold, the local similarity search provides a complementary method, identifying analogues that have a different fold but perform similar function because of the conservation of active/binding sites. The functional analogues from the global search results are ranked based on the conserved structural patterns present in the model. The local similarity search looks for conserved spatial motifs in the predicted I-TASSER model, with the candidates ranked based on their structure and sequence similarity to functional cavities (binding pockets) in known structures. Finally, the results from the global and local search are combined to present a comprehensive list of functional analogues.

2.3.5 Estimation of prediction accuracy

Assessing the quality of a prediction is important because this assessment eventually determines how biologists will use the prediction in their research. For estimating the accuracy of the structure predictions, a confidence score (C-score)[29] is defined based on the quality of the threading alignments and the convergence of the I-TASSER's structural assembly refinement simulations. C-score takes into account different parameters such as root mean square deviation (RMSD) from models to the cluster centroid, among others.

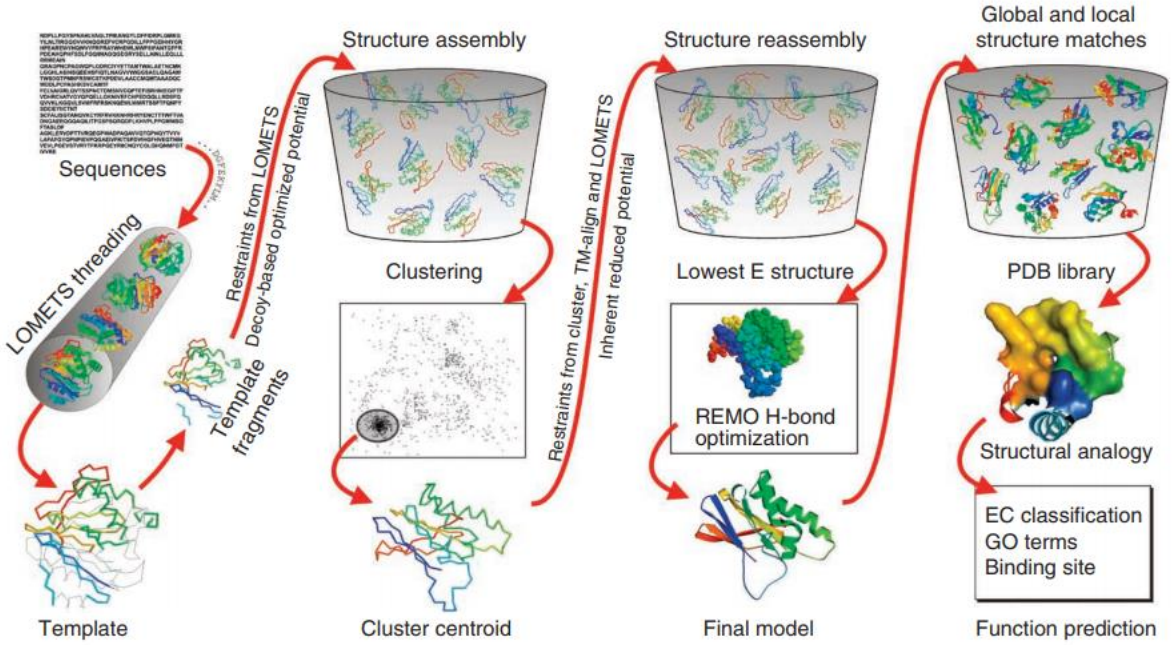


Figure 9 ITASSER structure prediction process

After all the structural and sequence study the next step is to synthesise the selected peptide.

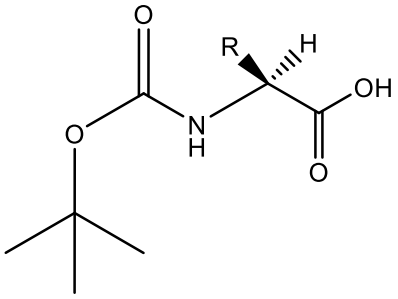
2.4 PEPTIDE SYNTHESIS

2.4.1 Peptide synthesis principles

Solid phase peptide synthesis (SPPS), originally developed by Bruce Merrifield in 1963 [30], has become the major technique for the synthesis of peptides. The growing peptide chain is assembled on a solid support, typically a polystyrene bead, allowing for the easy removal of reagents at each step during the synthesis.

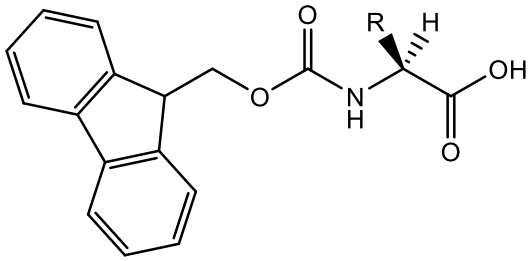
In solid phase peptide synthesis, the chain is built form C-terminus to N-terminus (unlike natural biological protein synthesis, which occurs from N-terminus to C-terminus). The amino acids are added to the chain one-by-one by reacting the free amine of the growing chain with the free carboxylic acid of the incoming amino acid.

To prevent unwanted reactions, the amine of the incoming amino acid is masked with a protecting group. There are two main strategies for peptide synthesis, named for the amine protecting group used in each: Boc and Fmoc. Fmoc synthesis utilizes the base-labile *9-fluoromethyloxycarbonyl (Fig.10)* protecting group. Boc synthesis utilizes the acid-labile *t-butoxycarbonyl (Fig.11)* protecting group.



t-butoxycarbonyl amino acid

Figure 11 Boc-protecting group structure.



[(9-fluorenylmethyl)oxy]carbonyl amino acid

Figure 10 Fmoc-protecting group structure.

The growing peptide chain is connected to the resin through a linker group. There are a wide variety of linkers available; the specific linker used for a given synthesis depends on the protection strategy as well as desired C-terminal modifications. For Fmoc syntheses, the linkers must be base-stable so that peptide is not removed from the resin during Fmoc removal. Certain linkers also allow for the removal of the peptide as the C-terminal amide instead of the free acid. The C-terminal amide is useful for peptides that will be used for biological testing as this protects them for degradation by endopeptidases (enzymes found in cells and tissues that degrade peptides and proteins).

In addition, when performing solid phase peptide synthesis, the R groups of the amino acids are masked with protecting groups. These groups prevent unwanted side reactions with the functional groups of the amino acid sidechains during the synthesis reactions. The particular side chain protecting groups depend on the amine protection strategy.

Since the Fmoc group is base-labile, base-stable (acid-labile) protecting groups are used on the side chains so that the protecting groups remain intact throughout the synthesis. Conversely, since the Boc group is acid-labile, acid-stable (base-labile) protecting groups are used for Boc synthesis.

In general, Fmoc synthesis is the more common synthetic strategy used. The cleavage step (the removal of the finished peptide from the resin) for Boc synthesis requires the use of hydrofluoric acid on a specialized cleavage manifold. With Fmoc chemistry, cleavage is normally done using trifluoroacetic acid (TFA), and can be performed in standard laboratory glassware in a fume hood.

2.4.2 The chemistry of peptide synthesis

Solid phase peptide synthesis is accomplished through the repetition of four main reactions. First of all, a coupling of the first amino acid to the resin, then the deprotection of the N-terminus followed by coupling of the incoming amino acid and finally a final cleavage step to liberate the peptide from the solid support. As this project has been done using Fmoc protected amino acids, discussion of the specific reactions involved will focus on Fmoc synthesis.

2.4.3 Amino acid coupling to resin

The first step (*Fig.12)* in this process is to attach the first amino acid to the resin in order to have just one binding site. There are two type of resins: polystyrene resins and ProTide ® resins. The latter ones are used for very long and difficult peptides, like our case. Based on PEG and polystyrene (PS) core with *trytilchloryde* derivative (TC) linkers, the previous necessity for purchasing resins with pre-loaded C-terminal amino acids is eliminated. The C-terminal amino acid reacts with the linker-chloride, in the presence of *potassium iodide* (KI), with an amine, like *diisopropyethylamine* (DIEA) in a microwave irradiation medium[31].

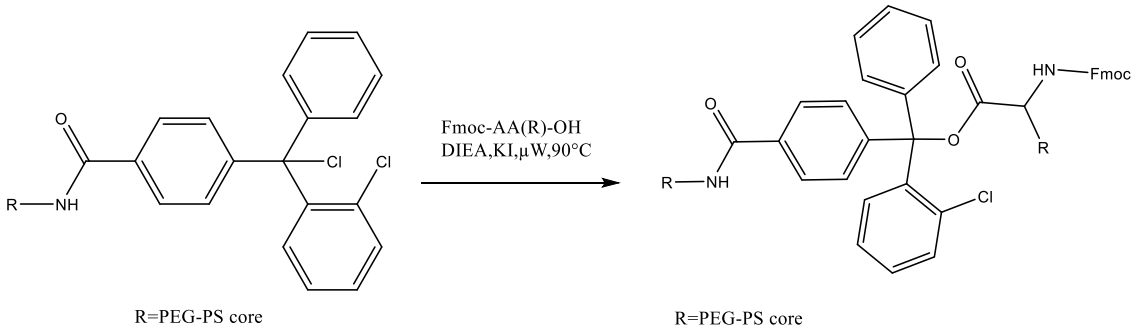


Figure 12 Amino acid -Resin coupling reaction.

Deprotection: The second step in any Fmoc coupling is the removal of the Fmoc group from the growing peptide chain (*Fig.13)*. Typically, Fmoc removal is accomplished with *piperidine*, although other reagents such as *piperazine* may also be used. *Piperidine* reacts with Fmoc-protected amino acid to remove the Fmoc group (which is converted to *dibenzofulvene*) yielding the free amine following decarboxylation.

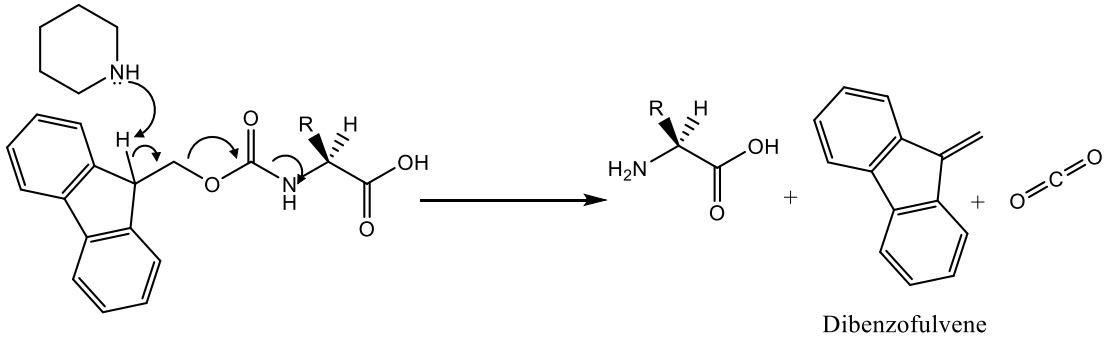


Figure 13 Deprotection step reaction.

The lone hydrogen residing on the α-carbon of the Fmoc group is very acidic due to the electron withdrawing effect of the *fluorene* ring system[32]. This proton at the 9-position of the *fluorene* ring system is therefore susceptible to removal by base during peptide synthesis.

Removal of the proton results in a β-elimination mechanism to generate a reactive *dibenzofulvene* intermediate, CO_2, and a deprotected amine group.

Carbodiimide coupling: The reaction of the incoming amino acid with the growing peptide chain to form an amide bond is accomplished by converting the acid into an activated form. The most commonly used and studied activation method for peptide synthesis have been based on the use of carbodiimides. The *carbodiimide* that was used in the peptide synthesis *was N,N'-diisopropylcarbodiimide* (DIC), unlike the case of silk PEGylation, that EDC was chosen because of its previously explained features.

Furthermore, additional *hydroxybentzotriazole (HOBt)* can be added in order to improve the c*arbodiimide* activation method. *HOBt* quickly converts the *O-acylisourea* into a OBt ester that is highly reactive with the amine group and avoids undesirable *N-acylisourea* and *oxazolone* formation. Other reagents, like *1-hydroxy-7-azabenzotriazole* (HOAt), has been tested as additives to *carbodiimide* coupling method obtaining better coupling yields[33]. The additive selected for this project was *[ethyl 2-cyano-2-(hydroxyimino)acetate]* (Oxyma®) due to the lower risk of explosion at high temperatures if we compare it to the previous mentioned reagents [34].

Cleavage and peptide isolation: In Fmoc synthesis, the removal of the peptide from the solid support is typically accomplished with *trifluoroacetic acid* (TFA). Because the side chain protecting groups used in Fmoc synthesis are acid labile, a single step both cleaves the peptide from the resin and removes the protecting groups. Various scavenger molecules are added to the TFA to prevent the cleaved protecting groups from reattaching to the peptide. The particular scavengers used depend on the specific peptide sequence. Common scavengers include water (scavenges t-butyl cations), *triisopropylsilane* (TIS, scavenges *trityl* and Pbf cations), *dioxa-1,8-octane-dithiol* (DODT, scavenges t-butyl cations, reduces oxidation of Cys/Met side chains), among others. The cleavage solution chosen for this project was a mix of different agents, TFA, TIS, DODT and water.

Following cleavage, the completed peptide is typically isolated by ether precipitation. The cleavage solution is filtered to remove the spent resin, and an 8-10 fold excess (by volume) of ice-cold ether is added to precipitate the peptide. The peptide will crash out of solution as a white or off-white precipitate, while the TFA, scavengers, and cleaved protecting groups will remain in solution. The peptide can then be recovered by centrifugation and decanting.

2.4.4 Microwaves in peptide synthesis

Since its introduction, microwave energy has demonstrated an ability to increase reaction rates, decrease side reactions, and allow for more solvent choices. The first reported use of microwave energy for peptide synthesis was reported in 90s using a domestic multi-mode microwave oven. In 2002[35], the application of microwave energy toward peptide coupling was again investigated using various onium salt activators at temperatures greater than 100°C. This study used a new generation single-mode microwave cavity that had recently become available and allowed for a more uniform and repeatable microwave field compared to domestic microwave ovens. In 2005, the first fully automated microwave peptide synthesizer became available. Since this time, the use of microwave energy for peptide synthesis has grown dramatically and several reviews on the technology have been published[36].

2.4.5 Side reactions

There are few side reactions that may occur during peptide synthesis. It is important to take them into account in order to understand further characterization results.

Racemization: The properties of proteins and peptides are critically dependent on the configuration of their chiral centres. The alteration of a single chiral centre can have a drastic effect on biological activity. With the exception of *glycine*, all 20 standard amino acids contain a chiral center at the α-carbon atom. Additionally, *isoleucine* and *threonine* contain a chiral center in their side-chains. In SPPS, racemization has been extensively documented during the coupling reaction, but also can occur during base catalysed deprotection steps. During the coupling reaction epimerization can occur through formation of an *oxazolone* intermediate, as it is explained before, or direct enolization. In conventional SPPS, *cystidine* and *histidine* appear to be the most susceptible to epimerization[37].

During coupling, conversion of the incoming amino acid to an activate ester increases the acidity of the α-carbon. This can tend to favour enolization, which leads to re-arrangement about the α-carbon. In SPPS, sterically hindered tertiary amines are used in an effort to minimize base-catalyzed removal of the α-carbon. Higher temperature coupling reactions by microwave or conventional heating have been shown to increase epimerization of *cysteine* and *histidine* residues[38].

Aspartimide formation: One of the most common side reactions in Fmoc SPPS is base catalyzed *aspartimide* formation that occurs during each deprotecting step (*Fig.14)*. This side reaction involves attack by the nitrogen atom attached to the α-carboxy group of either aspartic acid or asparagine on the side chain ester or amide group, respectively. Nucleophilic attack then causes subsequent ring opening, which gives rise to a mixture of α-aspartyl and β-aspartyl peptides.

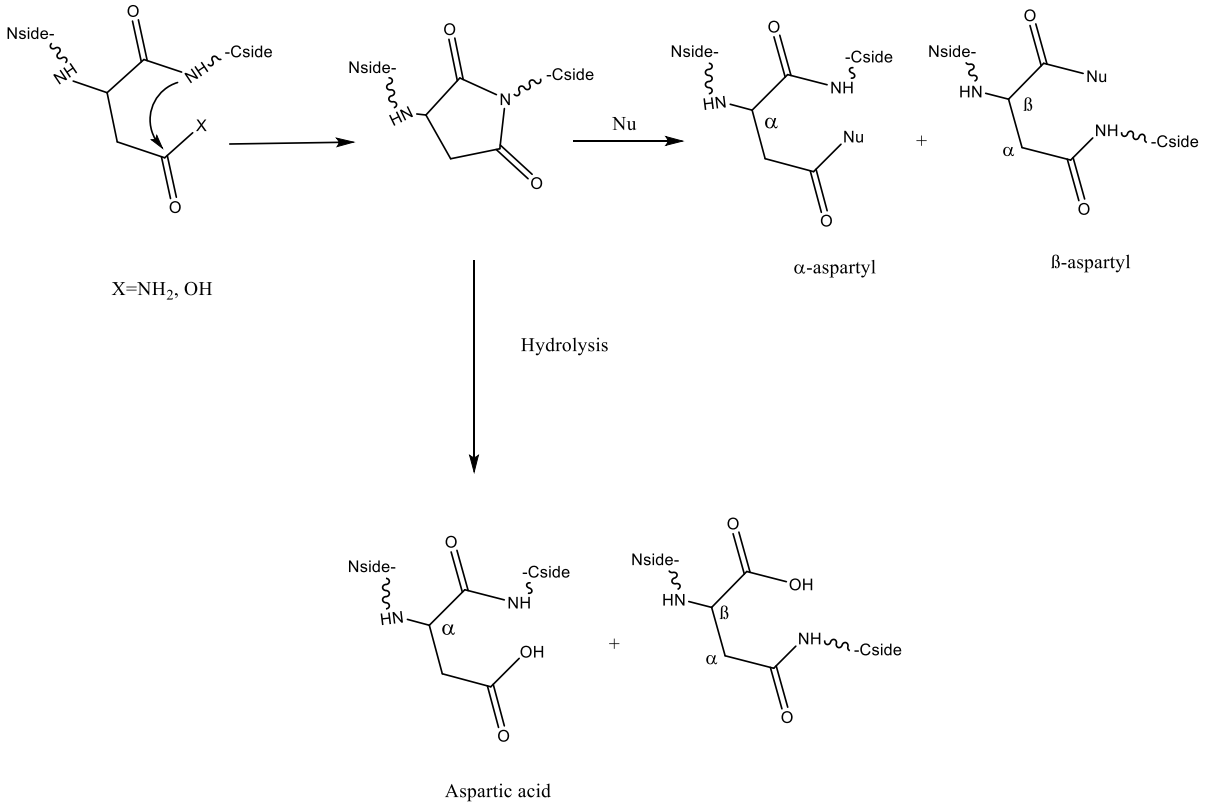


Figure 14 Aspartamide formation reaction.

Aspartimide formation has been shown to occur in sequences containing the "Asp-X" moiety, where X = [Gly, Asn, Ser,Thr]. Each subsequent deprotection cycle after the "Asp-X" sequence further increases aspartimide formation, resulting in a potentially serious problem in longer peptides with multiple Asp residues. The Asp-Gly segment in particular is the most susceptible for aspartimide formation due to the lack of steric hindrance from glycine to inhibit this reaction[39]. Inclusion of the *β–tert-butyl ester* protection on the aspartic acid side chain will reduce, but not eliminate *aspartimide* formation due its bulkiness.

The use *piperazine* in place of *piperidine* has demonstrated significantly lower levels of *aspartimide* formation[40]. *Piperazine* is advantageously a non-controlled substance unlike *piperidine*, which as a precursor for the synthesis of *phenylcyclidine* (Angel Dust), the use of it is regulated by the authorities (you need to indicate how much do you spend and make a report of it). However, *piperazine* will lead to a slower deprotection, due to its pKa of 9.8 compared to 11.1 for *piperidine*, and for conventional synthesis of hydrophobic sequences can lead to more incomplete Fmoc removal.

3 MATERIALS AND METHODS

3.1 Bombyx mori Silk Fibroin

The silk materials were prepared using reported procedures[2]. Briefly, silk cocoons provided by the laboratory supplies, were degummed using 0.5% sodium carbonate solutions and water washing. The separated silk fibers dried in air. The silk fibroin was regenerated by dissolving in lithium bromide solution (9M) to ~5% protein concentration, dialysing rapidly (<8h) against distilled deionized water to a conductivity of <10µS/cm and finally freeze-drying.

3.2 Methoxypolyethylene glycol 5,000 acetic acid >80% (mPEG-AcH)

mPEG-AcH acid was provided by Sigma-Aldrich®. Reference code 70718-1G-F. MDL number MFCD05665435.

3.3 PEGylated silk fibroin

Modified silk fibroin was obtained adapting reported procedures[13-16]. In brief, ~20mg (1mg/mL) regenerated silk was soaked in 20mL PBS buffer (0.1M NaH_2PO_4 and 0.5M NaCl) 30 min. After that, silk solution was activated with 10-fold excess of EDC (0.4 mols / 71µL) and the corresponding part of NHS (0.756 mols / 0.087 g) and stir the mix 15 min at room temperature. Then, add 200mg of mPEG-AcH and stir the mix 2h at room temperature.

3.4 Thin Layer Chromatography (TLC)

This technique was used to asses if the PEG functionalization process had been done correctly or not. Samples were dissolved on a volatile solvent were applied on a plate made of glass coated with silica gel (stationary phase) and let dry. A beaker with a solvent or solvent mixture (mobile phase) was prepared as well. After the sample had been applied on the plate was introduced in the beaker and the mobile phase started drawn up the plate via capillary action. Different samples rise at different rates due to their unlike polarities. Therefore, non-functionalized PEG's that have higher polarity due to the presence of the acid group, should have risen higher in the plate in the same time.

It is known that TLC is not the best technique to differentiate polymers due to the low mobility of these in the stationary phase because of their huge molecular weight. Even though, this technique was attempted to discern if there was still NHS unreacted in order to assign peaks in the NMR spectre. Samples

were taken at different reaction times to follow the path of it. Ethyl acetate, hexane, 10% methanol solution, Dichloromethane (DCM) and mixture of these were used as solvents on the mobile phase, but none of them could displace the polymers properly. Permanganate solution (10 g; 0.07 mol K_2CO_3 / 1.5 g; $9.5 \cdot 10^{-3}$ mol $KMnO_4$ / 1,25mL NaOH 10%; $3.4 \cdot 10^{-4}$ mol / fill up to 200 mL distilled water) was also prepared to notice a change in the samples dots in the plates, as the carboxyl dot was expected to disappear.

3.5 NMR (Nuclear Magnetic Resonance) Analysis

NMR was used to asses that PEGylating process of the silk fibroin was completed. ~20mg of each analysed component were dissolved in ~1ml of the proper solvent ($CDCl_3$, D2O, DMSO d-6) in a NMR tube. Solutions are stirred carefully to dissolve completely the samples. H^1, C^{13} and COSY analysis were made in a 400MHz equipment.

3.6 Peptide fragment of silk fibroin election

To decide which peptide was going to be synthesized it was needed to describe the secondary structure of the light chain from the silk fibroin protein. Therefore, found the main factors that help to stabilize structures such helixes, coils or turns was crucial. All the models obtained during this work have been obtained with I-TASSER webpage service. This service can predict secondary structure models from primary amino acid sequences. I-TASSER simulations generate a large ensemble of structural conformations, called decoys. To select the final models, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, using a RMSD cut-off between the query and the models, which is calculated iteratively until the first cluster includes less than 70% and more than 15% of all of the structures[41].

I-TASSER reports up to five models which corresponds to the five largest structure clusters. The confidence of each model is quantitatively measured by Checkerboard score (C-score)[42] that is calculated based on the significance of threading template alignments and the convergence parameters of the structure assembly simulations. C-score is typically in the range of [-5, 2], where a C-score of a higher value signifies a model with a higher confidence and vice-versa.

Template modelling score (TM-score)[43] and Root-mean-square deviation of atomic positions (RMSD) are standards for measuring structural similarity between two structures which are usually used to measure the accuracy of structure modelling when the native structure is known. In case where the native structure is not known, it becomes necessary to predict the quality of the modelling prediction, i.e. what is the distance between the predicted model and the native structures? To answer this question, ITASSER tried to predict the TM-score and RMSD of the predicted models relative the native structures based on the C-score.

TM-score is a recently proposed scale for measuring the structural similarity between two structures[41]. The purpose of proposing TM-score is to solve the problem of RMSD which is sensitive to the local error. Because RMSD is an average distance of all residue pairs in two structures, a local error (e.g. a misorientation of the tail) will give rise to a big RMSD value although the global topology is correct. In TM-score, however, the small distance is weighted stronger than the big distance which makes the score insensitive to the local modelling error. A TM-score >0.5 indicates a model of correct topology and a TM-score<0.17 means a random similarity. These cut-offs do not depend on the protein length. TM-score and RMSD are estimated based on C-score following the correlation observed between these qualities [44-45].

3.7 Construction of silk fibroin peptide fragment

The following construct was designed: 37mer (ACAAANVINSYTDGVRSGNFAGFRQSLGPFFGHVGQN). With a monoisotopic molecular weight of 3828,82 Da. This peptide is a fragment of the silk fibroin light chain from *Bombyx mori* cocoon silk. This protein fragment was chosen after the structure explained in this project and differs from those used in the studies by Tristan Giesa and Carole C. Perry[46]. Peptide was synthesized using HE-SPPS (High Efficiency Solid Phase Peptide Synthesis).This synthesis was performed by Liberty Blue™ Automated Microwave Peptide Synthesizer[47]. This equipment allows us to prepare our peptide by CarboMAX™ coupling chemistry (patented) and microwave deprotection and coupling (patented). All the amino acids used were Fmoc-protected. The activator used was N, N'diisopropylcarbodiimide (DIC), the activator base was ethyl cyanohydroxyiminoacetate (Oxyma). The solvent used for the main wash was N,N-dimethylformamide (DMF). The provider for all of these compounds was Sigma Aldrich®.

3.8 High performance Liquid Chromatography (HPLC)

This technique was used to purify the peptide once was prepared. Specifically, was an HPLC with Reverse Phase (RP) were the stationary phase used was a long hydrocarbon (C18) and the mobile phases used were: 95% Methanol for preparative chromatography, 0.1% Formic acid and double distilled water (10µS), which was called Solution A; 5% Methanol for preparative chromatography, 0.1% Formic acid and double distilled water (10µS), which was called Solution B. The columns used were a preparative AERIS™ 00B-4506-AN PEPTIDE XB-C18 100 (dimensions were 2.1x50 mm with a 1,7 µm particle size) and a semi-preparative ZORBAX™ 300SB-C18[48] (dimensions were 9.4 x 250 mm with a 5 µm particle size).

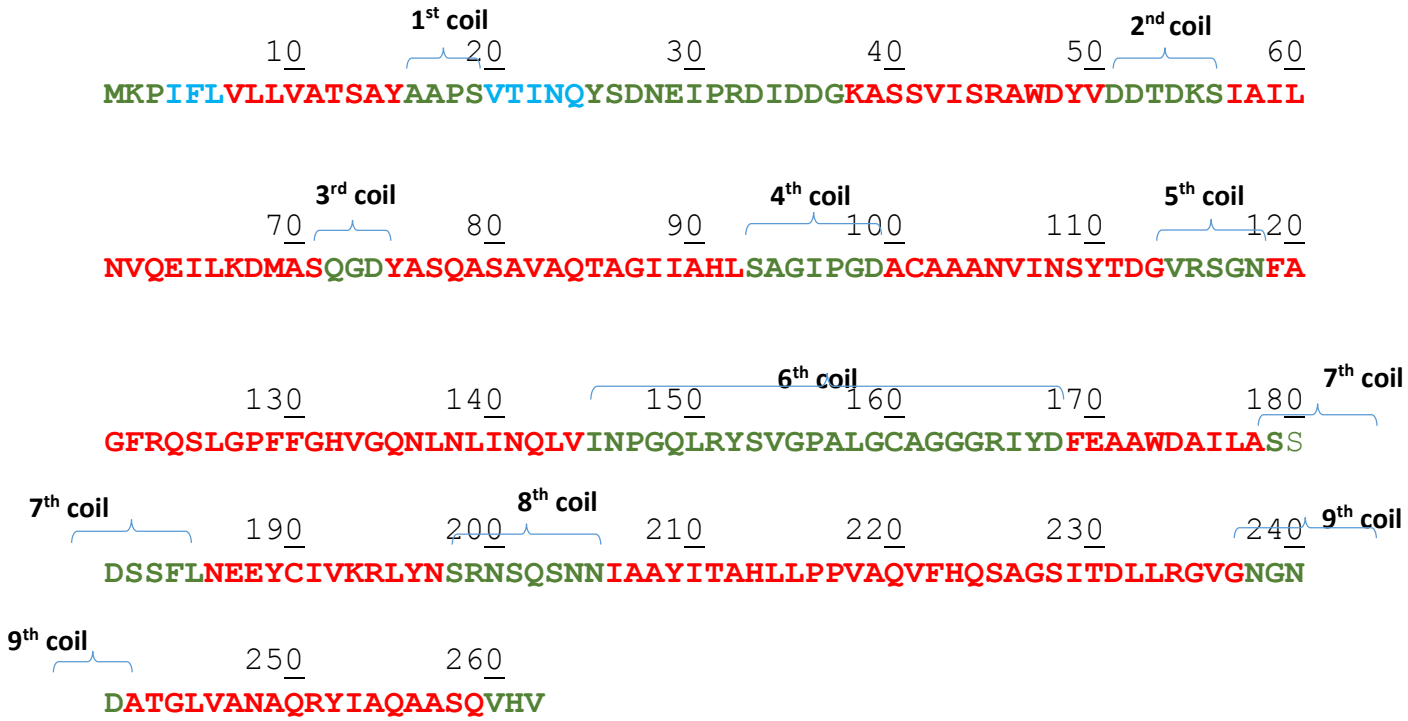
3.9 Circular Dichroism/Ultra-Violet (CD-UV)

This technique was used to check the composition in terms of secondary structure of the synthesized peptide. This may help to the characterization of the folding of the protein. Actually the calculatiosn that were made during the project were only theoretical, using the webpage service *Dichro Calc* University of Notthingham[49-50]. The supported chromophores for CD are: peptide bonds, aromatic side chains, among others. These parameters are derived from *ab initio* calculations.

4 RESULTS AND DISCUSSION – COMPUTATIONAL PART

4.1 Light chain sequence

This amino acid sequence corresponds to the light chain of the silk fibroin. From now onwards α-helixes have been written with red letters, blue for the β-sheets and green for the coils.



Important note:

All the figures in this section are labelled following this code to ease their locating and referencing:

AG-SLC-X-YYY(Z) NM Cs = number

This code stands for Alberto Giménez-Silk fibroin Light Chain.

X → Sequence fragment observed

"WS", "5", "9" stands for whole structure, 5th coil, and 9th coil, respectively

YYY → Amino acid length of the fragment.

Z → Features of the fragment

Is optional. "S" stands for symmetric in the case of coils and "A" for asymmetric.

N → ITASSER model number

M → Model.

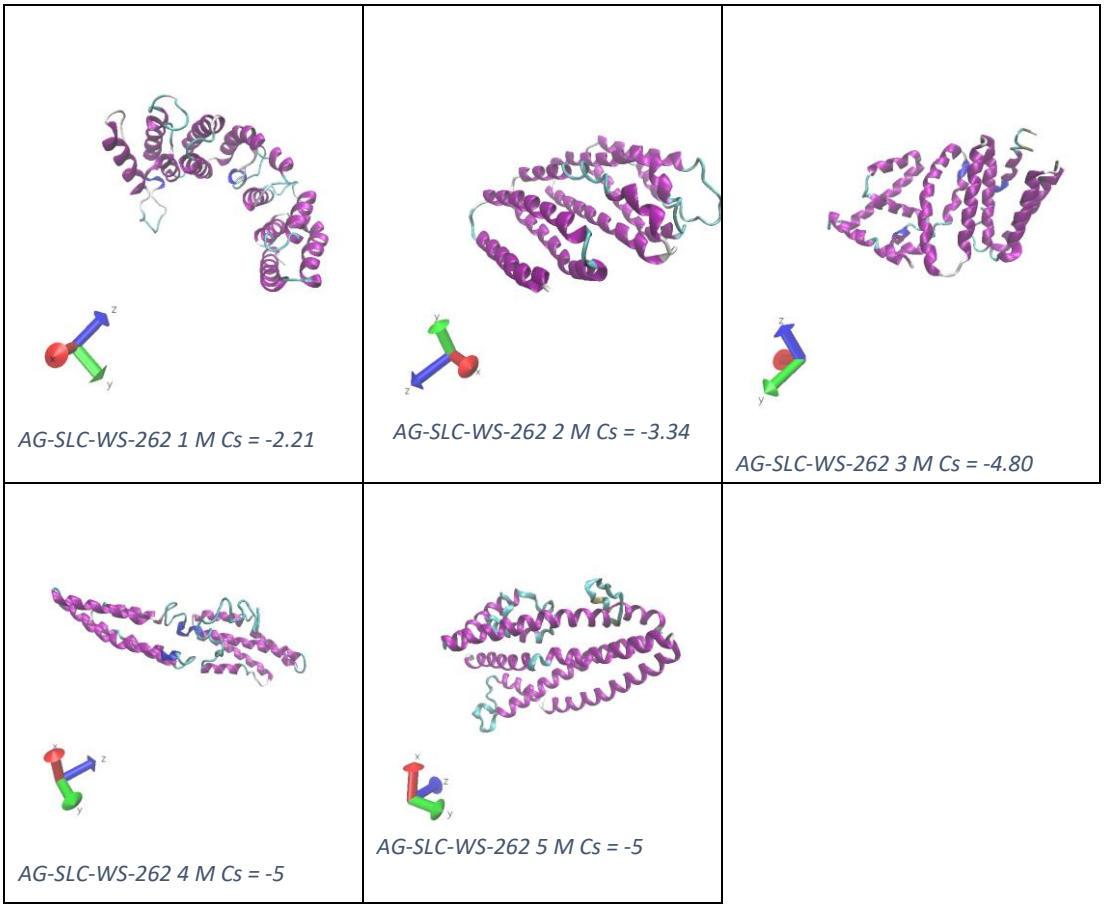
Cs→ C-score assigned by the ITASSER

4.2 Whole light chain models

Due to the presence of a variety of structures (α-helix, β-sheets or coils) in the sequence of the LC-SF we decide to focus just in one type of constructions, the coils. As it is said before, the presence of well-arranged β-sheets improves the toughness of the silk in the HC-SF, but this study is meant to be for the LC-SF. On the other hand, α-helixes were not studied due to the large amount of factors that can affect the formation of one single helix because it is bigger than just a coil. So it was decided that it was easier to study a single coil surrounded by other constructions.

ITASSER webpage service allow us to rotate the whole molecule and know which amino acid corresponds to the C- or N- terminal side of each structure. With this information we can check the position, the length and the differing structures that all the amino acids from the coils and its surroundings have formed. In most of the cases the amino acids that are supposed to form coils according to ITASSER adopt different structures. After a thorough analysis of each coil it has been concluded that the best coils to start doing calculations, due to their repetitively behaviour in terms of lengths, are the following (*see Table 1*).

- Val#114 and Asn#118 (5 aa long)
- Asn#238 and Asp#241 (4aa long)



AG-SLC-WS-262 1 M Cs = -2.21

AG-SLC-WS-262 2 M Cs = -3.34

AG-SLC-WS-262 3 M Cs = -4.80

AG-SLC-WS-262 4 M Cs = -5

AG-SLC-WS-262 5 M Cs = -5

Table 1 Coil´s length in the different models

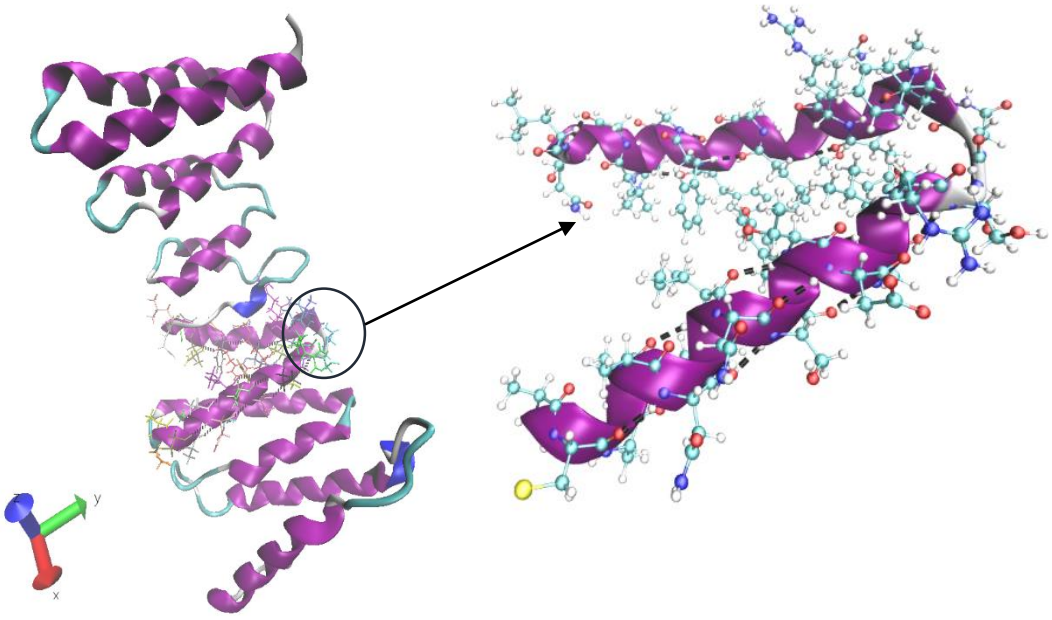
	5th Coil (Val#114 to Asn#118)		9th Coil Asn#238 to Asp#241	
Model 1	Ser#116	Gly#121	Asn#238	Thr#243
Model 2	Ser#116	Phe#119	Val#236	Asn#240
Model 3	Ser#116	Phe#119	Gly#237	Ala#242
Model 4	Ser#116	Gly#121	Gly#239	Thr#243
Model 5	Val#114	Gly#117	Asn#238	Asp#241

In order to visualize these models more easily Visual Molecular Dynamics program (VMD) was used. It provided more information about hydrogen bonding, interatomic distances, among others. It was worth noting that ITASSER and VMD does not provide the same data, e.g. 5th coil in ITASSER was 5 aa long (Val#114 to Asn#118) but with VMD it was just 3 aa long (Ser#116 to Asn#118). This fact does not mean that one of them is incorrect because they are using different calculation methods. ITASSER amino acids assignations were taken as the reference ones and VMD were used to compare structures.

For the next pages, the distance cut-off of the hydrogen bond is 3.2 Å and the angle cut-off is 30˚ by default if nothing else is indicated. This distance was chosen[51,52] because hydrogen bonds with donor-acceptor distances of 2.2-2.5 Å are categorized as "strong, mostly covalent", 2.5-3.2 Å as "moderate, mostly electrostatic", 3.2-4.0 Å as "weak, electrostatic". Energies are given as 40-14, 15-4, and, 4 kcal/mol respectively. Most H bonds in proteins are in the moderate category, strong H bonds require conditions that are rare within proteins. In α-helix structures the N-H group of an amino acid forms a hydrogen bond with the C=O group of the amino acid four residues earlier; this repeated *i → i+4* hydrogen bonding is the most prominent characteristic of an α-helix.

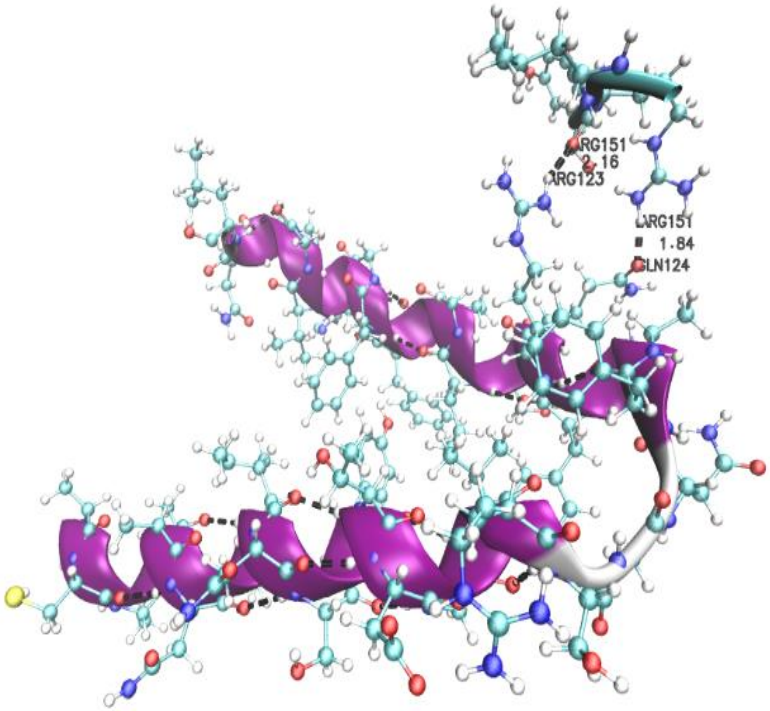
4.2.1 First model

(AG-SLC-WS-262 1)



AG-SLC-5 Fig. 1 Location of the 5th coil in the whole light chain (Model 1)

First, hydrogen bonds formed are displayed as black dot lines. If we look to the 5th coil **(AG-SLC-Fig 1)** we can see that the Arg#123 and Gln#124 forms a hydrogen 2.16 Å bond with Arg#151 from the 6th coil. The 6th and the 5th coil are parallel so they are close in the space, which is why the Arg#151 can interact with the Arg#123 and Gln#124. **(AG-SLC-Fig 2)**



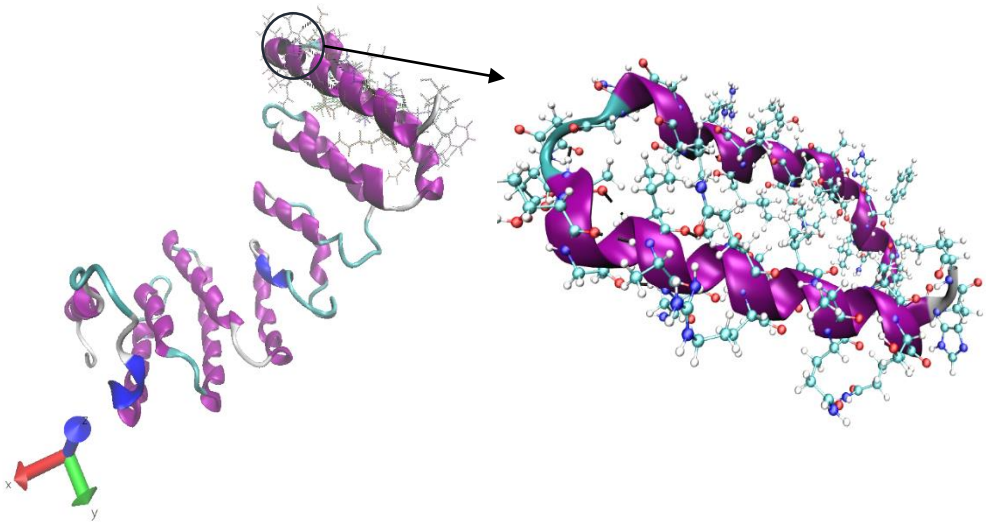
AG-SLC-5 Fig. 2 Inter-helixes hydrogen bonds between 5th and 6th coils.

We can also observe different hydrogen bonds among the α-helix of both sides of the coil. These hydrogen bonds help to stabilize the helix and consequently they maintain the coil stability as well. It is worth noting that most of the hydrogen bonds are made between amino acid *i* and *i+4*, confirming that we have α-helix and no 3_{10} or other different helixes. The main motif is an α-helix/turn/α- helix, the two helixes are on the same plane that may imply a supersecondary structure as it is discussed in the section *"3.5 Supersecondary structure"*. There are 21 hydrogen bonds with an average length of 2.04 Å. Even though there is more information that can be extracted from this models (e.g. torsion angles), this average length value will give a comparative approach to the contraction degree of the helixes.

I-TASSER does not give any other parameter for interatomic forces like steric hindrance between residues, van der Waals or London forces, with the exception of the relative "exposure" value to the solvent of each residues, which is not a proper interatomic force but can be considered as a property that reflects the previous named.

Even though, helicity, which was calculated by circular dichroism (CD) spectroscopy in further structures, assess that even without knowing the values for the diverse interatomic forces, different peptides will show different contraction degree of the helixes due to this forces.

If we look at the 9th coil **(AG-SLC-9 Fig 1)** there are hydrogen bonds in the turn between Gly#239-Ala#242 and Asn#240-Asp#241. There are a large number of hydrogen bonds within the helixes as well**.** Unlike the 5th coil of this model, 9th coil exhibit a roughly acute angle between helixes. This could lead to a different supersecondary structure as it is discussed in section *"3.5 Supersecondary structure".* This structure has 27 hydrogen bonds in it with an average length of 1.93 Å.



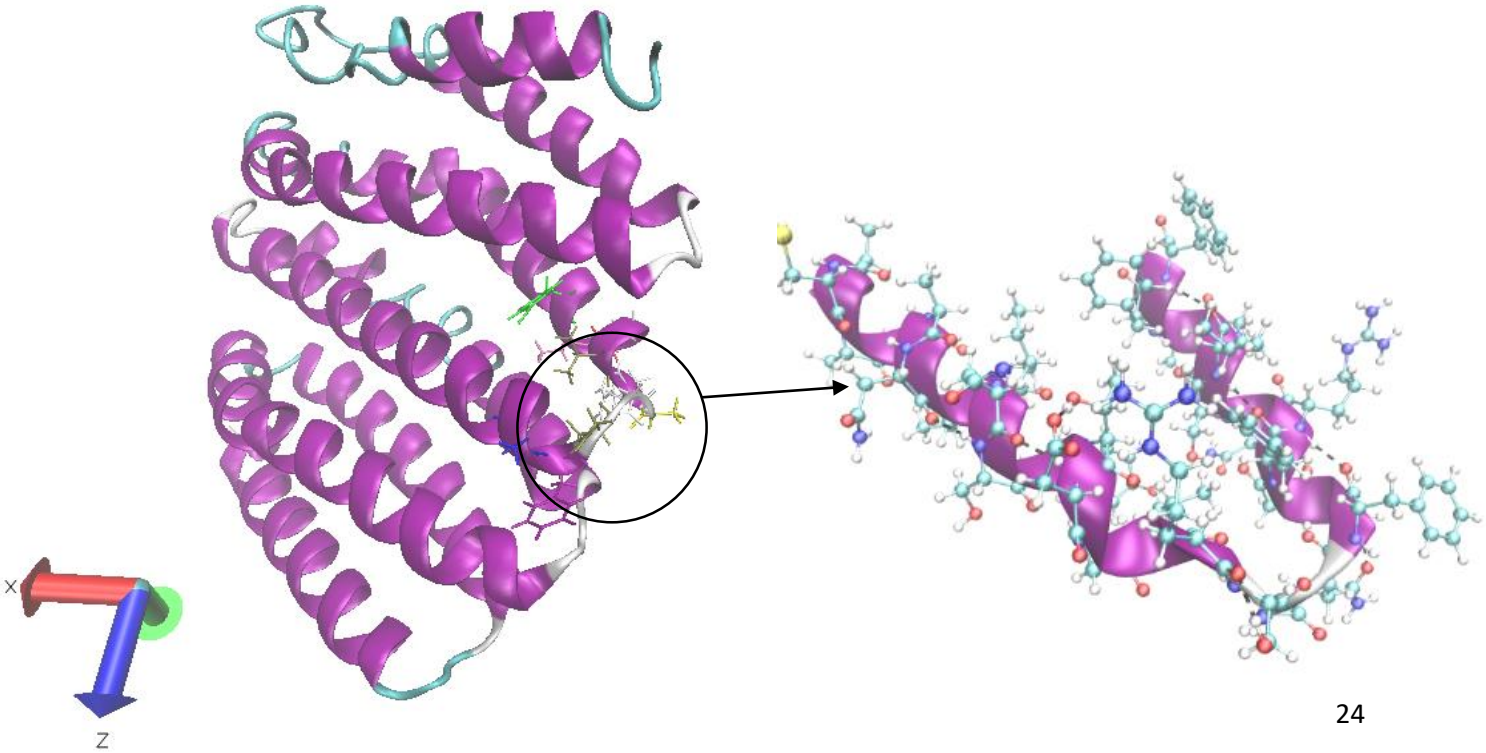
AG-SLC-9 Fig 1 Location of the 9th coil in the whole light chain structure (Model 1)

It is worth saying that the first model has the best description of the structure. Either way, all the other models had been taken into consideration to get probable different structures.

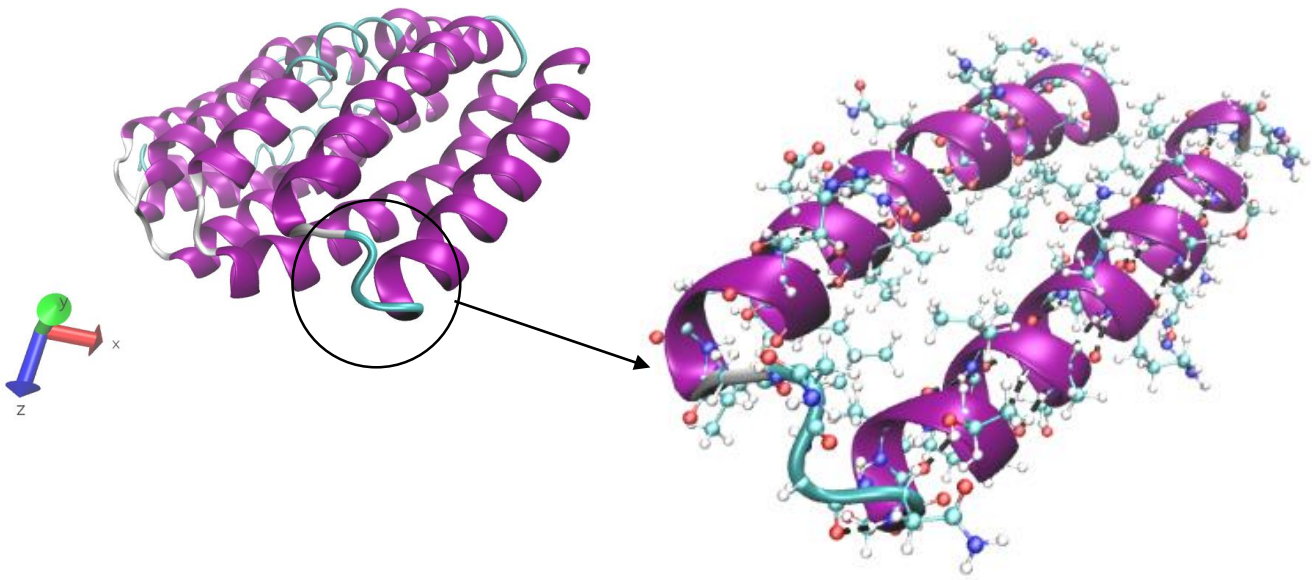
4.2.2 Second Model

(AG-SLC-WS-262 2)

In this model 5th coil **(AG-SLC-5 Fig3)** have no interactions between its amino acids and the amino acids of a near above or below structures. The main motif is still conserved with the two helixes on the same plane. There are 32 hydrogen bonds within the helixes with an average length of 1.95 Å.



AG-SLC-5 Fig 3 Location of the 5th coil in the whole chain structure (Model 2)



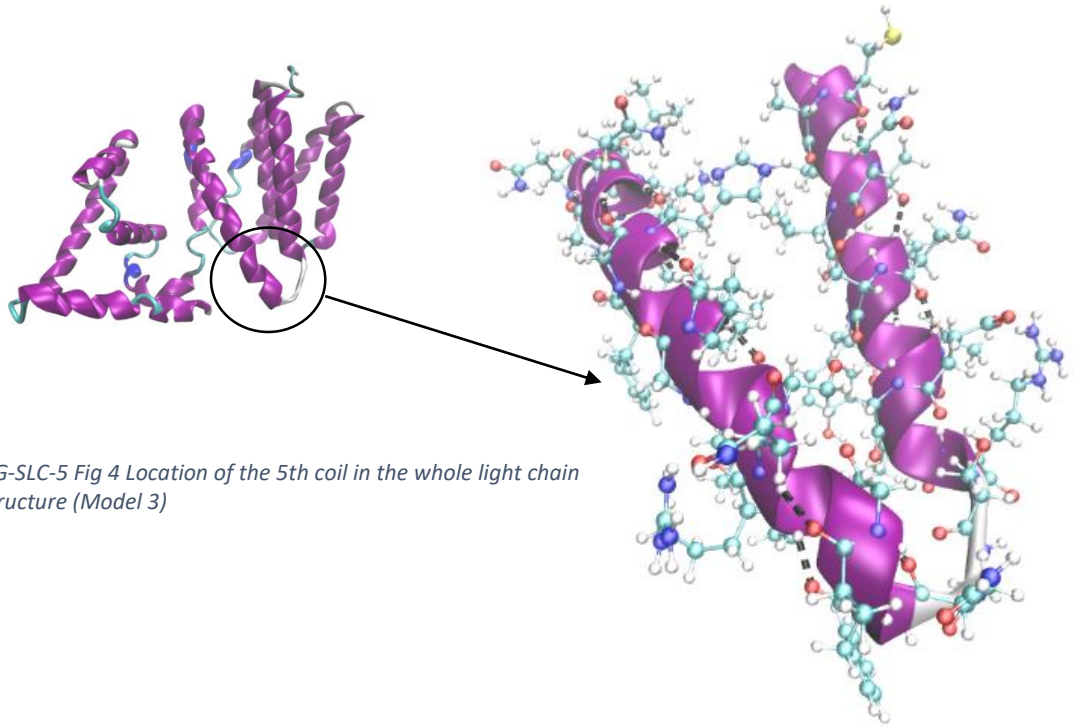
AG-SLC-9 Fig 2 Location of the 9th coil in the whole light chain structure (Model 2)

The same situation has been founded in the 9th coil. There are no interactions in the coil's amino acid **(AG-SLC-9 Fig 2).** In this case the angle between the helixes is more similar to the 5th coil than in the first model's case. There are 27 with an average length of 2.00 Å. Visually, the second model looks like a compressed version of the first one but the general disposition of the helixes is preserved.

4.2.3 Third model

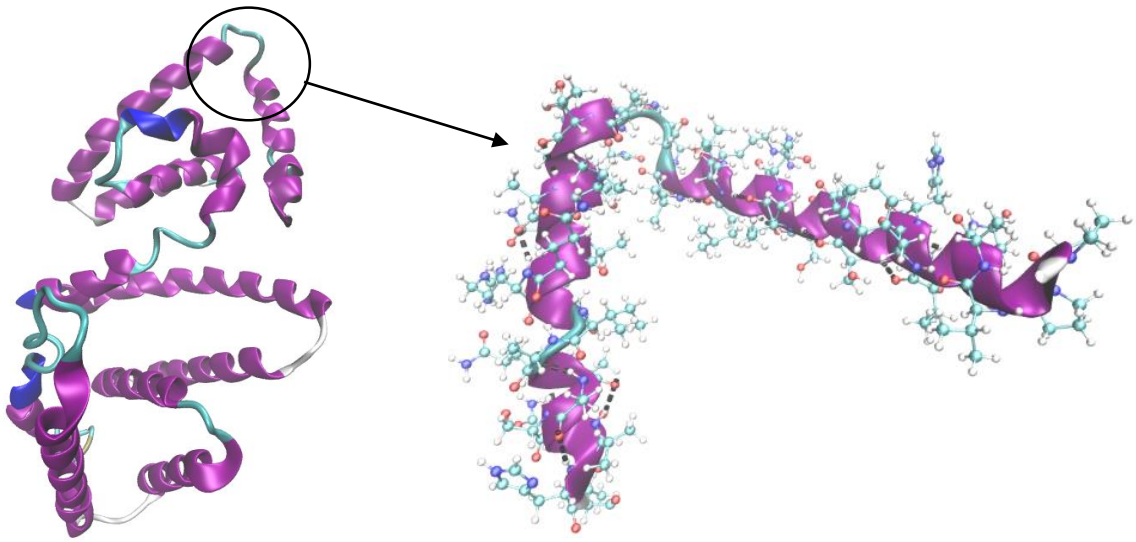
(AG-SLC-WS-262 3)

No hydrogen bonds are present in the turn structure **(AG-SLC-5 Fig 4). T**here are 22 hydrogen bonds in the helixes with an average length of 1.96 Å. Furthermore, the general structure of parallel helixes is barely conserved. As it is said before, the C-score of this model is pretty low (-4.80) so it should not be considered as a possible structure.



AG-SLC-5 Fig 4 Location of the 5th coil in the whole light chain structure (Model 3)

No hydrogen bonds are present in the turn structure **(AG-SLC-9 Fig 3).** There are 14 hydrogen bonds with an average length of 1.99 Å. The degree of compressity of the helix is similar to previous models. Nevertheless, as we can see in **(AG-SLC-9 Fig 3),** the orientation of the helixes in the 9th coil is almost perpendicular for this model. This coul lead to a different supersecondary structure. As it is said before this will be discussed in the section "*3.5 Supersecondary structure".*

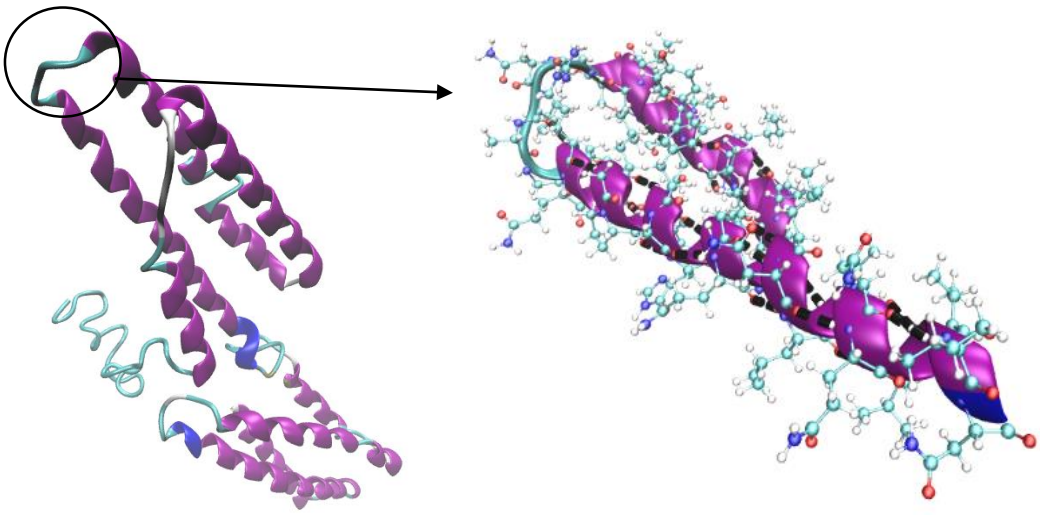


AG-SLC-9 Fig 3 Location of the 9th coil in the whole light chain structure (Model 3)

4.2.4 Fourth model

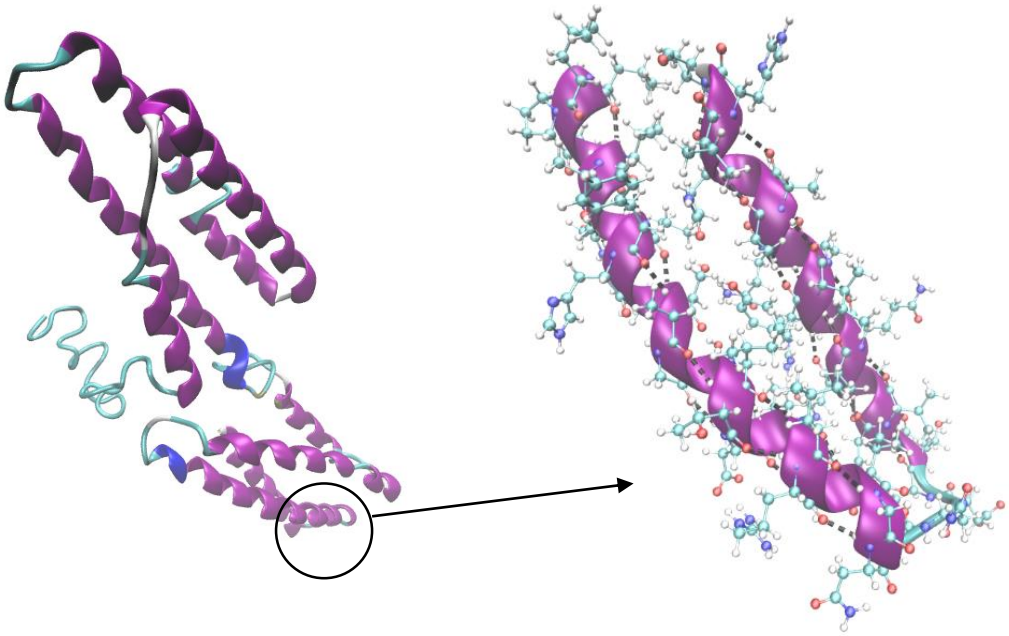
(AG-SLC-WS-262 4)

In this model if we look to the 5th coil we can see that there is a hydrogen bond between the Phe#119 and Arg#123. There are no more hydrogen bonds in the coil **(AG-SLC-5 Fig 5).** And if we look to the 9th coil, there is no hydrogen bonds in the coil by itself. **(AG-SLC-9 Fig 4).** In the first case there are 23 hydrogen bonds with a 2.01 Å average length, in the latter one there are 32 with a 2.02 Å average length.



AG-SLC-5 Fig 5 Location of the 5th coil in the whole light chain structure (Model 4)

In both cases there is a lack of planarity between the two helixes. Nevertheless, it has to be taken into account that the C-score for this model was 5, so it is not the best repressentation of the peptide folding.

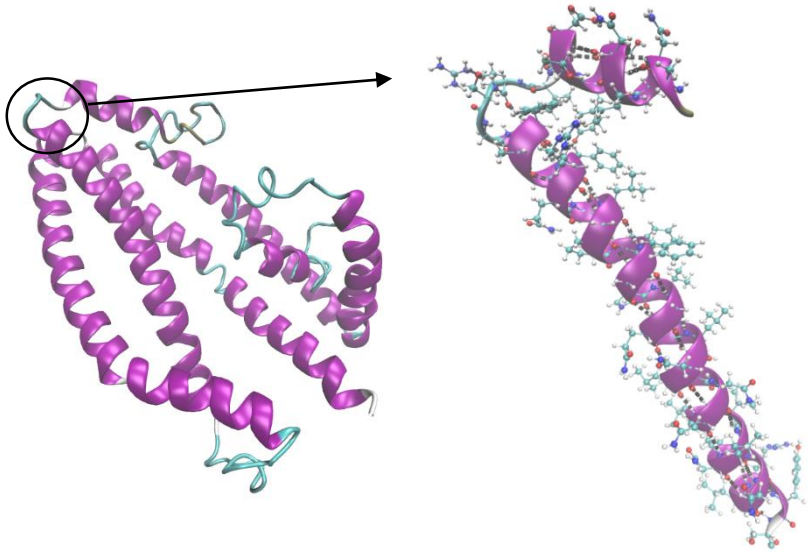


AG-SLC-9 Fig 4 Location of the 9th coil in the whole light chain structure (Model 4)

4.2.5 Fifth model

(AG-SLC-WS-262 5)

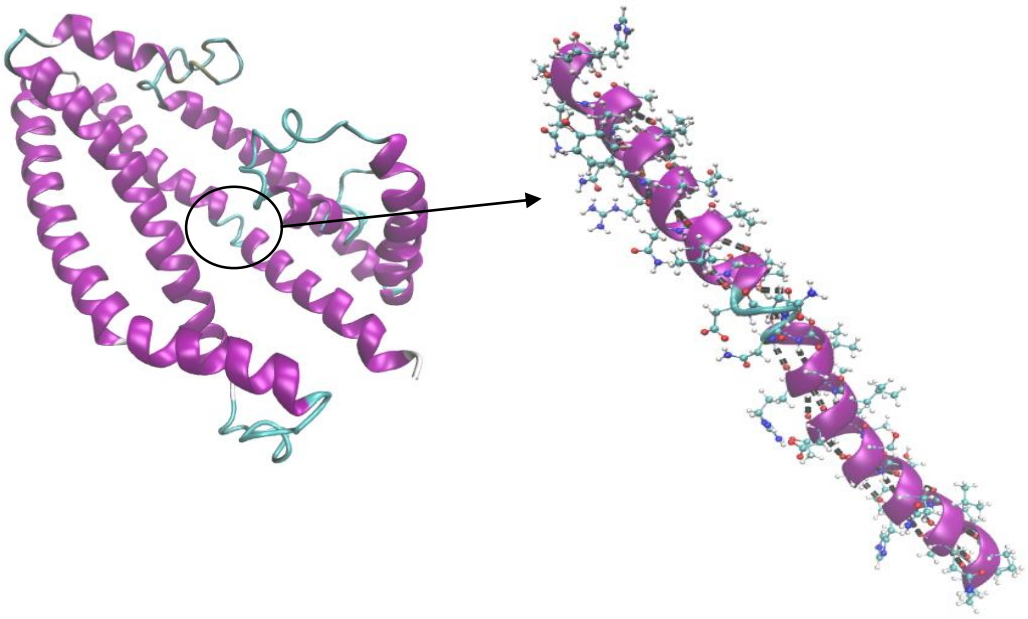
In the last model if we look to the 5th (**AG-SLC-5 Fig 6)** coil we have no hydrogen bonds between coil's amino acids again and there is an acute angle between helixes. There are 35 hydrogen bonds within the helixes with an average length of 2.00 Å.



AG-SLC-5 Fig 6 Location of the 5th coil in the whole light chain structure (Model 5)

The torsion between helixes could be related to the short size of one of them because of the proximity of a random coil. If that helix was completely straight the 5th coil might have parallel helixes in the same plane. But again, like the third and the fourth model, the C-score is so low (-5), so this model shouldn't be taken as good as the first two.

Moreover, if we look to the 9th coil (**AG-SLC-9 Fig 5)** we can see a few hydrogen bonds with the coil's amino acids. First we have and inside coil hydrogen bond: Gly#237 with Gly#239 and another one between Gly#239 and Ala#242. There are 27 hydrogen bonds within the helixes with an average length of 2.01 Å.



AG-SLC-9 Fig 5 Location of the 9th coil in the whole light chain structure (Model 5)

4.3 Charge factor

If we look carefully at each residue that form these coils we can see that we have a certain amount of charge in each of them, this may affect at its stability because the coils are more exposed to the solvent, so charged amino acids provide stability to these structures.

HP stands for hydrophilic residue, PC for positive charged, NC for negative charged, PU polar uncharged

5th coil: (#114) VRSGNFAG (#121)

Val-	-Arg-	-Ser-	-Gly-	-Asn-	-Phe-	-Ala-	-Gly
HP	PC	PU	PU	PU	HP	HP	PU

9th coil: (#236) VGNGNDAT (#243)

Val-	-Gly-	-Asn-	-Gly-	-Asn-	-Asp-	-Ala-	-Thr-
HP	PU	PU	PU	PU	NC	HP	PU

As we can see the 5th coil has a slight positive charge on it (Arg) and the 9th coil has a negative one (Asp). We cannot obtain any conclusion from this because there is just one from eight amino acids in both cases, so it will be a slight charge on them, but I do not reckon that this is the main factor that stabilize the coil.

4.4 Preference factor

Another factor that could induce the formation of the turns is the amino acids preference for certain secondary structures. The amino acids preferences has been studied for years[53] and are summarised in *Fig.15*

Amino acid	Preference α-helix	β-strand	Reverse turn
Glu	**1.59**	0.52	1.01
Ala	**1.41**	0.72	0.82
Leu	**1.34**	1.22	0.57
Met	**1.30**	1.14	0.52
Gln	**1.27**	0.98	0.84
Lys	**1.23**	0.69	1.07
Arg	**1.21**	0.84	0.90
His	**1.05**	0.80	0.81
Val	0.90	**1.87**	0.41
Ile	1.09	**1.67**	0.47
Tyr	0.74	**1.45**	0.76
Cys	0.66	**1.40**	0.54
Trp	1.02	**1.35**	0.65
Phe	1.16	**1.33**	0.59
Thr	0.76	**1.17**	0.90

Amino acid	Preference α-helix	β-strand	Reverse turn
Gly	0.43	0.58	**1.77**
Asn	0.76	0.48	**1.34**
Pro	0.34	0.31	**1.32**
Ser	0.57	0.96	**1.22**
Asp	0.99	0.39	**1.24**

Figure 15 This table shows the main secondary structure preference for each amino acid.

If we look at all the amino acids that can form the coils in the different models we can notice that we have few amino acids that are prone to make a turn, 4/8 in the 5th coil case and 5/8 in the 9th coil case.

5th coil: (#114) VRSGNFAG (#121)

```
Val-  -Arg-  -Ser-  -Gly-  -Asn-  -Phe-  -Ala-  -Gly
             1.22   1.77   1.34                  1.77
```

9th coil: (#236) VGNGNDAT (#243)

```
Val-  -Gly-  -Asn-  -Gly-  -Asn-  -Asp-  -Ala-  -Thr-
      1.77   1.34   1.77   1.34   1.24
```

It is worth it to say that Gly are more prone to make turns because it is so flexible, so it can easily make the sharp turns and bends needed in a β- turn. Proline is special as well because it is so rigid: we could say that it is pre-bend for the β-turn.

Aspartic acid, asparagine and serine have in common that they have short side chains that can form hydrogen bonds with the own backbone. These hydrogen bonds compensate the energy loss caused by bending the chain into a β-turn.

The preference of each amino acids is a factor to take into account whenever the folding of a peptide is discussed, as it is shown most the amino acids that form the coil have a high preference to form it rather than other structures.

4.5 Superstructure

The main motif that have shown the sequence that contains the 5th coil it is a super-secondary structure called αα hairpin or helix hairpin (*Fig.17)*[54]. This refers to two almost antiparallel and coplanar α-helical segments connected by a loop.

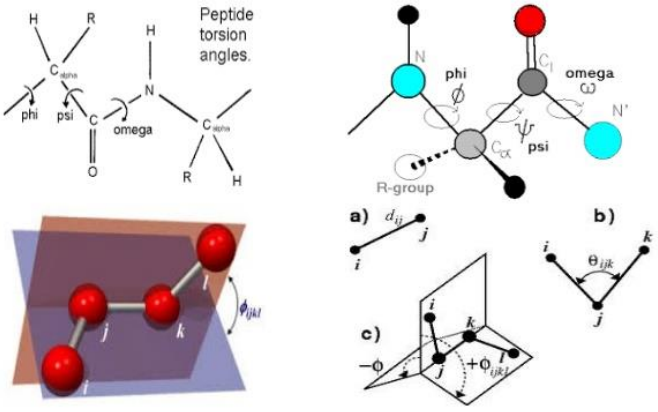


Figure 16 Scheme of the angles in a peptide bond

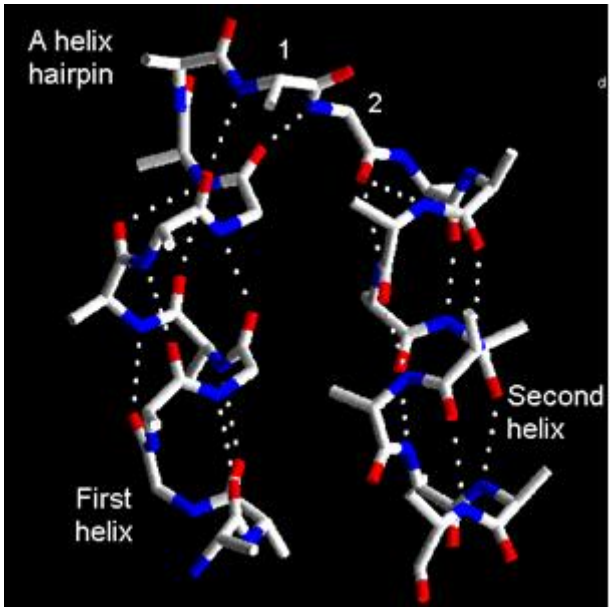


Figure 17 Helix hairpin supersecondary structure

Clearly, the longer the length of the loop the greater the number of possible conformations. However, for short connections there are limited number of conformations and for the shortest loops of two or three residues, there is only one allowed conformation.

Antiparallel and coplanar α-helixes will interact generally by hydrophobic interactions between side chains. Therefore, hydrophobic amino acids have to be appropriately positioned in the amino acid sequence, one per turn of each helix approximately, to generate hydrophobic interactions at the space between helixes.

The C-terminal residue of the first helix must have a short side chain to avoid steric hindrance and is observed commonly to be glycine. The first residue of the second helix frequently has a small polar side chain, such as Ser, which can form hydrogens bonds with the free NH groups at the amino-terminal end of the second helix. The central residue of the αα-corner is almost always hydrophobic as it's buried and interacts with other non-polar side chains buried where the ends of the two helixes contact to each other. The shortest α-helical connections involve two residues which are oriented roughly perpendicular to the axes of the helices. The first of these two residues adopts phi and psi angles in the bridging or alpha-helical regions of the Ramachandran plot. The second residue is always glycine and is in a region of the Ramachandran plot with positive phi which should not be easy-available to other amino acids due to the steric hindrance of their lateral chains. *Tables 3 & 4* show this features:

Table 2 Dihedrals angles from 5th coil in model 2

SECOND. STR.	PHI	PSI	OMEGA	AMINO ACID
HELIX	-53.90	13.54	156.25	ARG115
COIL	-149.39	-67.37	167.42	SER116
COIL	45.96	-134.07	170.66	GLY117
COIL	-132.40	126.25	-176.76	ASN118
COIL	-57.58	-19.90	170.33	PHE119
HELIX	-68.23	-42.03	170.03	ALA120

Table3 Dihedral angles from 9th coil in model 1

SECOND. STR.	PHI	PSI	OMEGA	AMINO ACID
COIL	-91.25	-25.07	164.37	GLY237
HELIX	-70.34	-34.33	165.35	ASN238
HELIX	40.37	33.5	-164.02	GLY239
HELIX	-56.22	-31.6	167.49	ASN240
HELIX	-135.65	25.20	-164.12	ASP241
HELIX	-84.50	46.97	174.28	ALA242
HELIX	-68.22	-63.52	174.48	THR243
COIL	-63.35	-34.72	-175.25	GLY244

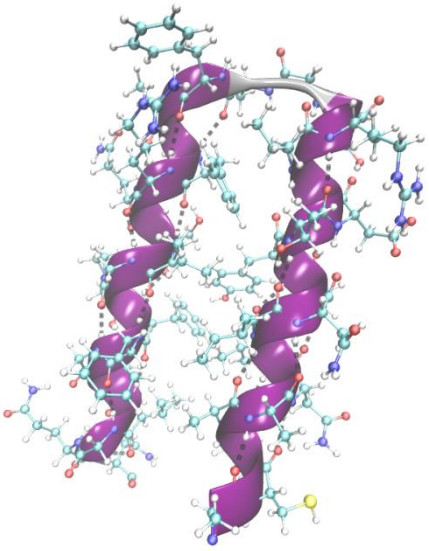
4.6 5th coil 37 amino acid long

A 37 amino acids fragment has been sent. It goes from Ala#100 to Asn#136. This sequence contains the amino acids that form the α-hairpin, two helixes with the 5th coil connecting them. If the same structure is obtained, it means that there is no need of inter-chain hydrogen bonds to stabilize this coil. If not, another sequence including the helix that has formed a hydrogen bond with the first one will be send.

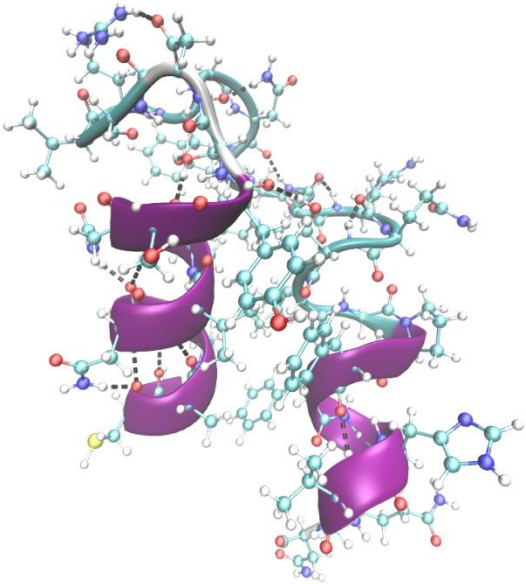
100 105 110 115 120 125 130 135

Submitted: **ACAAANVINSYTDGVRSGNFAGFRQSLGPFFGHVGQN**

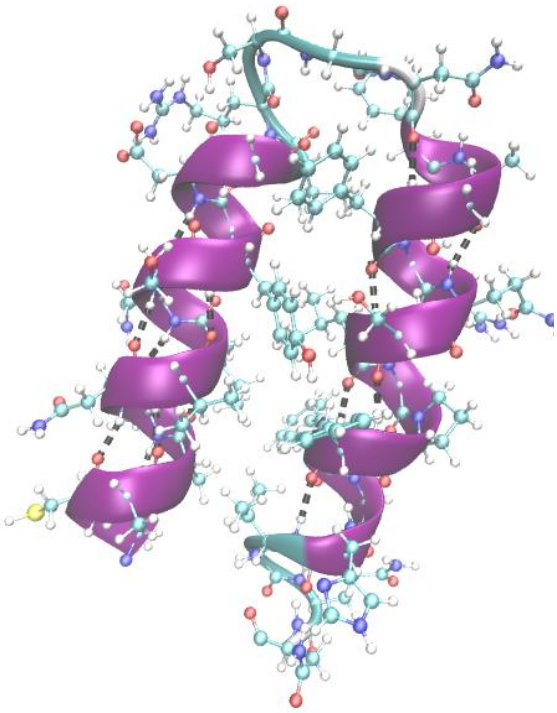
Obtained: **ACAAANVINSYTDGVRSGNFAGFRQSLGPFFGHVGQN**



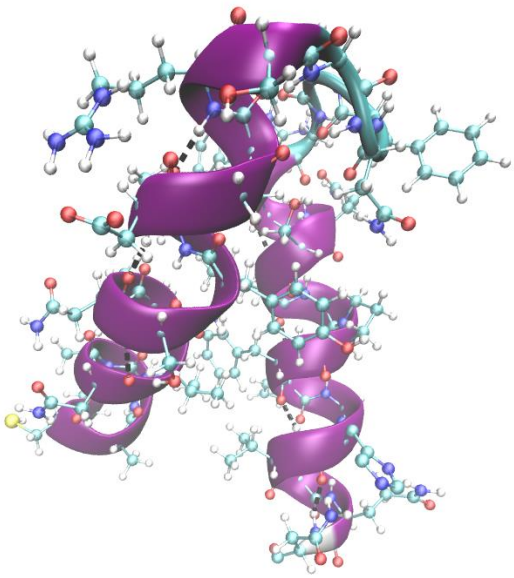
AG-SLC-5-37S M 1 Cs = -1.38



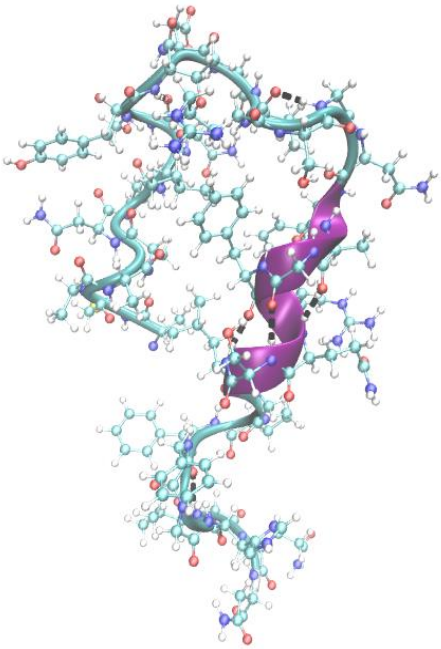
AG-SLC-5-37S M 2 Cs = -3.39



AG-SLC-5-37S M 3 Cs = -1.54



AG-SLC-5-37S M 4 Cs = -3.04



AG-SLC-5-37S M 5 Cs = -4.75

Arg#123 use to form an inter-chain hydrogen bond when it was in the whole light chain. Now instead, this extract from that sequence forms a hydrogen bond with the Phe#119 and Gly#127. There is a stable α-hairpin motif with a 2 amino acid turn (Gly#117 and Asn#118) in the case of 1st model; 5 amino acid turn (Val#114 to Asn#118) in the case of 3rd model; and 6 amino acid turn (Ser#116 to Gly#121) in the 4th model. It could be thought by looking at the 2nd and the 5th models of the sequence that other imaginable structures. However, the models that have better C-score, nearer to 2, are the ones that show this motif.

4.7 5th coil combination fragments

Two more calculations were sent due to the presence of this stable structure. The first one was made with the same sequence (1sthelix-turn-2ndhelix) plus a fragment that contains the same turn and the second helix.

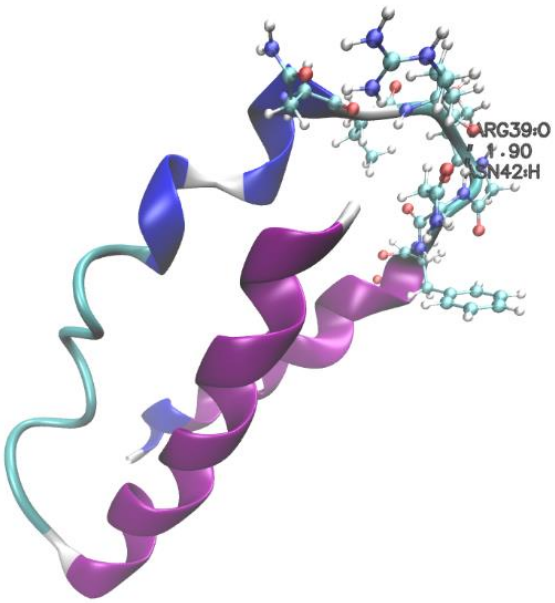
Submitted:

ACAAANVINSYTDGVRSGNFAGFRQSLGPFFGHVGQNVRSGNFAGFRQSLGPFFGHVGQN

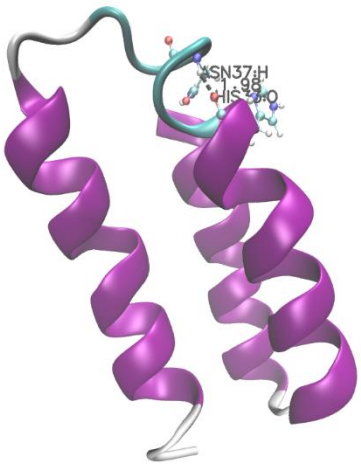
Obtained:

ACAAANVINSYTDGVRSGNFAGFRQSLGPFFGHVGQNVRSGNFAGFRQSLGPFFGHVGQN

Two models are plotted because the further 3 had a C-score = -5. Hence, they are supposed to be worse.

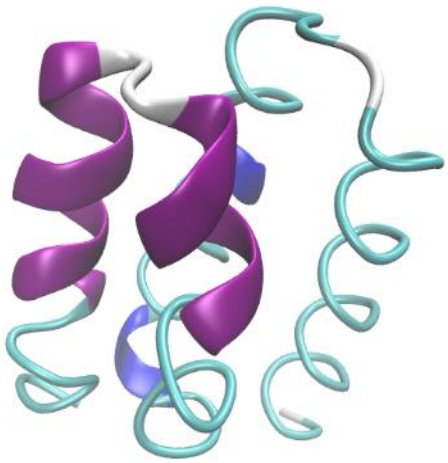


AG-SLC-5 Comb. Fig 1 There is only one hydrogen bond present in the turn. Arg#39-Asn#42. Cs = -0.91



AG-SLC-5 Comb. Fig 2 There is only one hydrogen bond present in the turn. His#33-Asn#37. Cs = -3.22

The second calculation tried was a three repetitive 5th coil fragment. This was runned just to test if the conection between two helixes (αhairpin— αhairpin) will be stable enough. Results were not as good as before, only the 2nd model was slightly similar as what was expected.



AG-SLC-5 Comb. Fig 3 No hydrogen bonds are present in the turn. Cs = -2.30

Although these chimeric structures were not part of the original sequence, they cannot be obtained from the natural SF, they were calculated anyway. Allowing to confirm that, for the specific case of the SF-LC sequence, using structures that appears to be stable repeatedly does not imply that a stable structure will be obtained.

4.8 9th coil containing fragments

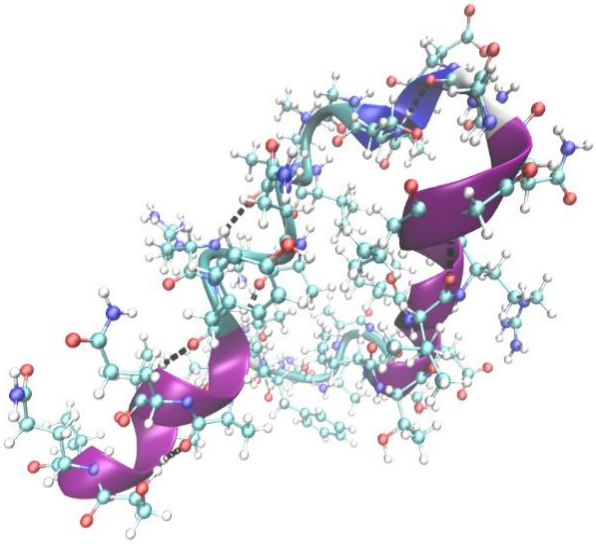
9th coil sequence has also been send to check if the unchanged structure could be obtained whithout the hydrogen bond between the helixes that was before (between Asp#231 and Lys#193 from the underneath helix). A fragment from Val#221 to Val#260 has been sent, it was 40 amino acid long.

225 230 235 240 245 250 255 260

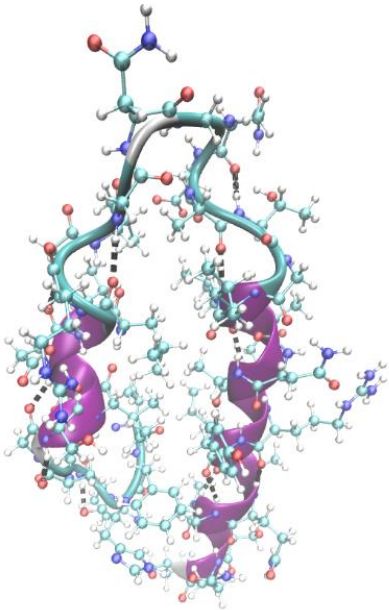
Submitted: **VFHQSAGSITDLLRGVGNGNDATGLVANAQRYIAQAASQV**

Obtained: **VFHQSAGSITDLLRGVGNGNDATGLVANAQRYIAQAASQV**

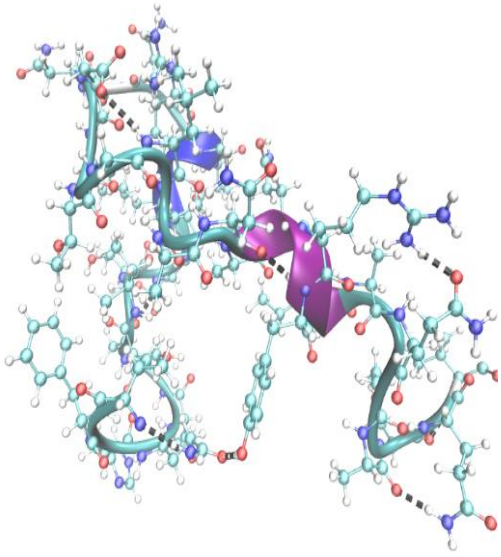
In this case the αhairpin motif from the whole chain have not been obtained. This could be due to the lack of the interchayn hydrogen bond. So the first helix (left) cannot be stabilized in any model as properly as before.



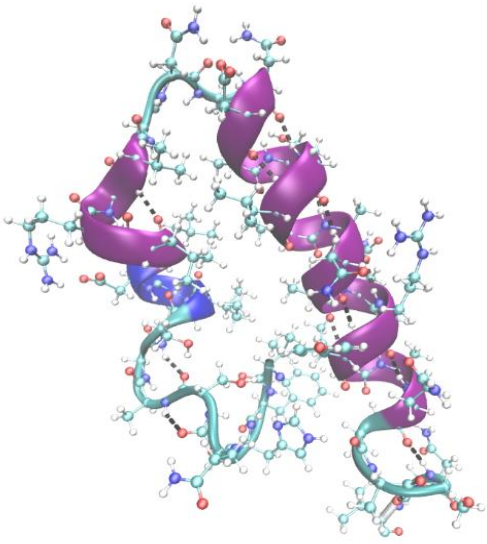
AG-SLC-40A Fig 1 There are two Hbonds in the turn. Ala#249 – Ile#253 and Ala#247 – Arg#251. Cs = -1.57



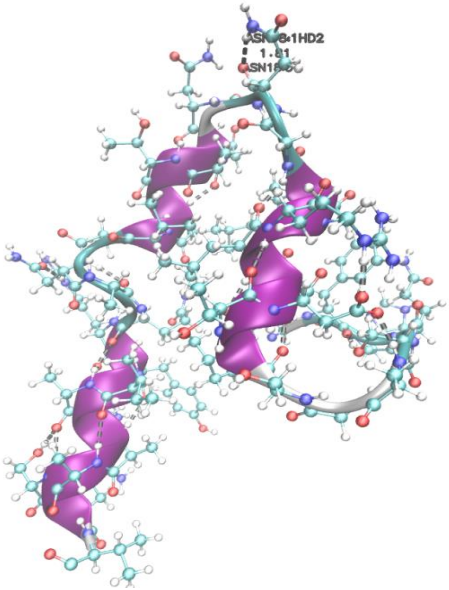
AG-SLC-40A Fig 2 There are three Hbonds in the turn. Leu#233 - Gly#237, Asn#240 - Thr#243 and Ala#242 Val#246. Cs = -2.68



AG-SLC-40A Fig 3 There is one hydrogen bond in the turn. Asp#241 - Leu#245. Cs =-3.92



AG-SLC-40A Fig 4 There is no hydrogen bonds in the turn. Cs = -4.18



AG-SLC-40A Fig 5 There is one hydrogen bond in the turn. Asn#238 - Asn#238. Cs = -5

In the 9th coil case the models are not that similar to the original light chain coils in terms of shape and structure distrifbution. So

We could say that the 5th coil is more stable by itself than the 9th coil. Meaning that it will conserve the same structure that has in the original chain if it is synthesized.

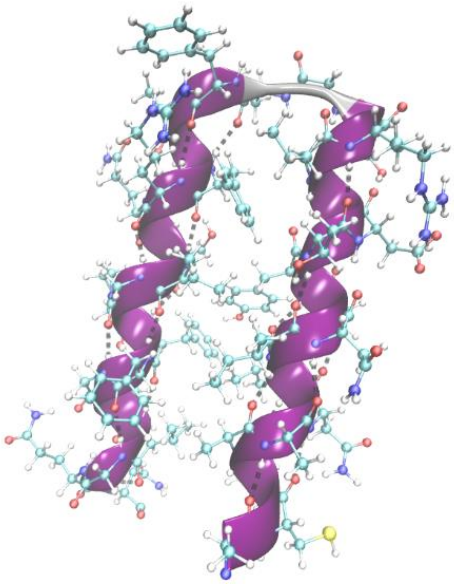
So due to this results ,as well as the dummy calculations for the combinated fragments, it can be concluded that the best sequence to be synthesized has to contain the 5th coil.

4.9 5th coil containing fragments

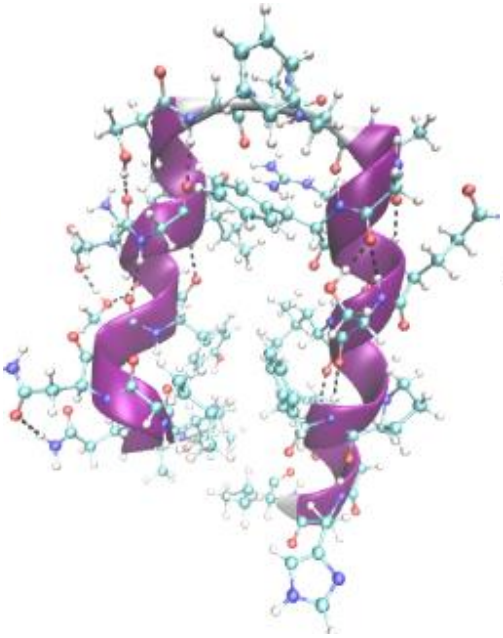
Two more sequence fragments containing the 5th were sent with few amino acids less. The first one with three amino acids less at each side of the sequence and the latter one with four less. Only the models with higher C-score are taked into account to compare with similar criteria.

Also the 5th coil previous fragment is plotted.

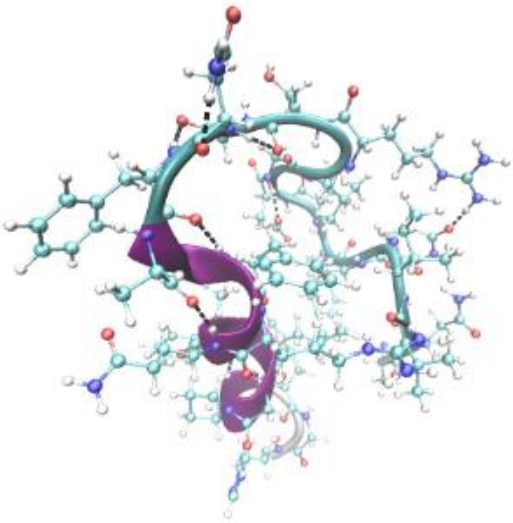
```
   105  110  115  120  125  130  135
ACAAANVINSYTDGVRSGNFAGFRQSLGPFFGHVGQN  5th coil 37aa (A)
---AANVINSYTDGVRSGNFAGFRQSLGPFFGHV---  5th coil 31aa (B)
----ANVINSYTDGVRSGNFAGFRQSLGPFFGH----  5th coil 29aa (C)
```



AG-SLC-5-37S M 6 Cs = -1.38



AG-SLC-5-31S M 1 Cs = -1.29



AG-SLC-5-29S M 1 Cs = -1.57

As we reduce the number of amino acids less hydrogen bonds can be formed. There are no similar hydrogen bonds between this structures. The results obtained also shown that if we remove three amino acids form each side the main structure is maintained but if we remove we loose one of the α-helixes.

This may due because of the interaction between Val#106 and Arg#123 that forces the fragment that is supposed to be the first loop to twist. Hence, the helix doesn't start and at least three loops are needed to form a stable helix, so we don't have enough amino acids to form the helix before the turn. This is the reason why we obtain a random distribution of the amino acids before the coil instead of having a helix.

4.10 Symmetric 27

(AG-SLC-5-27S 1M)

Another point to take into account is that,because of being an α-helix, each loop has 3.6 amino acids so a 5 loop helix, like the one that we have, is 18 amino acids long. The smallest helix thaht can be formed need 3 loops, so to make it feasable we would need 11 amino acids at least.

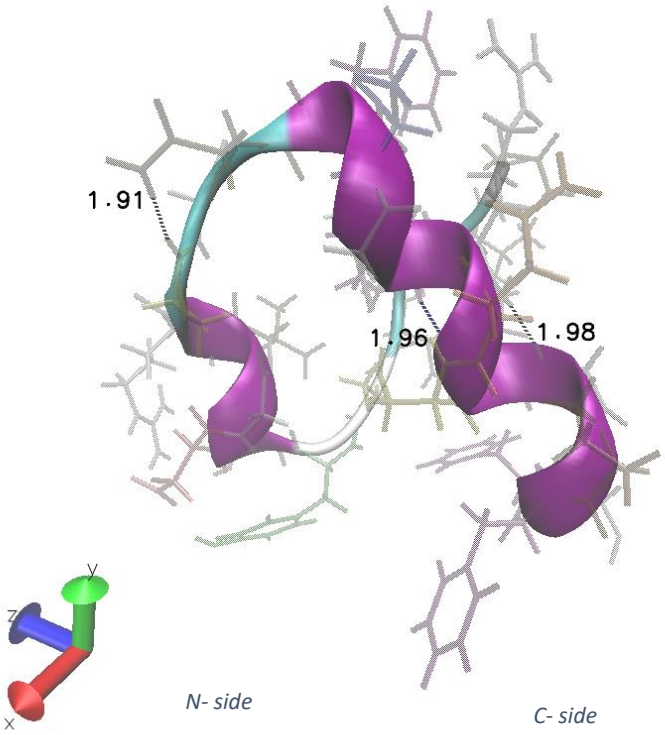
If we try to cut one more amino acid at each side, so we have 27 in total, the model obtained that has a structure like the previous is this one:

NVINSYTDGVRSGNFAGFRQSLGPFFG

As it can be observed, the fragment before the turn is formed by eleven amino acids but we cannot see a properly helix, we can appreciate just a five amino acid loop, the rest is a random coil.

With our cutoff parameters only three hydrogen bonds are formed. Right helix stabilized with two hydrogen bonds within.

Only first model is plotted because the other 4 have a C-score below -3.



AG-SLC-5-27S 1M Cs = -1.49

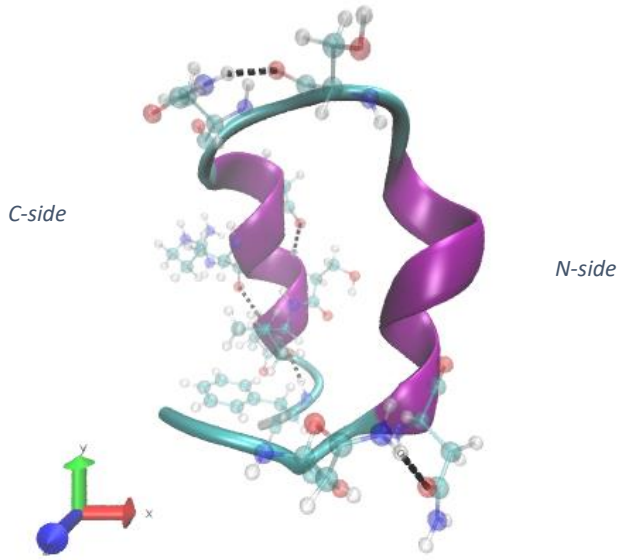
4.11 Asymmetric 29aa

(AG-SLC-5-29A)

The fact that C-terminus helix is the stable one has been assessed. Therefore, trying to cleave amino acids from that side instead of doing at each sides of the peptyde at the same time, is the next step. A 29 amino acid fragment is obtained, it seems to retain the same structure:

AANVINSYTDGVRSGNFAGFRQSLGPFFG

As it was expected the general structure is more analogous to the 37aa model than in the 29aa symmetric case. It has practically the same number of hydrogen atoms. As the number of amino acids is reduced the less parallel are the helixes. Hence, we lose the α-hairpin supersecondary structure. As we can see in *Table 4*the second amino acid is not a glycine but a serine. Nevertheless, the second amino acid still has a positive phi angle so the motif is similar.



AG-SLC-5-29A 1M Cs = -1.72

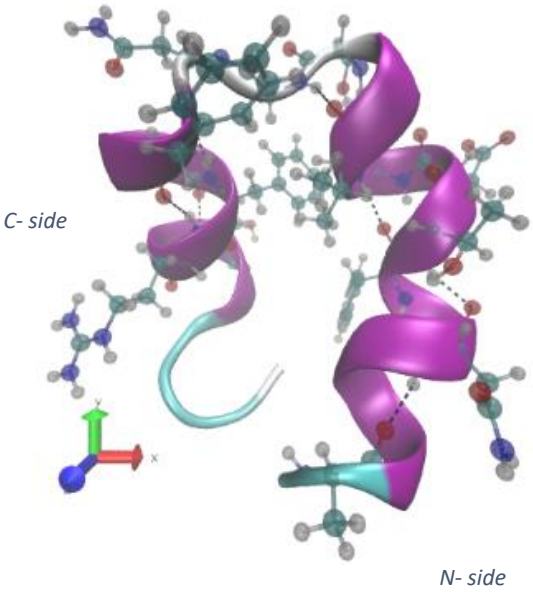
Table 4 Peptide´s bond angles

SECOND. STR.	PHI	PSI	OMEGA	AMINO ACID
COIL	5031	47.89	-173.19	ARG115
COIL	56.44	15.30	174.97	SER116
COIL	-57.30	36.30	-169.23	GLY117
COIL	-47.25	112.69	174.19	ASN118
COIL	39.00	3.19	177.27	PHE119

4.12 Asymmetric 28aa & 27aa

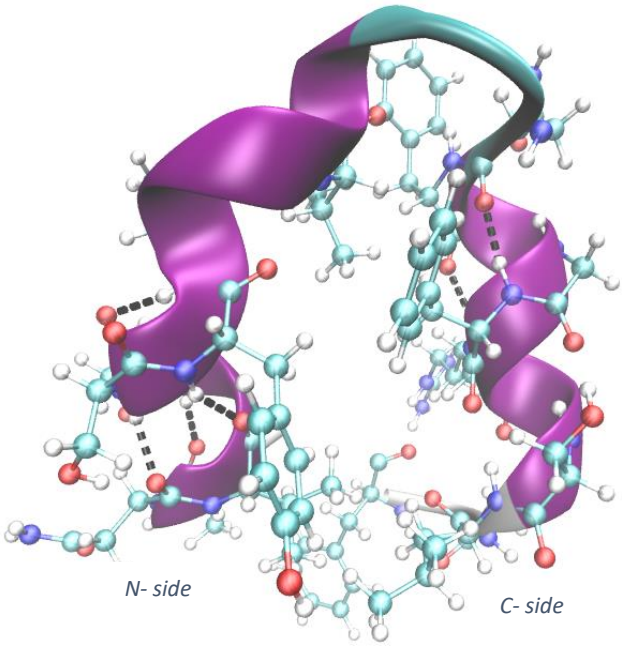
Cleaving another pair of amino acids does not make any change in the folding pattern of the peptide. It just increases the angle between the two helixes. Surprisingly, if we chop another amino acid leading to a 27aa segment, the helixes seems to be coplanar again coplanar again. What really happens is that the chains are so short that they share a plane, even though they are pointing different directions in space.

AANVINSYTDGVRSGNFAGFRQSLGPFF



AG-SLC-5-28A 1M Cs = -1.69

AANVINSYTDGVRSGNFAGFRQSLGPF



AG-SLC-5-27A 1M Cs = -1.8

After all the calculations the only structure that seems to keep the two helixes antiparallel and coplanar was the 37 amino acid segment. Therefore, it was decided to synthesize this peptide.

5 RESULTS AND DISCUSSION – LABORATORY PART

5.1 Peptide synthesis

The peptide synthesis was carried out using SPPS technique (see *section 1.5*). Liberty Blue™ was the chosen device for this purpose. Normally the sequences that are synthesized using this device are shorter (10-15 amino acids), for that reason it was resorted to a technician expert in the field. Although the challenge of the length, there were other problems such as: the presence of glutamine between bulky amino acids, the DGV fragment that was prone to aspartimide formation, more than 3 hydrophobic amino acids in a row, among others... All of these problems were sorted out by changing certain parameters of the device such as the time of the reactions or the temperature (check section *1.5.4* for more side reaction problems). Although all of that we managed to synthesize the sequence. After the cleavage and freeze drying of the peptide and before any further purification a yield calculation was made to check if the reaction has gone well

The peptide synthesizer was configured to prepare 1 mmol of peptide, which according to the monoisotopic molecular weight would give 0,3829 g of peptide. The weight of the peptide obtained prior any purification was 0,1685 g.

$$\%Yield = \frac{0.1685\ g\ peptide\ (obtained)}{0.3829\ g\ peptide\ (theoretical\ weight)} \cdot 100 = 44{,}01\ \%$$

5.2 Peptide purification

After synthesizing the peptide with the solid phase peptide synthesis technique the sample was brought to the HPLC to check its purity. First, a preparative column was used to look for the presence of other species (see *Fig.18*). Concentration used was 1 mg/mL and mobile phase was 50% A and 50%B. We could see that we have a mix of compounds but either way we also analysed the sample with MALDI-TOF technique to check the molecular mass of the impurities (see *Fig.19*). With this column we could not separate the peptide from the impurities so we brought it to a semi preparative column.

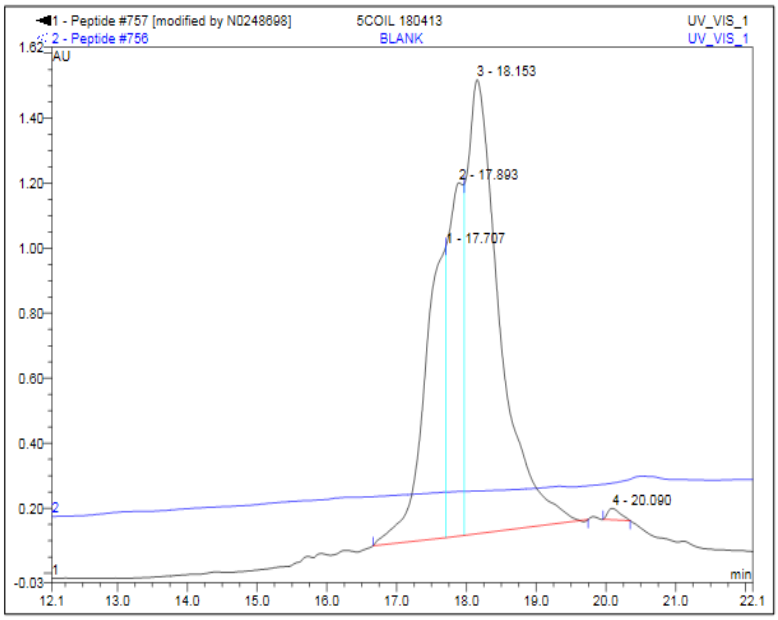


Figure 18 Preparative HPLC. It can be observed that we have a mix of compounds that appear in a range from 17-20 min.

To purify this an HPLC-RP was carried out. The optimal conditions of purification were found after a tedious method-optimization process (see *Table 5)*. One of the main problems during the optimization of the process was that high concentrations of the peptide could not be used. If the concentration of the sample exceeded the 5 mg/mL in double distilled water it started to gelled. It was observed that the gelation process takes place within 4h after making the solution, this gelation might be due to the rearrangement of the helixes of the peptide made, trying to reach a more stable conformation. This phenomenon was not observed before because the sample was freeze dried after the quenching from the ether, so there was no solvent.

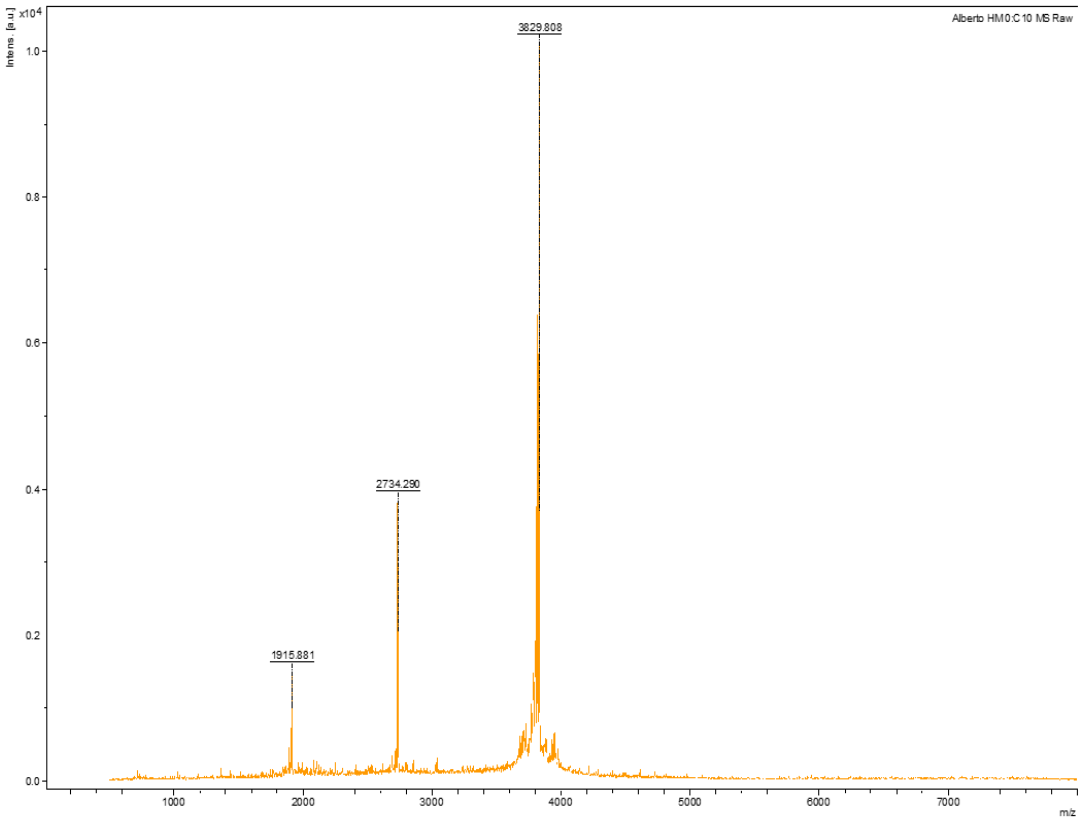


Figure 19 Peptide sample before purification contains two other compounds with lower mass.

The gelation of the sample in the vial was not a problem because the sample could be re-diluted if we reached a concentration of 1 mg/mL. The issue came when, after few more analysis, we realised that the peptide also has gelled inside the needle, injector and few connections prior to the column. We tried to flush away the gel from all the components with just B solution at high pressure. It came off from all of them but a few connections were needed to be replaced afterwards. Using mobile phase B instead of water allowed to reach a ~5mg/mL concentrated solution without gelation.

Table 5 Purification method. Temperature (44°C) and flux (4.0mL/min) remain constant during the whole process.

Time (min)	% Solution A	%Solution B	Step
1-4	60	40	**Column equilibration**
20-23	45	55	**Elution of the different fractions**
25-30	5	95	**Flushing**
30-35	60	40	**Column equilibration**

The different fractions were obtained at diverse times (*see Fig.20)*. Once the regions of time were defined, the automated collector pour every fraction in vials so it could be analysed by MALDI-TOF technique again. According to MALDI-TOF the fraction that contains the peptide was Fraction 6 (*see Fig.21)*

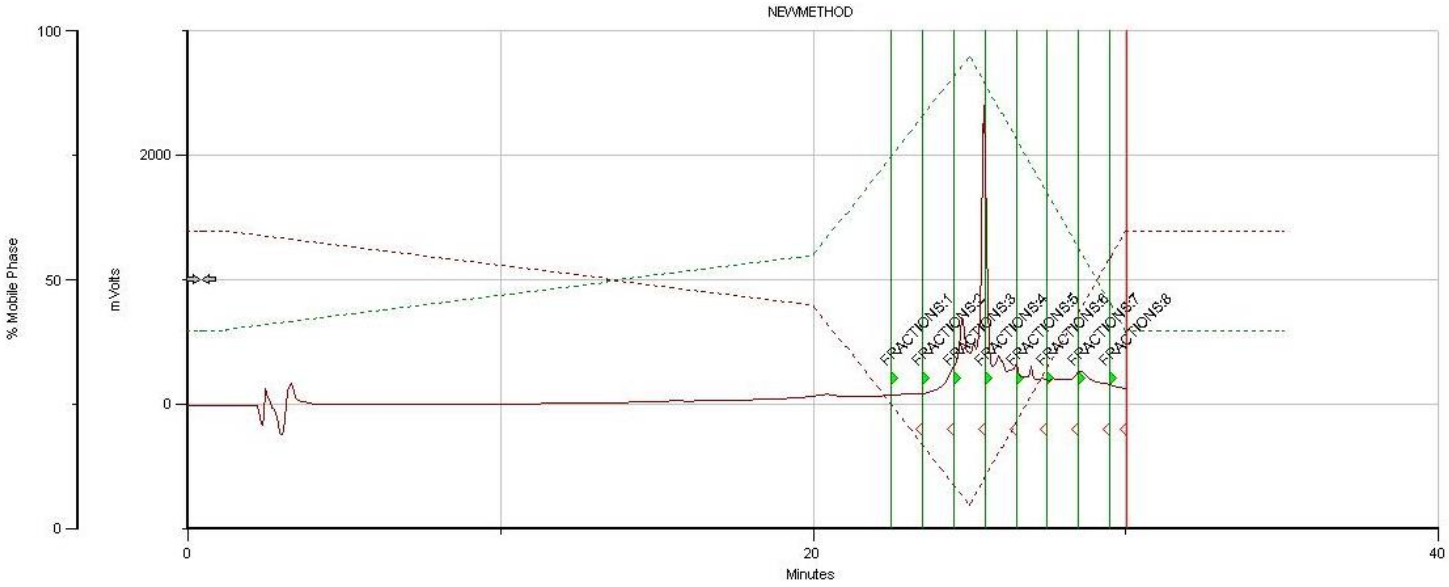


Figure 20 Chromatogram of the optimised peptide separation. Solvent A in red dotted line, Solvent B in green dotted line.

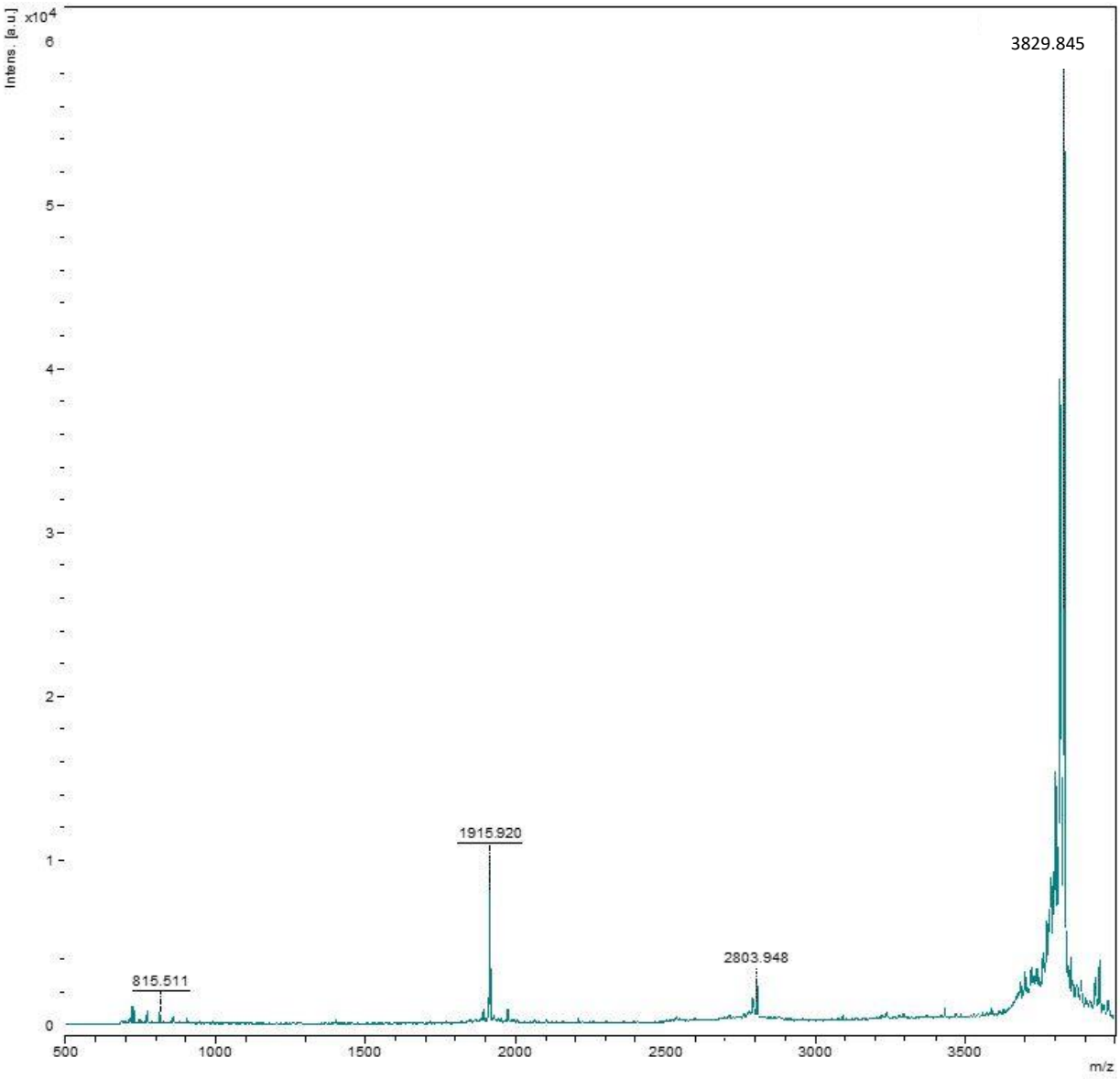


Figure 21 Fraction 6 MALDI-TOF spectra. There are still some other compounds that could be adducts formed during the synthesis. They could not be separated properly with the available tools. The major peak is 3829.845, according to the theoretical molecular mass.

5.3 CD-UV Spectra

This technique is used to get information about secondary structure composition (%helix, sheet, turns, etc.) from the peptide bond region. Absorption in 240nm and below region is due principally to the peptide bond. There are two main transitions: n → π* (weak but broad, centred around 220nm) and π→ π*(more intense, centred around 190nm). Aromatic amino acids side chains may affect to the spectra as well. The different types of regular secondary structure found in proteins provide characteristic CD spectra in the far UV.

Databases of proteins with different types of folding that have been characterised with X Ray crystallography are employed to make algorithms that provide the secondary structure of our sample using its CD spectra. CD spectra in this project have been calculated using *Dichro Calc*[49] with the *ab initio*

parameter set for the backbone chromophores including charge-transfer and aromatic side chain transitions.

One point to be noted is that databases do not include structures of oligopeptides and thus CD spectra of these compounds, except when particular secondary structures are predominant, cannot be analysed reliable at present. In our case the predominance of the α-hairpin motif benefits the CD characterization because it is a well-known structure. If we look back at the C-Score of the models 1 &3 of our peptide, in *section 3.6, were* -1,38 and -1,54. This implies that models 1 & 3 are most likely our synthesized peptide and so do they CD-UV spectra will be (*Fig 22 & 23 respectively).*

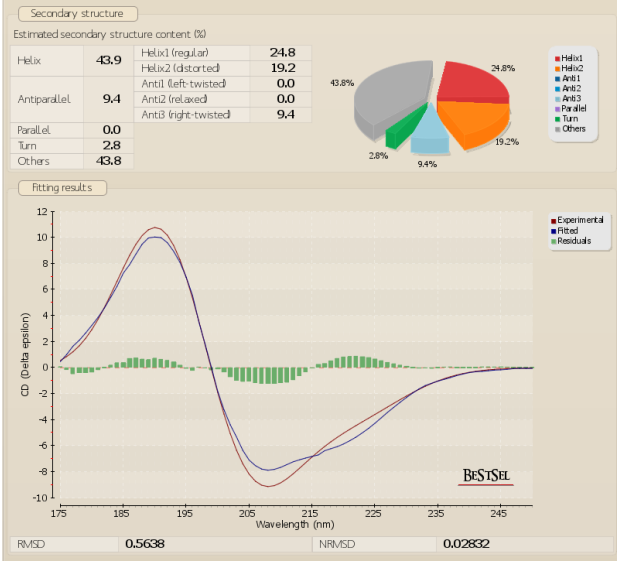


Figure 22 CD-UV Model 1 5th coil

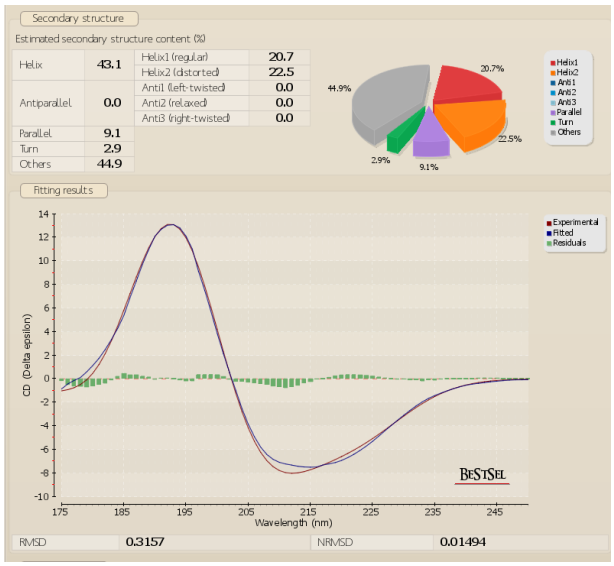


Figure 23 CD-UV Model 3 5th coil

If we look the theoretical spectra that have been obtained, we can see that the percentage of "Others" type of structure is around 45% of the whole peptide. This may be due to the parameters that *Dichro Calc* uses to calculate these CD-UV spectra.

6. CONCLUSIONS

To conclude this project, it can be said that the main structure of the LC-SF has been thoroughly studied and a 37mer sequence of this protein (ACAAANVINSYTDGVRSGNFAGFRQSLGPFFGHVGQN) has been synthesized using SPPS and purified by HPLC. This peptide has shown gelation properties in saturated solutions that may imply the preservation of the expected supersecondary structure, helix hairpin. Nevertheless, this fact could not be confirmed by CD/UV-Vis, only theoretical calculations were done. Not enough information to confirm the structure have been compiled but further investigation will be done in this field. On the other hand, a bibliographical study of silk fibroin functionalization with PEG derivatives has been done. The attempted NMR analysis could not be interpreted properly due to the lack of signal intensity of the peaks and a lack of time. During the recent years there has been a vast research on these silk related biomaterials. However, there is still plenty of possibilities that have not been tried yet. Hopefully this project will continue in the more than capable hands of the Biomolecular Materials Interface Research Group.

7. BIBLIOGRAPHY

(1) Dental Materials. In *The Chemistry of Medical and Dental Materials*; Nicholson, J. W., Connor, J. A., Eds.; The Royal Society of Chemistry, 2002; pp 148–185.

(2) Rockwood, D. N.; Preda, R. C.; Yücel, T.; Wang, X.; Lovett, M. L.; Kaplan, D. L. Materials Fabrication from Bombyx Mori Silk Fibroin. *Nat. Protoc.* **2011**, *6* (10), 1612–1631.

(3) Murphy, A. R.; John, P. S.; Kaplan, D. L. NIH Public Access. **2009**, *29* (19), 617–627.

(4) Lawrence, B. D. Processing of Bombyx Mori for Biomedical Applications. *Silk Biomater. Tissue Eng. Regen. Med.* **2014**, 78–99.

(5) Wenk, E.; Merkle, H. P.; Meinel, L. Silk Fibroin as a Vehicle for Drug Delivery Applications. *J. Control. Release* **2011**, *150* (2), 128–141.

(6) Hardy, J. G.; Torres-Rendon, J. G.; Leal-Egaña, A.; Walther, A.; Schlaad, H.; Cölfen, H.; Scheibel, T. R. Biomineralization of Engineered Spider Silk Protein-Based Composite Materials for Bone Tissue Engineering. *Materials (Basel).* **2016**, *9* (7).

(7) Fredriksson, C.; Hedhammar, M.; Feinstein, R.; Nordling, K.; Kratz, G.; Johansson, J.; Huss, F.; Rising, A. Tissue Response to Subcutaneously Implanted Recombinant Spider Silk: An in Vivo Study. *Materials (Basel).* **2009**, *2* (4), 1908–1922.

(8) Krüger, S.; Schwarze, M.; Baumann, O.; Günter, C.; Bruns, M.; Kübel, C.; Szabó, D. V.; Meinusch, R.; Bermudez, V. de Z.; Taubert, A. Bombyx Mori Silk/Titania/Gold Hybrid Materials for Photocatalytic Water Splitting: Combining Renewable Raw Materials with Clean Fuels. *Beilstein J. Nanotechnol.* **2018**, *9* (1), 187–204.

(9) Kaplan, D.; Adams, W. W.; Farmer, B.; Viney, C. Chapter 1 Silk : Biology , Structure , Properties , and Genetics. **1994**, 2–16.

(10) Schacht, K.; Scheibel, T. Processing of Recombinant Spider Silk Proteins into Tailor-Made Materials for Biomaterials Applications. *Curr. Opin. Biotechnol.* **2014**, *29* (1), 62–69.

(11) Matsumura, S.; Hlil, A. R.; Lepiller, C.; Gaudet, J.; Guay, D.; Shi, Z.; Holdcroft, S.; Hay, A. S. Ionomers for Proton Exchange Membrane Fuel Cells with Sulfonic Acid Groups on the End-Groups: Novel Branched Poly(Ether-Ketone)S. *Am. Chem. Soc. Polym. Prepr. Div. Polym. Chem.* **2008**, *49* (1), 511–512.

(12) Hermanson, G. T. PEGylation and Synthetic Polymer Modification. In *Bioconjugate Techniques*; Elsevier, 2013; pp 787–838.

(13) Veronese, F. M. Peptide and Protein PEGylation: A Review of Problems and Solutions. *Biomaterials* **2001**, *22* (5), 405–417.

(14) Carraway, K.; Koshland, D. E. United States Patent (19). **1977**, No. 19, 616–622.

(15) Sperinde, J. J.; Martens, B. D.; Griffith, L. G. Tresyl-Mediated Synthesis: Kinetics of Competing Coupling and Hydrolysis Reactions as a Function of PH, Temperature, and Steric Factors. *Bioconjug. Chem.* **1999**, *10* (2), 213–220.

(16) Hermanson, G. T. Zero-Length Crosslinkers. *Bioconjugate Tech.* **2013**, 259–273.

(17) Suri, S.; Ruan, G.; Winter, J.; Schmidt, C. E. *Microparticles and Nanoparticles*; 2013.

(18) Nakajima, N.; Ikada, Y. Mechanism of Amide Formation by Carbodiimide for Bioconjugation in Aqueous Media. *Bioconjug. Chem.* **1995**, *6* (1), 123–130.

(19) Williams, A.; Ibrahim, I. T. Carbodiimide Chemistry: Recent Advances. *Chem. Rev.* **1981**, *81* (6), 589–636.

(20) Carraway, K. L.; Triplett, R. B. Reaction of Carbodiimides with Protein Sulfhydryl Groups. *BBA - Protein Struct.* **1970**, *200* (3), 564–566.

(21) Cuatrecasas, P.; Parikh, I. Adsorbents for Affinity Chromatography. Use of N-Hydroxysuccinimide Esters of Agarose. *Biochemistry* **1972**, *11* (12), 2291–2299.

(22) Fiser, A. Computational Biology. **2010**, *673* (3).

(23) NCBI. BLAST Homepage and Selected Search Pages. *Ncbi* **2016**, 1–8.

(24) Wu, S.; Zhang, Y. LOMETS: A Local Meta-Threading-Server for Protein Structure Prediction. *Nucleic Acids Res.* **2007**, *35* (10), 3375–3382.

(25) Wu, S.; Skolnick, J.; Zhang, Y. Ab Initio Modeling of Small Proteins by Iterative TASSER Simulations. *BMC Biol.* **2007**, *5*, 17.

(26) Zhang, Y.; Kolinski, A.; Skolnick, J. TOUCHSTONE II: A New Approach to Ab Initio Protein Structure Prediction. *Biophys. J.* **2003**, *85* (2), 1145–1164.

(27) Yang, Z.; Daisuke, K.; Jeffrey, S. Local Energy Landscape Flattening: Parallel Hyperbolic Monte Carlo Sampling of Protein Folding. *Proteins Struct. Funct. Bioinforma.* **2002**, *48* (2), 192–201.

(28) Wu, S.; Zhang, Y. A Comprehensive Assessment of Sequence-Based and Template-Based Methods for Protein Contact Prediction. *Bioinformatics* **2008**, *24* (7), 924–931.

(29) Zhang, Y.; Skolnick, J. Automated Structure Prediction of Weakly Homologous Proteins on a Genomic Scale. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (20), 7594–7599.

(30) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.

(31) Sandhya, K.; Ravindranath, B. A Protocol for Racemization-Free Loading of Fmoc-Amino Acids to Wang Resin. *Tetrahedron Lett.* **2008**, *49* (15), 2435–2437.

(32) Carpino, L. A.; Han, G. Y. 9-Fluorenylmethoxycarbonyl Amino-Protecting Group. *J. Org. Chem.* **1972**, *37* (22), 3404–3409.

(33) Carpino, L. A. 1-Hydroxy-7-Azabenzotriazole. An Efficient Peptide Coupling Additive. *J. Am. Chem. Soc.* **1993**, *115* (10), 4397–4398.

(34) Subirós-Funosas, R.; Prohens, R.; Barbas, R.; El-Faham, A.; Albericio, F. Oxyma: An Efficient Additive for Peptide Synthesis to Replace the Benzotriazole-Based HOBt and HOAt with a Lower Risk of Explosion. *Chem. - A Eur. J.* **2009**, *15* (37), 9394–9403.

(35) Kappe, C. O.; Dallinger, D. Controlled Microwave Heating in Modern Organic Synthesis: Highlights from the 2004-2008 Literature. *Mol. Divers.* **2009**, *13* (2), 71–193.

(36) Collins, J. M.; Leadbeater, N. E. Microwave Energy: A Versatile Tool for the Biosciences. *Org. Biomol. Chem.* **2007**, *5* (8), 1141.

(37) Han, Y.; Albericio, F.; Barany, G. Occurrence and Minimization of Cysteine Racemization during Stepwise Solid-Phase Peptide Synthesis1,2. *J. Org. Chem.* **1997**, *62* (13), 4307–4312.

(38) Bacsa, B.; Horváti, K.; Bõsze, S.; Andreae, F.; Kappe, C. O. Solid-Phase Synthesis of Difficult Peptide Sequences at Elevated Temperatures: A Critical Comparison of Microwave and

Conventional Heating Technologies. *J. Org. Chem.* **2008**, *73* (19), 7532–7542.

(39) Palasek, S. A.; Cox, Z. J.; Collins, J. M. Limiting Racemization and Aspartimide Formation in Microwave-Enhanced Fmoc Solid Phase Peptide Synthesis. *J. Pept. Sci.* **2007**, *13* (3), 143–148.

(40) Wade, J. D.; Mathieu, M. N.; Macris, M.; Tregear, G. W. Base-Induced Side Reactions in Fmoc-Solid Phase Peptide Synthesis: Minimization by Use of Piperazine as N(α)-Deprotection Reagent. *Lett. Pept. Sci.* **2000**, *7* (2), 107–112.

(41) Yang, Z.; Jeffrey, S. Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins Struct. Funct. Bioinforma.* **2004**, *57* (4), 702–710.

(42) Stone, L.; Roberts, A. The Checkerboard Score and Species Distributions. *Oecologia* **1990**, *85* (1), 74–79.

(43) Xu, J.; Zhang, Y. How Significant Is a Protein Structure Similarity with TM-Score = 0.5? *Bioinformatics* **2010**, *26* (7), 889–895.

(44) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5* (4), 725–738.

(45) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat. Methods* **2014**, *12* (1), 7–8.

(46) Giesa, T.; Perry, C. C.; Buehler, M. J. Secondary Structure Transition and Critical Stress for a Model of Spider Silk Assembly. *Biomacromolecules* **2016**, *17* (2), 427–436.

(47) Microwave, A.; Synthesizer, P. Liberty Blue™ 600291 ·. 2016.

(48) Description, G.; Characteristics, C.; Considerations, S.; Guidelines, O. Agilent Zorbax Datasheet. *Control* **1996**, 5–6.

(49) Bulheller, B. M.; Hirst, J. D. DichroCalc—circular and Linear Dichroism Online. *Bioinformatics* **2009**, *25* (4), 539–540.

(50) Besley, N. A.; Hirst, J. D. Theoretical Studies toward Quantitative Protein Circular Dichroism Calculations. *J. Am. Chem. Soc.* **1999**, *121* (41), 9636–9644.

(51) Martz, E. Protein Explorer: Easy Yet Powerful Macromolecular Visualization http://proteinexplorer.org.

(52) Jeffrey, G. A. An Introduction to Hydrogen Bonding. *Oxford Univ. Press* **1997**.

(53) Zhong, L.; Johnson, W. C. Environment Affects Amino Acid Preference for Secondary Structure. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (10), 4462–4465.

(54) Efimov, A. V. Structure of α-α-Hairpins with Short Connections. *"Protein Eng. Des. Sel.* **1991**, *4* (3), 245–250.