POLITECNICO DI MILANO

School of Industrial and Information Engineering

Master's Program in Biomedical Engineering

Structural characterization of Self-Assembling Peptides via molecular dynamics simulations

Supervisor: Prof. Alfonso GAUTIERI

Author:

Marianna POLENGHI Matr. 915783

Academic year 2019-2020

RINGRAZIAMENTI

Desidero ringraziare il Professor Alfonso Gautieri per la completa disponibilità in tutto il periodo di svolgimento della tesi e per aver reso possibile la sensazione di appartenenza ad un gruppo, seppur a distanza.

Ringrazio Tiziana, Andrea, Antonio e Federico, i ragazzi del "group meeting" del martedì, che hanno saputo regalare dei momenti di spensieratezza e leggerezza in un periodo difficile e dei momenti di confronto che hanno mantenuto alta la motivazione.

Ringrazio Jacopo per aver facilitato le prime fasi del lavoro, aiutandomi ad acquisire delle competenze pratiche, per il rapporto di amicizia che si è instaurato e per aver sempre trovato un consiglio saggio.

Ringrazio i miei amici e il mio ragazzo Stefano per il costante incoraggiamento.

Ringrazio infine la mia famiglia: i miei nonni Luigino e Carmen, che non potrebbero essere più orgogliosi, mio fratello Luigi, per essermi stato d'esempio e d'aiuto in questi anni, la mia mamma Sabrina, per la dolcezza e resilienza che ha saputo trasmettermi sostenendomi sempre nel mio percorso.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

INTRODUCTION

Self-assembling peptides are a class of supramolecular biofunctional materials, held together mainly by noncovalent interactions (hydrogen bonds, hydrophobic interactions, van der Waals, electrostatic interactions, salt bridges, pi-stackings). They serve for different biomedical purposes, such as scaffolds for regenerative medicine and cell cultures, carriers for drug and gene delivery, hydrogels for probe immobilization in biochemical assays, nanosensors and semiconductors in bioelectronics. Given their property of excellent biocompatibility, self-assembling peptides are an emerging field of biomaterials.

Since their discovery in the ‘90s, self-assembling peptides have seen a gradually broader field of application. Many products including self-assembling peptides are patented, and various commercial products are nowadays available on the market for different biomedical and clinical purposes.

Researchers have observed the formation of fibrils, but the crystallographic structure is hard to obtain due to experimental limitations. Molecular Dynamics simulations can be used to assess the precise structure and chemical bonds involved in the formation of self-assembling peptides.

Different amyloid-like structures may be possible for a given peptide under different experimental conditions, allowing its folding. Therefore, there might be more than one stable configuration for each peptide. The aim of this work is to evaluate the stability of different possible configurations for seven peptide sequences, starting from pre-ordered structures that allow to observe the formation of fibrils within the simulation time available. The end goal is to identify the most probable structures of the fibrils observed experimentally through a stability analysis of the hypothesized configurations. Structures that tend to maintain the original configuration and interactions are considered stable. Furthermore, the investigation includes a comparison of different peptides, considering the influence of an amino acid substitution on self-assembly propensity and on the tendency to form a specific configuration.

MATERIAL AND METHODS

The peptide sequences are investigated through Molecular Dynamics simulations. Computational modelling is a tool to understand complex supramolecular interactions at the nanoscale level. It allows to overcome some experimental limitation, in terms of space and time resolution, and provides insights that can be combined with structural characterization techniques to obtain a complete picture.

Seven peptide sequences were chosen according to directions given by National Research Council (CNR) researchers, who are currently working on the experimental characterization of these self-assembling peptides. The specific computational approach was suggested by a previous thesis work, that established that the only way to observe fibrils formation in acceptable computational times was to start from pre-ordered organizations of peptides in space.

The work flow is analogue for all peptides and follows these main steps: the peptide construction as a pdb file containing the amino acidic sequence, the construction of β-strands and β-sheets with specific scripts with the instructions for translations and rotations of peptides in space, to obtain 8x8 models (where 8 β-sheets are composed by 8 β-strands each). For each peptide, 10 possible amyloid-like configurations are considered. Every configuration is then solvated in a water box with dimensions in the order of hundreds of Å and simulated for a time of 150 ns, which is a trade-off between the maximum time to contain computational cost and the minimum time to observe fibrils formation.

The peptides that require a particular construction procedure are those containing a halogenated-phenylalanine and peptides with modifications of a phenylalanine residue.
In the first case the topology and parameters of halogenated phenylalanine need to be obtained in order to build the sequence.

The stability analysis includes both quantitative parameters and a qualitative observation of the simulations. In particular, the indicators of stability are: Root Mean Square Deviation, Root Mean Square Fluctuations, hydrogen bonds, β-sheets, hydrophobic contacts, native contacts. The Stability Index is a measure that classifies configurations in terms of stability and mainly accounts for backbone interactions. Native contacts provide information on lateral residues interactions.

RESULTS AND CONCLUSIONS

Different peptide sequences have a different stability for each possible configuration. For every peptide, a few stable configurations - that maintain the original hypothesized structure - can be identified. Similarly, a few configurations that depart from the initial structure, and are therefore considered unstable, can be recognized.

It is also possible to compare configurations among the aminoacidic sequences and find similarities of behavior (in terms of shape, interactions and residue disposition). An example of the most common shapes assumed by the configurations is reported in Figure 1.

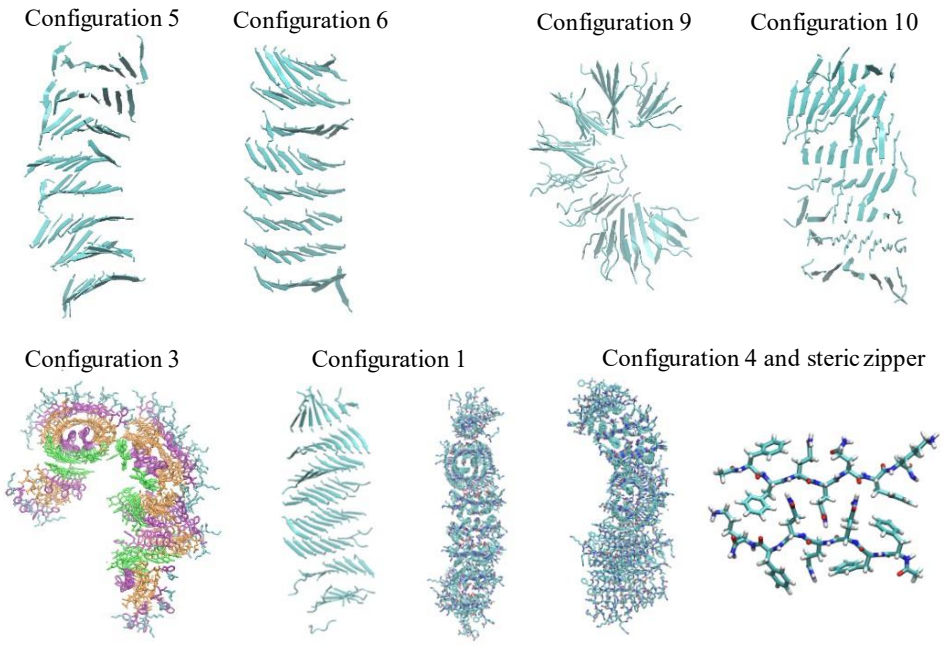

Figure 1 examples of configurations

Configuration 5, 6 and 7 result to be stable for most peptide sequences, indicating a preference of antiparallel β-strands to form compact fibrillar structures composed of β-sheets.

Configuration 9 and 10 always result unstable. Configuration 10, in fact, visibly loses compactness and order and the β-sheets structure disintegrates. However, configuration 9 seems to be prone to forming vesicle-like structures with a hydrophobic core, that might result stable even if they differ from the initial fibrillar form. Configuration 3 also displays a similar U-shape and analogue residue disposition in most peptides.

Configuration 1 and 4 are often classified with an intermediate stability from the analysis, but the visual observation and native contacts percentage reveal that they are characterized by a very ordered lateral packing, where residues are disposed in a steric zipper.

Future developments of this work can consist in longer simulations or more complex fibril models to obtain even more realistic systems. Another aspect might be the experimental characterization of the investigated peptide sequences, to validate the results. Observations extracted from simulations on the behavior of different configurations for each sequence can be useful in combination with peptide synthesis and structural characterization. In fact, the comparison allows to confirm whether the predicted stable structures are actually those of the experimentally observed fibrils.
Other future developments might be to introduce specific modifications to the fibrils to make them even more stable or tunable to environmental conditions. This would further expand the possible field of applications.

SOMMARIO

INTRODUZIONE

I peptidi autoassemblanti sono una classe di materiali sopramolecolari biofunzionali, tenuti insieme principalmente da interazioni non covalenti (legami idrogeno, interazioni idrofobiche, aromatiche, elettrostatiche, di van der Waals, ponti salini). Sono sfruttati in diverse applicazioni biomediche: come scaffold per medicina rigenerativa e colture cellulari, carriers per rilascio controllato di farmaci e molecole, idrogeli nei saggi biochimici e nanosensori in bioelettronica. Data l'eccellente biocompatibilità, i peptidi autoassemblanti sono un campo di studio emergente tra i biomateriali.

Dalla loro scoperta negli anni 90, i peptidi autoassemblanti hanno visto uno sviluppo e un ampiamento nei campi di applicazione. Attualmente vi sono diversi brevetti e prodotti commerciali disponibili che includono i peptidi autoassemblanti per vari scopi clinici.

Sperimentalmente è stata osservata la formazione di fibrille, ma la struttura cristallografica è spesso difficile da ottenere, aspetto che limita la caratterizzazione strutturale. Le simulazioni di dinamica molecolare possono investigare la struttura e i legami chimici coinvolti nel processo di auto-assemblaggio delle fibrille.

Sono possibili diverse strutture simil-amiloidi per un dato peptide, a seconda delle diverse condizioni sperimentali, che ne permettono il folding. È quindi possibile che ci sia più di una struttura stabile per ciascun peptide. Lo scopo del lavoro è valutare la stabilità di diverse possibili configurazioni per sette sequenze peptidiche, partendo da strutture ordinate, che consentono di osservare la formazione di fibrille nel tempo di simulazione. L'obiettivo finale è identificare le strutture più probabili delle fibrille osservate sperimentalmente, attraverso un'analisi di stabilità delle configurazioni ipotizzate. Le strutture che tendono a mantenere la configurazione e le interazioni originali sono considerate stabili. Inoltre, lo studio include diversi peptidi, in modo da valutare l'effetto di varie sostituzioni amminoacidiche sulla propensione ad autoassemblare e sulla tendenza ad assumere una determinata configurazione.

MATERIALI E METODI

Le sequenze peptidiche sono studiate tramite simulazioni di dinamica molecolare. I modelli computazionali costituiscono uno strumento per comprendere le complesse interazioni intermolecolari alla nanoscala. Permettono di superare alcuni limiti di risoluzione spaziale e temporale dei metodi sperimentali, fornendo delle informazioni complementari alle tecniche di caratterizzazione strutturale.

Sono stati selezionati sette peptidi, su indicazione di attuali studi sui peptidi autoassemblanti svolti da alcuni ricercatori del Consiglio Nazionale delle Ricerche (CNR). Il tipo di approccio computazionale è stato scelto sulla base di un precedente lavoro di tesi, che ha stabilito la necessità di partire da configurazioni precise ed ordinate (anziché da una disposizione casuale dei residui) al fine di osservare la formazione di fibrille nei tempi di simulazione.

Il processo del lavoro è analogo per tutti i peptidi e segue i seguenti passaggi fondamentali: la costruzione di un peptide come file pdb contenente la sequenza amminoacidica, la costruzione di β-strands e β-sheets con specifici script di rotazione e traslazione nello spazio dei peptidi per ottenere dei modelli 8x8 (in cui 8 β-sheets sono composti ciascuno da 8 β-strands). Per ogni peptide si considerano dieci possibili configurazioni di fibrille simil-amiloidi. Ogni configurazione viene solvatata in un box d'acqua di dimensioni dell'ordine di centinaia di Å e si svolge una simulazione di 150 ns, un compromesso che consente di osservare la formazione di fibrille e minimizzare il costo computazionale.

I peptidi che richiedono una procedura di costruzione più complessa sono quelli contenenti la fenilalanina alogenata e residui di varianti della fenilalanina (enantiomero D, omofenilalanina, β-fenilalanina). Questo rende necessario rispettivamente la costruzione o la modifica di file contenenti topologia e parametri per la fenilalanina alogenata o modificata, al fine di costruire la sequenza di amminoacidi.

L'analisi di stabilità include parametri quantitativi e l'osservazione qualitativa delle simulazioni. In particolare, gli indicatori di stabilità sono: Root Mean Square Deviation, Root Mean Square Fluctuation, legami idrogeno, contatti idrofobici, contatti nativi. Questi ultimi tengono conto delle interazioni dei residui laterali dei peptidi. Lo Stability Index

fornisce un numero che consente di classificare le configurazioni per stabilità e tiene conto principalmente delle interazioni del backbone.

RISULTATI E CONCLUSIONE

Diverse sequenze peptidiche hanno una diversa stabilità per ogni possibile configurazione. Si possono identificare, per ciascun peptide, alcune configurazioni stabili che mantengono la struttura originariamente ipotizzata e alcune instabili che vi si discostano.

È inoltre possibile comparare le configurazioni tra le varie sequenze amminoacidiche e trovare delle similitudini in termini di forma, disposizione dei residui e interazioni.

Le configurazioni 5, 6 e 7 risultano stabili per la maggior parte dei peptidi, indicando una preferenza dei β-strands antiparalleli a formare fibrille compatte e stabili.
Le configurazioni 9 e 10 risultano sempre instabili. La configurazione 10 perde la struttura a β-sheets. La configurazione 9, tuttavia, sembra avere la tendenza a formare strutture a vescicola con un interno idrofobico, che potrebbero risultare stabili pur discostandosi dalla struttura a fibrilla inizialmente ipotizzata. La configurazione 3 ha un comportamento simile. Le configurazioni 1 e 4 sono spesso classificate dall'analisi con una stabilità intermedia, ma l'osservazione visiva permette di apprezzare una disposizione laterale dei residui molto ordinata, che minimizza l'ingombro sterico.

Sviluppi futuri possono consistere in simulazioni più lunghe o modelli di fibrille più complessi per ottenere sistemi ancora più realistici. Un altro aspetto può essere la caratterizzazione sperimentale dei peptidi, per validare i risultati. Le osservazioni tratte dalle simulazioni sul comportamento delle varie configurazioni possono essere utili in combinazione con la sintesi dei peptidi e la caratterizzazione sperimentale. Un simile confronto, infatti, potrebbe confermare se le strutture predette come stabili siano effettivamente quelle che originano le fibrille osservate sperimentalmente.

Ulteriori sviluppi potrebbero essere l'introduzione di specifiche modifiche alle fibrille per renderle ancora più stabili o regolabili per rispondere a diversi tipi di stimoli esterni, in modo da espandere ulteriormente i campi di applicazione.

1 INTRODUCTION

1.1 SUPRAMOLECULAR BIOFUNCTIONAL MATERIALS: SELF-ASSEMBLY

Supramolecular chemistry is a multidisciplinary field that integrates chemistry, physics and biology to investigate supramolecular systems, formed by smaller molecular building blocks, that originate ordered 3D structures through noncovalent interactions. [1]

Supramolecular biofunctional materials are a class of biomaterials in which small molecules are bonded through self-assembly mainly by noncovalent interactions: hydrogen bonds, hydrophobic interactions, van der Waals forces, electrostatic interactions, salt bridges, and pi-stackings in case of noncovalent interactions between aromatic rings. [2]

Self-assembly is defined as the spontaneous and reversible aggregation of molecules via noncovalent bonds into hierarchical ordered structures. The process is driven by the intrinsic information contained in the components themselves. Nature presents many examples of supramolecular materials: the DNA double helix is formed by self-assembly of two strands, held together through hydrogen bonds and aromatic interactions and each strand itself is formed by nucleotides. [3]

Noncovalent interactions make self-assembling peptides particularly interesting and versatile, because they allow materials to have responsiveness to external stimuli and reversibility, so that they adapt to local changes. Furthermore, supramolecular biofunctional materials can be tuned to respond to different kinds of stimuli, such as temperature, pH variations and ionic strength among others. Another important feature is adaptiveness, which means that certain properties can be changed in a predictable way, making them useful in a wide variety of applications. In case of the biomedical field, using building blocks like peptides, nucleic acids or saccharides can give rise to biomimetic materials, able to better mimic the natural function and environment. [2]

The most important class of supramolecular biofunctional materials is constituted by supramolecular hydrogels, an emerging field that overcomes some limitations of polymeric hydrogels. Conventional polymeric type II hydrogels are exploited as scaffolds in regenerative medicine and tissue engineering and cell cultures thanks to their

biocompatibility and the ability to mimic the extracellular matrix. The noncovalent nature of the interactions among chains not only confers tunability to external conditions, but also makes the process of hydrogel formation reversible, limiting their use for some applications. Supramolecular biofunctional hydrogels form as a result of the self-assembly process of hydrogelators in water. The self-assembly of small building blocks via noncovalent interactions originates a supramolecular chain or nanofiber (or other nanostructures). These chains further assemble in a 3D network, again though noncovalent interactions, that retains and immobilizes water molecules inside. In case of self-assembly of hydrogelators in water, the main noncovalent interactions are hydrophobic interactions.

The most important advantage is reversibility, which means that the material can restore its mechanical properties after deformation, but the main disadvantage is the mechanical strength. However, supramolecular hydrogels have excellent properties of tunability and responsiveness to stimuli, biocompatibility, degradation/clearance and allow the incorporation of multiple biological functionalities.

Compared to polymeric hydrogels, supramolecular hydrogel have better responsiveness, which facilitates the control over assembly and disassembly, allowing a regulation of the process of formation or degradation of the hydrogel through external stimuli *in situ*. [2]

Supramolecular biofunctional materials can be constituted by a variety of building blocks, including the basic biological building blocks like proteins, nucleic acids, polysaccharides. They are versatile and particularly suitable for biomedical applications. Amino acids are the most frequently used and investigated and constitute self-assembling peptides. [2]

1.2 SELF-ASSEMBLING PEPTIDES

Self-assembling peptides started to be object of research since their serendipitous discovery in the 90's by studying a yeast protein, zuotin, having a short segment (EAK16-II) that can assemble in nanoscaffolds. Zahng et al. [4] found that this sequence Ac-(AEAEAKAK)2-NH2, with the alternance of oppositely charged (positively charged lysine and negatively charged glutamic acid) and hydrophobic residues (alanine) can originate β-sheets in which hydrophobic and charged residues are arranged in opposite sides of the backbone.

These peptides consequently have a hydrophilic and a hydrophobic side, and in water they assemble into nanofibers by disposing two hydrophobic sides towards each other, while the

hydrophilic sides towards the solvent constitute the outside of nanofibers. The formation of nanofibers starting from a single peptide is shown in Figure 2.

After over 20 years, in 2011, self-assembling peptides began to be successfully used in human clinical trials.

Nowadays, the large-scale production allows commercialization of various products for different purposes, spanning from fields of regenerative medicine, to biological research, to biosensors and bioelectronics. [4]

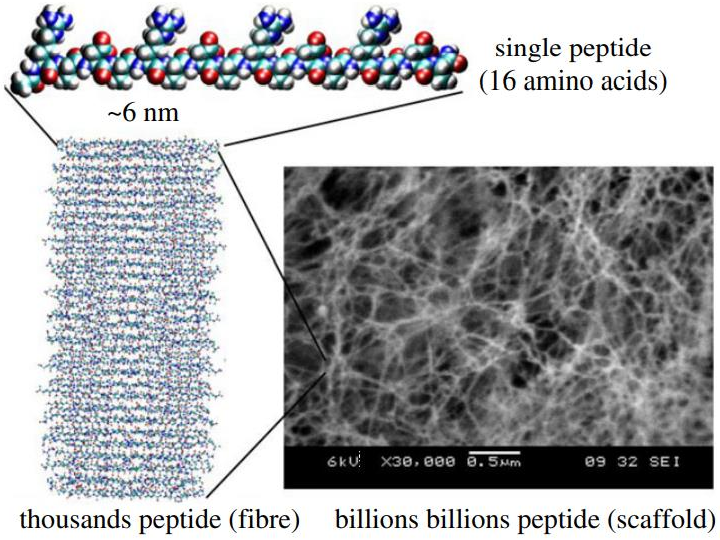

Figure 2 RADA16-I self-assembling peptide: peptide and nanofiber modelling and SEM image (scale bar 500 nm), Zhang et al. 2017

It was shown that many self-assembling peptides have β-sheets secondary structures. Even α-helix is possible as it is exploited in *de novo* design of fibrillar nanostructures [5], but β-sheet is prevalent and widely studied.

Once the monomers (amino acids composing the primary structure) are assembled into their secondary structure, the formation of different types of nanostructures is possible: nanofibers, nanotubes and nanovesicles have been observed. Short peptides can also form film monolayers and gels.[6]

Amino acids conformation might induce a preferred secondary structure, but also environmental conditions and interactions with neighboring residues affect the self-assembling peptide secondary structure and supramolecular architecture. Therefore, different variables - that might differ from the case of folded proteins - must be accounted for. This aspect makes it difficult to predict the final structure of a self-assembling peptides,

even when structural information on the amino acids conformation in the folded protein is given. [7]

There are many advantages of using building blocks of life like peptides in supramolecular biofunctional materials. Chemical synthesis is rapid and affordable, and both structural and functional information is available for most peptides building blocks. This simplifies their production. Self-assembling peptides are easily tunable to respond to different kinds of environmental condition such as pH, ionic strength, and temperature. Other advantages of using peptides as building blocks is the adoption of precise and ordered secondary structures (α-helix, β-hairpin or more often β-sheets) that can energetically favor the self-assembly process. Moreover, even though the assembly and disassembly of self-assembling peptides is a dynamic process, the reassembly happens spontaneously without the need for catalysts or mechanical processing, unlike conventional polymers. [4]

The main interactions that guide the process of self-assembly in water are hydrophobic and van der Waals interactions, arrays of electrostatic interactions between oppositely charged residues, and hydrogen bonds, especially between peptide backbones. [4] Aromatic residues are known to produce ordered structures through π-π stackings of aromatic rings.[6] In case of peptides displaying a halogenated aromatic ring, it is also possible to observe the formation of halogen bonds between the partial positive charge of the covalently bonded halogen atom and another electronegative atom like nitrogen.

1.3 CHEMICAL BONDS: NONCOVALENT INTERACTIONS

Noncovalent interactions are weaker than covalent interactions, since their energy ranges respectively from 2 kJ mol^{-1} to 300 kJ mol^{-1} compared to 150-450 kJ mol^{-1}. [3] However, when combined cooperatively, noncovalent interaction can stabilize a supramolecular structure. They include hydrogen bonds, hydrophobic interactions, van der Waals interactions, electrostatic interactions, salt bridges and aromatic interactions. A brief description of each interaction is reported from a chemical perspective.

1.3.1 Hydrogen bonds

Hydrogen bonds have a key role in the formation of supramolecular architectures and can be classified as electrostatic interactions. A hydrogen atom is shared between two other atoms. The hydrogen bond donor is the group containing the atom (usually oxygen or nitrogen) that is more strongly linked to the hydrogen. The electronegativity of the atom bonded to the hydrogen induces a partial positivity (δ^+) of the hydrogen itself. This allows another electronegative atom (δ^-), belonging to the hydrogen bond acceptor, to attract the hydrogen through electrostatic interaction. [8]

Hydrogen bonds are fundamental in the formation of protein secondary structures.
Self-assembling peptides can arrange in different secondary structures (β-sheet, α-helix, random coil), but the most frequent secondary structure experimentally observed seems to be β-sheet. In fact, peptides that contain the RAD motif, which are similar to peptide EAK16, were shown to assemble into ordered nanofibers with a β-sheets secondary structure. The role of β-sheets is more important especially in supramolecular hydrogels, whereas α-helix secondary structure is less likely to form hydrogels. [2]

Hydrogen bonds lead to β-sheets formation. In this case different β-strands are held together by noncovalent interactions and amino acid residues belonging to two strands are brought in close proximity, forming a hydrogen bond between the C=O group of one residue and the N-H group of the other.

In Figure 3 it is possible to observe hydrogen bonding in parallel and antiparallel β-sheets. This diagram shows two parallel (A) and antiparallel (B) β-strands, interacting through hydrogen bonding. The black boxes display three different amino acid pairs. In case of parallel strands, one amino acid is hydrogen bonded while the corresponding amino acid in the pair is non hydrogen bonded. On the contrary, in case of antiparallel β-strands both amino acids in the pair are either hydrogen bonded or non-hydrogen bonded, giving rise to a more ordered and compact structure. [9]

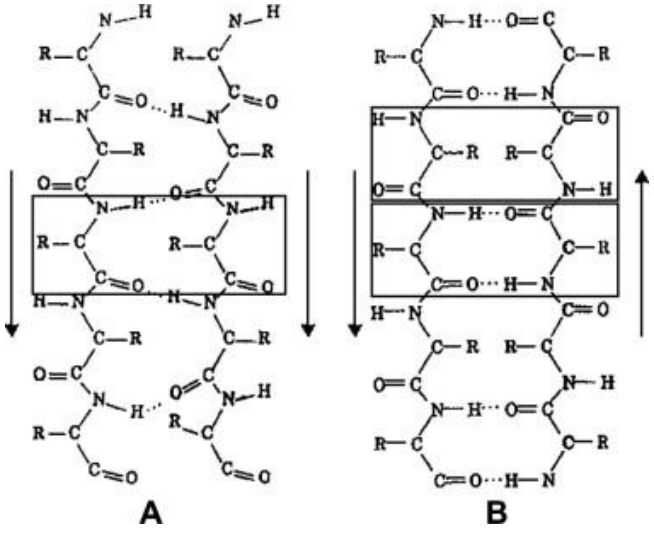

Figure 3 hydrogen bonds in parallel and antiparallel β-sheets, Zhang et al. 2009

Other interactions between peptides in β-sheets, a part from noncovalent hydrogen bonds along the backbones are: arrays of electrostatic interactions between positive and negative charges, hydrophobics interactions and van der Waals interactions, water-mediated hydrogen bonds. [4]

1.3.2 Hydrophobic interactions

Hydrophobic interactions are generated from the repulsion of non-polar groups toward the aqueous solvent. Non-polar groups tend to exclude themselves from the environment and locate at the core of the structure creating a hydrophobic cavity (Figure 4), so that the external part exposes polar hydrophilic group and this disposition minimizes the energy. [3]

The hydrophobic effect, that causes non-polar groups to orient towards the inner core of a protein to avoid being in contact with water molecules, is the major driving force in protein folding and strongly contributes to protein stability. [10]

Hydrophobic contacts also have a role in stabilizing the protein structure once the process of folding is completed.

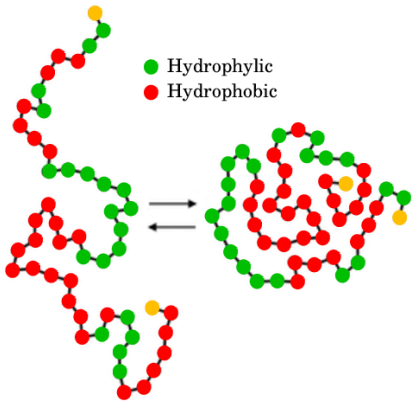

Figure 4 disposition of hydrophobic residues at the inner core in protein folding, A. Sengupta, 2011

1.3.3 Van der Waals

Van der Waals forces rely on the fact that the distribution of the electronic cloud around an atom is transient, changing in time because it is subjected to fluctuations. An instantaneous dipole is formed within the molecule itself. The asymmetry of charge distribution in one atom causes an asymmetry of charge distribution in a neighboring atom, so that they attract: the partial positive charge of one atom attracts the complementary partial negative charge of the other one.

The closer the atoms, the higher the attraction force between the two, up to a point (van der Waals contact distance, r_0 in Figure 5, beyond which repulsion occurs, as a consequence of overlapping electronic clouds. These interactions are stronger if the species are easier to polarize. [8]

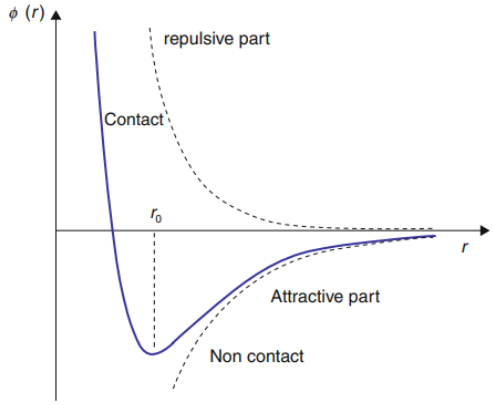

Figure 5 van der Waals forces and intermolecular potential as a function of atoms distance (Zhang, 2013)

1.3.4 Electrostatic interactions

Electrostatic interactions include ion-ion interactions, ion-dipole interactions, dipole-dipole interactions.

Ion-ion interactions are the strongest and the energy follows the Coulomb's law:

$$E = k * \frac{q_1 * q_2}{Dr}$$

Where q_1 and q_2 are the charges of the atoms (expressed in units of electronic charge), D is the dielectric constant of the medium, r is the distance between the two atoms (expressed in Angstrom) and k is a proportionality constant (k=332 to give energies in Kcal/mol or k=332 to give energies in KJ/mol). [8]

Ion-ion interactions are represented in Figure 6a, in which alternated cations and anions attract.

Ion-dipole and dipole-dipole interactions are weaker (the energy ranges from 50-200 and 5-50 kJ mol^{-1} respectively) and they have directionality, so they require a specific alignment of the two species to be established, therefore they form in a narrower variety of cases compared to ion-ion interactions. These interactions are represented in Figure 6b. However, dipole-dipole have the most relevant role in bringing alignment, since it is necessary for both species to be spatially oriented in a certain way, even though they are the weakest. [3]

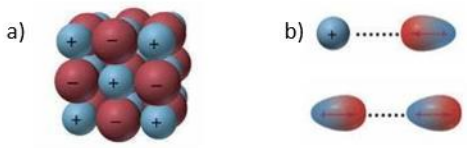

Figure 6 a) ionic interactions, b) ion-dipole (top) and dipole-dipole (bottom) interactions, https://chem.fsu.edu/chemlab/chm1046course/interforces.html

There are two types of noncovalent interactions in proteins: specific and non-specific. Specific interactions are mainly electrostatic, whereas non-specific interactions are hydrophobic and van der Waals. Electrostatic interactions have an important role in the process of protein folding, and directly impact protein structure, stability and function. [11]

1.3.5 Salt bridges

Salt bridges, which can be considered a special form of hydrogen bonds, are composed of negative charges from Asp, Glu, Tyr, Cys and the C-terminal carboxylate group, and of positive charges from His, Lys, Arg and the N-terminal amino group. Since the side chain charge of these residues depends on pH, the free energy contributions of salt bridges to protein stability are pH dependent. An example of salt bridge is reported in Figure 7.

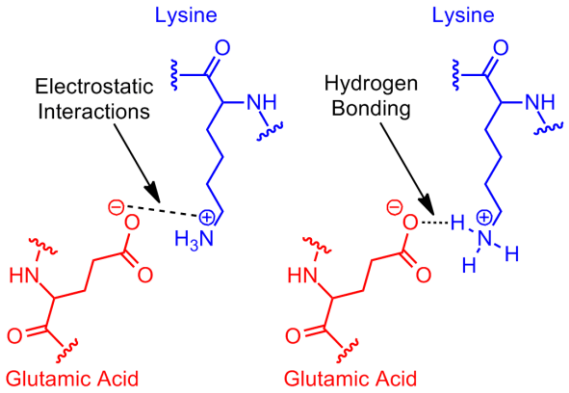

Figure 7 examples of salt bridge between amino acid lysine and glutamic acid, source: https://en.wikipedia.org/wiki/Salt_bridge_(protein_and_supramolecular)

The charge of amino acids side chain depends on pH, therefore the ability to form salt bridges is pH dependent. The pH affects the charge, which in turn affects protein stability. The different contributions in free energies from salt bridges depending on pH explains why salt bridges can be stabilizing or destabilizing for the structure. [10]

Moreover, these bonds can give favorable or unfavorable energetic contributions because the net contribution of a salt bridge is given by electrostatic interaction and desolvation and structural ordering. These interactions are respectively favorable and unfavorable and their balance determines the overall contribution. [10]

Salt bridges are found in proteins secondary and tertiary structure and form between two oppositely charged residues that are close enough (so that two or more heavy atoms are within hydrogen bond distance) to form electrostatic interactions.

Salt bridges are generally located in the parts exposed to the solvent, so they interact with the external environment, specifically with water but also with cosolvents like ions. Therefore, the electrostatic attraction is influenced by the interposed water molecules. [12]

Salt bridges form a network of electrostatic interactions in native proteins and are thought to confer rigidity to the protein structure because they are generally not found in flexible regions. Despite the fact that they seem to constrain motion, salt bridges can either stabilize or destabilize the structure. [11]

1.3.6 Aromatic interactions

Aromatic-aromatic interactions are of great importance since aromatic peptides containing aromatic amino acids are shown to have a high propensity for self-assembly and drive this process of amyloid fibrils formation. Studies have proven that simple phenylalanine is able to form amyloid-like fibrils in an aqueous environment. [13]

Amino acids like phenylalanine, containing an aromatic ring, are particularly interesting because phenylalanine is found in β-amyloid fibrils, involved in amyloid plaques formation, that lead to the development of Alzheimer disease. In this case the aromatic interactions are so that the peptide backbone orients itself perpendicular to the fibril axis. The peptides are in β-sheets structures.

The substitution of aromatic residues to aliphatic residues with similar hydrophobicity in a peptide sequence (all else being equal) is thought to have an impact on the nanofiber morphology, that can be explained by the different hydrogen bonding orientation (antiparallel instead of parallel). As a consequence, this further influences the higher hierarchical structures that form. [14]

The π-π stacking (also called pi-stacking) are noncovalent interactions among aromatic groups. They can be classified in three categories, as showed in Figure 8, based on the disposition of the rings: (A) edge-to-face stacked (or T-shaped), (B) offset stacked, (C) face-to-surface stacked. [15]

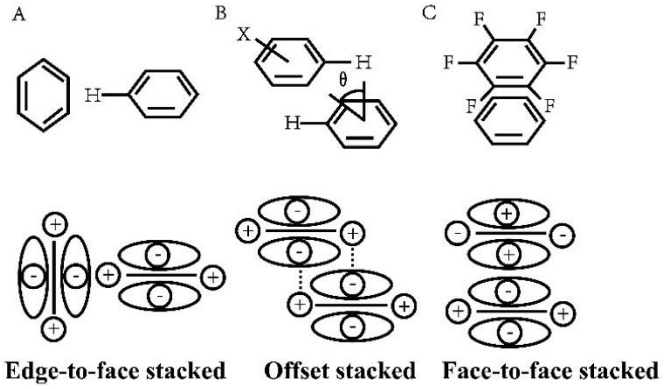

Figure 8 classification of pi-stacking interactions, Zhuang et al. 2019

The first two configurations, edge-to-face and offset stacked, have higher energy than face-to-face. The energy range is 1-50 kJ mol^{-1}. The distance between centroids of two parallel structures is around 3.5 Å and the angle (Θ) is less than 20°. [15]

These interactions are largely exploited in drug delivery systems (together with hydrogen bonding) because many drugs have aromatic rings and π-π stacking are influenced by environmental conditions like pH, so they simplify the process of triggering for assembly or disassembly, making it possible to tune the delivery as a function of pH. [15]

1.4 HALOGENATED PEPTIDES

Several modifications of the aromatic ring have been investigated in recent years in order to increase the gelation efficiency of self-assembling peptides, including the introduction of halogen atoms.

The work of Nilsson and co-workers suggests that the number and the position of halogen groups influences the behavior of hydrogels by acting on the noncovalent interactions between the gelator molecules in the aromatic amino acids. [13]

The work of Singh et al. evaluates the impact of a modification in the aromatic ring on the process of self-assembling, by introducing a nitro group. In particular, the residue is 4-nitrophenylalanine (4NP) was investigated in DMSO (Dimethyl Sulfoxide), in water, and in mixed solvents of both water and DMSO, to assess its behavior. The residue 4NP is already known to form crystals in water, but it was discovered that it forms a gel in DMSO. In a

mixed solvent, the self-assembly depends on the chemical composition of the solvent, which interacts mainly via hydrogen bonding and charge-transfer interactions with 4NP. [13]
The inclusion of a halogen atom in the aromatic ring changes the supramolecular interactions, affecting the physical and rheological properties of the resulting self-assembling peptides gel. [13]

Halogen atoms are often used in aromatic residues, especially Phenylalanine, to tune aromatic interactions in peptides and regulate the process of self-assembly, by acting on the supramolecular interactions. [13] Phenylalanine is relevant because it is one of the main residues involved in amyloid fibrils. It was observed that 'Phe-Phe' is the core residue for amyloid fibril formation. [13] It was shown that halogenation at the *p*-position of the benzene ring of Phenylalanine can promote amyloid self-assembly. [16] This studies confirms that the type and the position of the halogen substitution influences the supramolecular interactions and therefore the ability to form gels. [16]

1.4.1 Halogen bond

The strategy of introducing halogen atoms in aromatic residues is particularly effective to favor self-assembly thanks to halogen bonds.
A halogen bond is a directional, noncovalent, attractive interaction between a halogen atom and an electronegative atom. Halogen atoms usually display a negative charge and cannot interact through electrostatic interactions with an electronegative atom. However, when bound covalently to another atom, halogen atoms are subjected to the formation of a region of positive charge, defined as σ-hole, located at the opposite side of the covalent bond. The anisotropic distribution of charge around the halogen atom, which gives a positive region, makes it possible for the halogen atom to interact with another electronegative atom though halogen bond. [17]

The halogen bond, schematized in Figure 9, arises between a donor halogen atom X (covalently bonded to another atom A) and an acceptor that is a Lewis base Y. Halogens - that play an important role in noncovalent interactions - therefore behave as a Lewis acids in halogen bonds, compared to hydrogen bonding, in which they act as bond acceptors. [18] This peculiar charge disposition and the partial positive region is one of the main challenges

for the parametrization of halogen atoms in molecular modelling, since the charge is usually represented as unitary and negative.

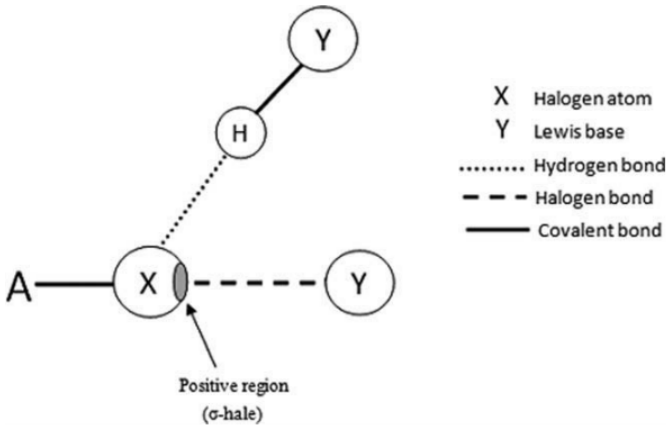

Figure 9 schematization of the halogen bond, Ibrahim 2011

1.5 OVERVIEW OF EXPERIMENTAL TECHNIQUES FOR SELF-ASSEMBLING PEPTIDES CHARACTERIZATION

The most robust technique for structural characterization of proteins is X-ray crystallography. There are also other experimental techniques that can be combined to obtain a complete picture of the structure: solution-phase Nuclear Magnetic Resonance (NMR), light absorbance measures such as Circular Dichroism (CD), Wide Angle X-ray Scattering, Small Angle X-ray Scattering (WAXS, SAXS).

X-ray crystallography is technique that allows to determine the 3D structure of a molecule, a small organic molecule or material (and it is typically used for proteins and biological macromolecules), starting from X-ray diffraction patterns.

The pattern of diffraction provides information on the packing symmetry and the size of the repeating cell in the crystal, the intensity of the spots makes it possible to reconstruct a map of electron density, from which the resulting structure is fitted.

The analyzed substance is required to be in the crystalline form, therefore different crystallization procedures are previously implemented. It is not always possible to obtain a crystal because if the process – based on the principle of extracting the crystal from a high-concentration solution of the protein – happens too fast, precipitation occurs. In the correct

conditions, crystals can grow instead. To be scanned by the X-ray beam, the longest dimension of the crystal must be at least 0.1 mm, so that the volume is sufficient. [19]

NMR is a characterization technique that provides information at the atomistic level and, in addiction to X-ray diffraction, it has been used to obtain many structures of known proteins for years. Many structures present in the Protein Data Bank were obtained from this technique. While crystallography requires the formation of crystals, NMR is usually carried out in solution, under conditions close to physiological conditions. The strategy of structure determinations follows a few steps. First, protein solution preparation, followed by NMR spectroscopy. The next step is the obtainment of NMR frequency spectrum through the Fourier transform (where atoms are represented by their resonance frequencies), and the assignment of NMR signals to each atom. Lastly, the collection of conformational constraints from which the 3D structure can be calculated. [20]

Circular Dichroism (CD) is a light absorbance measure that, unlike the previous two techniques, does not provide residue-specific information. It is exploited for the determination of proteins secondary structure, protein binding and folding. The basic principle is the difference in light absorption (right-handed or left-handed circularly polarized light). Different structural features (for instance α-helix or β-sheets) produce different CD spectra. [21]

As far as scattering techniques are concerned, SAXS and WAXS are among the most used for biomolecular complexes. Wider angles of scattering correspond to more detailed structural features of proteins, small angle data usually reflect the global shape and size. The samples are exposed to a collimated, focused X-ray beam and from the resulting scattering intensity pattern the tridimensional structure can be obtained. [22]

1.6 LIMITATIONS OF EXPERIMENTAL TECHNIQUES

Experimental techniques for the structural characterization of self-assembling peptides lack the ability to give complete information when used alone.

The spatial resolution is not enough to provide the exact conformation of single building blocks composing the supramolecular structure. The molecular packing investigation usually

requires the synergic combination of various experimental techniques in order to be fully understood.

Not every self-assembling peptide can be characterized by crystallography for example, because the starting crystal structure is not always available. The nature of the nanostructures in their solvent itself can prevent researchers from using some structural determination analysis, such as X-Ray crystallography or solution-phase NMR (Nuclear Magnetic Resonance). Many times, different experiments, dealing with smaller parts of molecular structures, are carried out to provide a complete information together with crystallographic studies. For instance, measure of light absorbance and scattering can be useful (using UV light, Infrared IR, Circular Dichroism CD for light absorbance and Dynamic Light Scattering DLS, Wide Angle/Small Angle X-Ray Scattering WAXS/SAXS).

There are three main limitation in the experimental approach. First, every technique measures molecules conformational means, making it possible to obtain wrong models or idealized interactions for each molecule, that might differ from real ones. Secondly, data from X-Ray experiments is usually interpreted based on empirical models that were developed for those macromolecular system and do not have general validity. Lastly, laboratory experiments with a temporal scale of weeks make it hard to catch the first few steps of self-assembly, in the order of nanoseconds. Molecular Dynamics simulations provide the necessary temporal resolution instead.[23]

1.7 ROLE OF MOLECULAR DYNAMICS SIMULATIONS

Self-assembling peptides can spontaneously form different types of nanostructures like tubes, filaments, fibrils, vesicle, hydrogels and monolayers.

The crystallographic structure of peptides is not always available, so it is difficult to establish the secondary and tertiary structure that a given peptide sequence will assume. One of the main goals of Molecular Dynamics (MD) simulations is to identify chemical bonds and molecular interactions that determine the final nanostructure of the material. In case the crystallographic structure can be obtained, simulations can provide an important feedback to design and evaluate candidates self-assembling peptides.

Molecular interactions that characterize self-assembling peptides determine their final nanostructure, and subsequently the morphology in solution. It is not possible to predict which peptides self-assemble, the exact mechanism and the final structure, given the amino acid sequence. Molecular dynamics simulations can provide a first screening tool to select the best suited candidates among many possible amino acidic combinations.

A possible way to overcome experimental limitations is to use classical molecular simulation. This computational method traces the motion of individual atoms or molecules in time. It has been used in the past to add qualitative observations on the behavior of supramolecular polymers. It has gradually gained relevance also in the field of self-assembling peptides, both in their design and in the validation of experimental results as a supporting strategy.
Computational modelling is a tool to understand complex supramolecular interactions.

1.8 EXAMPLES OF COMPUTATIONAL STUDIES

Two practical examples of Molecular Dynamics simulations used for self-assembling peptides are reported. The first one is the work of Fredrix et al.[23, 24], who simulated the 8000 possible combinations of tripeptides using the Martini Force Field to understand the design rules for short self-assembling peptides. From the *in silico* screening the sequences that can self-assemble were predicted. These results were then confirmed by experimental evidence, through the synthesis of the best candidates, that seemed more likely to form hydrogels at neutral pH. This work focused on the propensity to assemble of different amino acids, as a function of their position in the peptide sequence (closer to the N-terminus or C-terminus). Some indications on the selection of building blocks can be deducted.

It is also possible to exploit MD simulations for the screening of experimental conditions and select the optimal pH or salts concentration for example. The goal in this case is to identify those experimental conditions that favor self-assembly of a specific molecule. The second example is the work of Fu et al. [25], that allowed the study of 800 peptide amphiphiles (PAs). The Phase diagrams represented in Figure 10 were constructed, showing different nanostructures formed as a function of temperature and hydrophobicity.

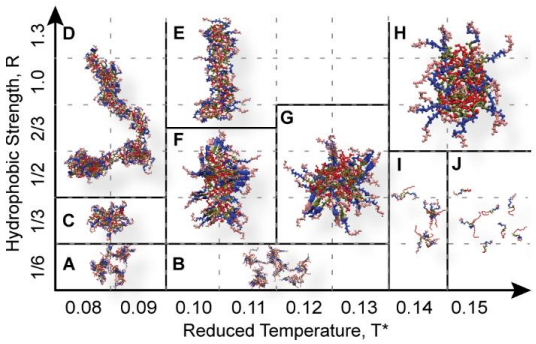

Figure 10 phase diagrams, Fu et al. 2014

1.9 COMBINATION OF EXPERIMENTAL AND COMPUTATIONAL APPROACH

Recent advancements in Force Fields and sampling methods have led to improvements in speed and accuracy of information that can be obtained from simulations. Likewise, also experimental techniques have seen improvements in relating results with computational outputs. Experimental methods are going in the direction of smaller spatial and temporal scales to meet molecular simulations, going in the opposite direction. The more the two approaches allow the investigation of the same molecular process, the higher the accuracy in drawing conclusions. [6]

1.10 FIELD OF APPLICATIONS OF SELF-ASSEMBLING PEPTIDES

1.10.1 Short self-assembling peptides

Among self-assembling peptides, the most prevalent type is constituted by short self-assembling peptides, which are composed of less than 20 amino acids. [26]

The focus on short peptides is due to the fact that they hold a great potential of use in biomedicine and nanotechnology, like antimicrobial agents, vectors for controlled drug delivery, bioelectronics, materials for cell cultures. [26]

Short self-assembling peptides can generate a variety of nanostructures. The field is attracting growing interest, because of the great versatility of the resulting hierarchical architectures that present excellent biocompatibility and good functionality. [26]

Amino acids are building blocks presenting many physical properties, such as polarity, hydrophobicity, charge. Having short peptides with less than 20 amino acids reduces the complexity both in peptide design and synthesis. Therefore, it is easier to establish quantitative structure-function relationships. Short peptides can be linear, cyclic or branched. The different amino acids lateral chains also introduce variability, multiple types of interactions can take place even in short peptides, but they are easier to handle in terms of folding. One of the main driving forces of short peptides assembly in water is the hydrophobic effect, to minimize the structure free energy. [26]

There are 20 amino acids in eukaryotes, so the possible combinations of tripeptides are about $20^3 \times 2 = 16\,000$, where the factor 2 is given by the directionality of the peptide bond. The number is slightly overestimated because of the presence of symmetrical sequences (for example the sequence IKI is counted twice). The topology of short peptides can be modified, so it is possible to have a linear chain or a circular ring in case of cyclic peptides, in which the first amino acid is linked to the last one. Another possibility is a branched peptide, in case of peptides having more than one active group. Design rules are mainly focused on analysis of amino acid substitution, sequence variations and chirality. [26]

The peptide sequence can form α-helix or β-sheets secondary structures, that interact through noncovalent interactions and generate supramolecular structures. Different nanostructure result from linear, cyclic or branched peptides. Linear peptides often assemble into nanofibers, nanoribbons, nanotubes, nanovesicles. Cyclic peptides generally form nanotubes. Branched peptides normally organize themselves in structures such as micelles or vesicles. [26]

These first nanostructures give rise to higher hierarchical structures, to obtain final length scales ranging from a few nanometers to hundreds of micrometers. Many external environmental factors can affect the nanostructure, including pH, temperature, ionic strength, solvent polarity, UV or visible light radiations, metallic ions and enzymatic reactions. [26]

The different nanostructures are more useful in different applications. In particular, nanofibrous architectures are broadly exploited in cell cultures and tissue engineering, nanotubular structures find application in artificial membrane channels and biosensors, whereas small micelles can serve as effective gene or drug carriers. [26]

1.10.2 Fields of application of SAPs: regenerative medicine and tissue engineering

Self-assembling peptides are beginning to find applications in tissue engineering and regenerative medicine, because they meet one of the main challenges in this field: finding 3D scaffold that mimic the extracellular matrix structure. Molecular self-assembly is a bottom-up technique to produce stable structures, ordered at the nano or micro scale and able to interact favorably with cells. [27]

In particular, the most investigated type of scaffold is the one generated by nanofibrous networks, an example of which is represented in Figure 11. Self-assembling peptides with a tendency to form nanofibers are privileged as an alternative to traditional natural or synthetic biomaterials currently used in regenerative medicine. These new scaffolds can be specifically designed to mimic both the structure and the function of the extracellular matrix. They can give mechanical support to the organs and provide stimuli to cells though a precise nanoscale topology. Moreover, it is possible to introduce many different bioactive motifs that would be naturally present in proteins in order to reproduce chemical stimuli. Hydrogels made of functionalized peptides can be used as scaffold for the regeneration of damaged tissues and allow cellular viability, proliferation and differentiation. The most used classes are α-elical, β-sheet peptides and peptide amphiphiles. [27]

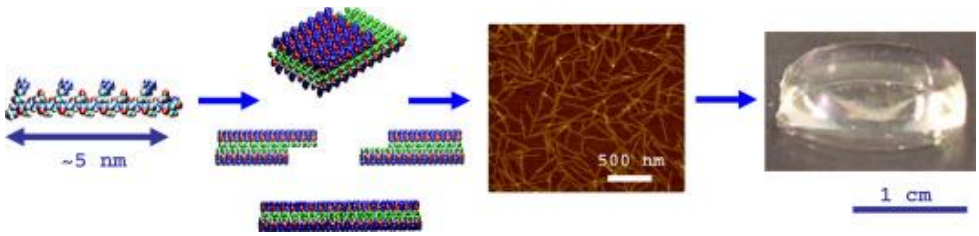

Figure 11 self-assembling peptides in nanofibrous architecture and hydrogel formation, Koutsopoulos S., MIT 2018

Hydrogels of nanofibrous self-assembling peptides are exploited in tissue engineering thanks to their capacity of biomimicry of the ECM and the possibility of different functionalization, two important advantages. They can promote biological activity in terms of protein adhesion [28], nerve regeneration [29, 30], osteoblasts proliferation, differentiation and migration [31] among others.

The focus is on applications in angiogenesis, bone regeneration, cartilage and nerve regeneration. Nowadays, most studies are at the level of animal model (mainly rat or mouse) and for cartilage the advancement is still at the level of in vitro research. The most used peptide is RADA. [32]

Some disadvantages, still limiting the clinical application, are poor mechanical properties and low reproducibility because it is hard to standardize nanofibers morphology and dimension. Other critical aspects are long-term stability and sterilization methods. [33]

1.10.2.1 Tissue engineering: β-sheets

As far as β-sheets are concerned, an example of commercialized peptide is RADA16-I (under the name of PuraMatrix), represented in Figure 12.

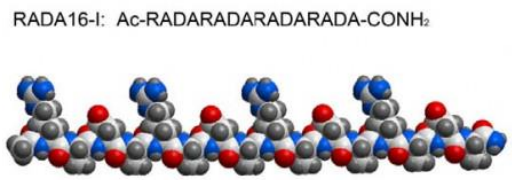

Figure 12 peptide RADA16-1, Tang et al. 2009

It has found many applications thanks to its excellent properties of biocompatibility and the enhancement of cell proliferation and tissue regeneration. Its structure resembles the extracellular matrix, mimicking the natural cell environment. [27] For instance, PuraMatrix peptide hydrogel was investigated by McGrath et al. [34] as a scaffold for peripheral nerve regeneration, showing good axonal regeneration and migration of Schwann cells in the membrane conduit.

1.10.2.2 Tissue engineering: peptide amphiphiles

Peptide amphiphiles (PAs) are a class of molecules made of peptide sequences linked to a hydrophobic alkyl chain. Various peptide amphiphiles were studied by Stupp et al. [35] The chemical structure of peptide amphiphiles is made of four regions: a long alkyl chain, hydrophilic peptide sequences with a β-sheet structure, charged amino acids as a linker and an active epitope.

The process of self-assembly of PAs in aqueous solvent is driven by hydrophobic interactions of alkyl chains and hydrogen bonding between β-sheets, similarly to the process of protein folding. They generate various type of structure (nanofibers, nanotubes, nanovesicles, micelles and ribbons), being a subclass of self-assembling peptides. [27]

Among peptide amphiphiles, it is possible to distinguish the class of aromatic peptides amphiphiles. They contain a short peptide sequence and they usually have a capping of fluorenyl-9-methoxycarbonyl (Fmoc) at the N terminus. For example, Fmoc-dipeptide Fmoc-FF (where F is phenylalanine) can spontaneously assemble in fibrous hydrogels in physiological conditions. The pi-stacking of aromatic fluorenyl rings is a particularly relevant driving force in assembly of short peptides. [27]

1.10.3 Fields of application of SAPs: drug and gene delivery

Self-assembling peptides can be employed in the realization of platforms for smart delivery of drugs or genes. Various bioactive agents, including drugs, proteins, DNA or RNA can be charged in nanostructures formed by self-assembling peptides and transported to the desired site. Once they reach a specific target, they must disassemble to release the molecules, so it is necessary to have a triggering mechanism that allows the release. There are different possible mechanisms for peptides to respond to external biological changes, such as enzymatic cleavage.

In a recent study of He et al. [36, 37], a self-assembling peptide was conjugated to a specific tag protein, with the function of incorporating a drug into the micellar structures composed by the peptide. The sequence in this case is Nap-FFK(ᵉG-FLAG)Y, where Nap stands for 2-acetylnaphthyl group, while the FLAG-tag corresponds to the peptide sequence DYKDDDK, that is linked to a ε-glycine residue. The formula and the mechanism of action are reported in Figure 13. Once the micelles get inside mitochondria, they are triggered by an amino acids sequence (FLAG-tag) to change their shape and turn into nanofibers, also thanks to specific enzymatic reactions, so that the drug gets released inside the mitochondria. In particular, the mechanism involves the enzymatic cleavage of the FLAG-tag by the enterokinases.

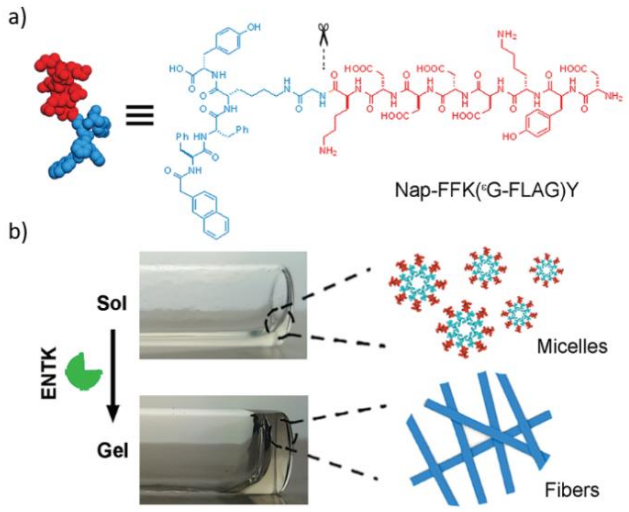

Figure 13 self-assembling peptide with flag-tag (a) and delivery mechanism (b), He et al. 2018

Another study by Cao et al. [38] investigated the peptide Nap-FFGPLGLARKRK, which forms long nanofibers able to incorporate anticancer drugs in its hydrophobic core. Even in this case, the release is allowed by an enzymatic cleavage reaction. Some enzymes that are present at the tumor site activate the fragment subjected to cleavage (GPLGLA) and cause the fiber disruption in the desired place.

1.10.4 Fields of application of SAPs: antimicrobial agents

Antimicrobial peptides work through a rapid mechanism of bacterial membrane disruption. They are often in the form of nanofibrous architectures. Antimicrobial peptides can be natural or they can be the product of *de novo* design of self-assembling peptides. The latter have the advantage of easier synthesis and scale-up, the sequences can be optimized and designed specifically to be selective for a certain type of pathogenic agent. Self-assembling peptides with antimicrobial properties generally do not have unexpected side effects and present low toxicity for the host, given their selectivity. One of the main advantages is the extremely rapid mechanism of action compared to traditional antibiotics treatments, which take longer to function. Antimicrobial peptides action first involves electrostatic interactions, the hydrophobic interactions and finally membrane rupture, that causes bacterial death and usually prevents bacterial resistance. [26]

1.10.5 Fields of application of SAPs: nanosensors and bioelectronics

Bioelectronics is another emerging field of self-assembling peptides. Nanotubes can be used as biosensors in bone mineralization, as probes for imaging and in bioelectronics devices.

Research has been carried out on peptide nanotubes to control the process of mineralization in bone regeneration. For instance, peptide amphiphiles can be conjugated with sequences that naturally have a capacity of biomineralization, such as the RGD sequence (Arginine-Glycine-Aspartate). This allows to nucleate metallic materials or semiconductor materials on the assembled nanotubes surface. [6] It was shown that this RGD sequence also has a role in integrin-mediated cell adhesion. [35]

The work of Mitchison et al. [35] used this sequence in the design of peptide amphiphiles, that assemble in nanofibers in this case, and have the capacity of nucleating hydroxyapatite (HA) crystals. They exploited acidic groups, which aid this process of HA crystals formation (anionic groups are prone to accumulating inorganic cations, leading to nucleation of crystals) and phosphorylated groups, which are important for calcium phosphate mineralization. A precise peptide design seems to favor the process of bone mineralization and investigating the key features of peptides, to identify the design principles, can extend the use of scaffold in this application. Reproducing the chemical composition and structure of natural bone HA is the main challenge, together with mechanical performance.

As far as diagnostic imaging is concerned, these nanofibrous bioactive scaffolds can also be used in magnetic resonance imaging (MRI). [35] Moreover, nanomaterial peptides can be tagged with radiometals for in vivo detection and tracking of specific targets. This techniques generally use fluorescence in the range of infrared or MRI. [39]

Short self-assembling peptides are a potential semiconductor, especially in aromatic dipeptides assembly containing phenylalanine or tryptophan, in which pi-stacking interactions play a role in semi-conductivity. In bioelectronics, peptide design research is aimed at understanding whether the electricity in peptides is conducted by transferred electrons or by protons (H^+ associated to acidic groups). This aspect still remains unclear and it is object of investigation, to deepen the understanding of peptides conductivity. [26]

1.11 PATENTS AND COMMERCIAL PRODUCTS

The relevance of self-assembling peptides also emerges form a patent research on the site "Espacenet", which provides more than 12 000 results.

Figure 14 reports an example of patent: a peptide hydrogel for insulin delivery which regulates its disassembly based on the blood glucose concentration. By glucose binding it changes from hydrophobic to charged and dissociates.

An example of a commercial product is reported in Figure 15, which represents the haemostatic material "Purastat" in a preformed syringe.

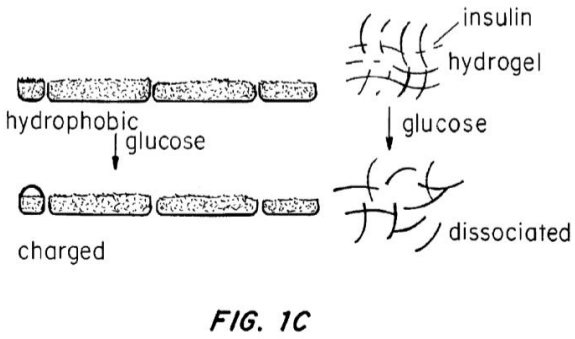

Figure 14 Self-regulated peptide hydrogel for insulin delivery: the delivery is triggered by glucose concentration in the blood (MIT patent, 2013)

Figure 15 haemostatic material in a preformed syringe, (https://3dmatrix.com/products/purastat/)

For each application field previously described, Table 1 reports a few significant examples of patents, that give a general view of the possible practical use of self-assembling peptides. (i)

PATENT	DESCRIPTION
Bone formation promoter *(2012)* KUMADA YOSHIYUKI [US]; ZHANG SHUGUANG [US]	Self-assembling peptide used for bone regeneration which forms β-sheet structure in aqueous solutions at neutral pH.
Modified self-assembling peptides *(2008)* GELAIN FABRIZIO [IT]; HORII AKIHIRO [US]; MASSACHUSETTS INST TECHNOLOGY [US]; WANG XIUMEI [US]; ZHANG SHUGUANG [US]	Self-assembling bone filler with two domains: a domain with assembling hydrophilic amino acids and a domain with hydrophobic amino acids that do not assemble.
Cerebrospinal fluid leakage occlusion *(2016)* 3-D MATRIX LTD [JP]	Peptide forming a hydrogel in physiological conditions and occludes the leakage.
Self-assembling nanoparticle drug delivery system *(2006)* CHIMEROS INC [US]	System including self-assembling peptides, proteins, nucleic acids or synthetic drugs.
self-regulated peptide hydrogel for insulin delivery *(2013)* MASSACHUSETTS INST TECHNOLOGY [US]	Hydrogel of peptide amphiphiles which links glucose and releases insulin in response of an increased glucose concentration.
Self-assembly antibacterial peptide *(2016)* CHINA UNIV OF PETROLEUM (EAST CHINA)	Cationic amphiphilic oligopeptide with antibacterial properties, selective for Gram+ bacteria.
Injectable self-assembling antibacterial peptide hydrogels *(2020)* NEW JERSEY INST TECHNOLOGY [US]	Set of amphiphilic cationic self-assembling peptides which use the high charge density to disrupt bacterial membranes (Pseudomonas biofilms).

Table 1 self-assembling peptides patents

In addition to the ongoing clinical trials and chemical research, there are also several products already available on the marked which are based on self-assembling peptides. The most famous example is PuraMatrix, commercialized by the company 3D Matrix [(ii)] Technology, which is leading clinical studies on self-assembling peptides, especially as bone

fillers, both in USA and in Europe. For instance, PuraMatrix found a clinical application in 2014 to limit peritoneal effusion in patients after pelvic surgery.

Dealing with industrial production, self-assembling peptides up-scaling costs must be considered. Solid phase synthesis is one of the main techniques that provide small scale production. New production technologies are gradually required as SAPs become more clinically relevant and a realistic treatment alternative for patients. [32]

Designer self-assembling peptides hydrogels are the most exploited type of self-assembling peptides on the market and they are particularly relevant. They present several advantages, including good biocompatibility, biodegradability in non-toxic products, predictable interactions with cellular/tissue environment. Furthermore, they are very versatile and adaptable and show high bioavailability, meaning they can exert their function without the risk of immediate degradation in the organism.

Self-assembling peptide hydrogels are particularly well suited for in vitro studies to gain insights into the interactions between cells and the extracellular matrix, or to observe stem cells behavior and cell fate, when cultured in 3D scaffolds. Self-assembling peptides can also serve as scaffold that carry out specific biological functions, such as providing chemical stimuli for cell proliferation, mimicking the extracellular matrix natural environmental conditions. Mechanical and physical properties might also be tuned to introduce topographical stimuli. [40]

Since the Zhang group in 2001 commercialized PuraMatrix – utilizable in cell cultures, regenerative medicine and tissue engineering – various other products were released on the market: PuraMatrix (Corning), PGmatrix (PepGel LLC), HydroMatrix (Sigma), Biogelx, Purastat and Curodont are some examples. [(iii, iv, v)]

Corning PuraMatrix Peptide Hydrogel [(vi)] is used to reproduce 3D microenviroments that mimic the extra-cellular matrix thanks to their nanofibrous structure created by self-assembling peptides. The composition has a high amount of water (99%) and standard amino acids which assemble in physiological conditions. The resulting hydrogel enhances cellular adhesion, proliferation of stem cells, differentiation of some cell types (such as endothelial cells, hepatocyte progenitor cells). It is useful also in studies with cancer cells to study tumor cell migration and invasion, and to study *in vivo* tissue regeneration.

Biogelx [vii] is a technology based on short self-assembling peptides. The system includes two peptides: a "gelator" hydrophobic peptide, and a "surfactant" hydrophilic peptide. These are Fmoc-diphenylalanine and Fmoc-serine respectively. In water, they assemble into nanofibers, which – in the presence of calcium ions – crosslink to give rise to a nanofibrous network. The resulting peptide hydrogel, with 95% of water, is similar to the extracellular matrix. The nanofibers surface is hydrophilic and it allows appropriate cellular adhesion. It is also possible to add functional motifs, like fibronectin, collagen, laminin, to further mimic the natural ECM composition and function.

The matrix PuraStat [viii] is formed by three synthetic peptides, arranged in a sequence that is repeated four times: RADA-16-I. These peptides assemble as a consequence of a pH change, so the exposition to ionic solution can trigger the process of assembly. This product is a hemostatic material, in the form of a pre-formed syringe. It is used in case of small vessels bleeding in the gastrointestinal tract and for vascular anastomosis.

Despite the many different biomedical applications, researchers still have difficulty in predicting higher order hierarchical structures, their properties and function, starting from the individual constituent building blocks.

With a multidisciplinary approach and the integration of computational and experimental studies it is possible to investigate chemical-biological aspects in their complexity, with the aim of transferring the obtained knowledge to clinical and industrial application of the realized products.

1.12 AIM OF THE WORK

Self-assembling peptides can assemble into higher order hierarchical structures and can form different possible architectures at the nanoscale: nanotubes, nanofibers, nanoribbons, micelles, layers. It is not possible to determine the structural features only based on the peptide sequence. Researchers have experimentally observed the formation of fibrils from some self-assembling peptides through imaging techniques, but the precise molecular configurations are still undetermined. They are often hard to obtain due to experimental limitations in techniques of structural characterization, such as X-ray crystallography. The crystallographic structure of a peptide might not always be available, but structural

knowledge would be important to understand the principles of self-assembly and to deduct design principles for peptides.

The aim of this work is to investigate, from a computational point of view, seven different peptide sequences of self-assembling peptides. Molecular dynamics simulations are used to assess the behavior of peptide models in time. Starting from pre-ordered configurations, that are hypothesized to be amyloid-like, different possible structures for each peptide are evaluated. Simulations are examined from a visual perspective and through quantitative analysis, to establish which structures are more stable and more likely to be the molecular configurations underlying experimentally observed fibers.

The reason why simulations start from pre-ordered structures and not randomly organized peptides is because otherwise it is hard to observe self-assembly in the generally used simulation time, both because of computation resources and of time limitations. All configurations present peptides in β-sheet secondary structure. Single peptide chains organize in β-stands and β-sheets, giving rise to different possible configurations which differ in the combination of strands and sheets disposition in space. Different configurations consequently cause different intermolecular interactions and might have a different stability and propensity for assembly. Computational modelling is the strategy used to assess the best configurations and deepen the understanding of self-assembling peptides at the nanoscale level.

2 MATERIAL AND METHODS

2.1 COMPUTATIONAL METHOD: MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics consists in the calculation of an atomic system motion using molecular mechanics, including interactions and bonds among atoms. It is possible to assess the time-dependent behavior of a system and their stability and intermolecular interactions.

The basic algorithm of molecular dynamics is schematized in Figure 16. The force field allows the potential energy calculation. The force is defined as the potential gradient with a sign change. Using Newton's second law it is possible to obtain the acceleration. Integration in time leads to velocity calculation and a second integration to the position. Reiterating the process for each atom *i*, spatial coordinates of each atom in the system can be calculated. The system coordinates and velocities at several time points allow the reconstruction of atoms trajectories and the final output is a sequence of simulation frames. [41]

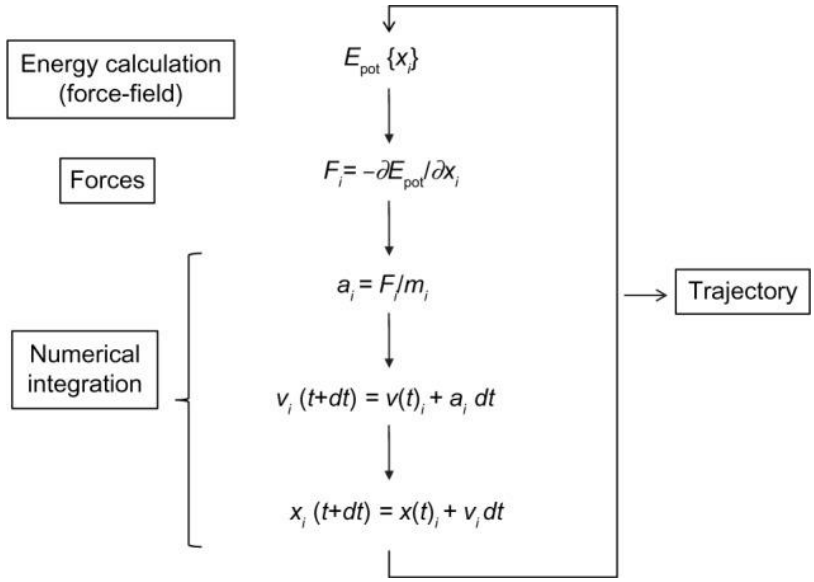

Figure 16 basic algorithm of molecular dynamics, Hospital 2015

Molecular dynamics is the step-by-step solution of equations of motion and the timestep is the parameter that defines the time span between two consecutive passages, which determines the frequency of calculations. A timestep of the order of fs is generally used, since it has to be inferior to the smallest oscillation period of the system atoms (caused by thermic vibrational motions).

The algorithm starts from the calculation of the potential energy U of the system. The mathematical expression of potential energy is referred to as Force Field. The total potential energy is given by different contributions: covalent bonds (bond stretching, bending, rotation), van der Waals forces, electrostatic interactions. In particular, the expressions of the energy contributions are the following:

- Covalent bonds energy between atoms pairs:

$$U(l) = \sum_{bonds} \frac{1}{2} k(l - l_0)^2$$

The two atoms are modelled as masses joined by a spring, so the energy corresponds to the elastic energy: l and l_0 are the length and the resting length respectively, corresponding to the distance between atoms, and k is the elastic constant of the spring.

- Angle energy due to interactions of three atoms covalently bonded:

$$U(\theta) = \sum_{angles} \frac{1}{2} k(\theta - \theta_0)^2$$

Where θ and θ_0 are the angles formed by three atoms in the considered time instant and at rest, and k is once again the elastic constant.

- Dihedral angle energy (or torsional energy) due to the interaction of four atoms covalently bonded:

$$U(\varphi) = \sum_{dihedrals} \frac{1}{2} k\,(1 - \cos\varphi)$$

Where φ is the torsional angle and k is the constant that reflects the torsional capacity.

- Van der Waals energy:

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

Where ϵ is the minimum interaction energy and σ is the minimum atomic distance at which the interaction potential results equal to zero, r is the two atoms distance.

- Electrostatic interactions energy:

$$U(r) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}$$

Where q_i and q_j are the charges of the two atoms, ϵ_o and ϵ_r are the dielectric constants respectively of the vacuum and the mean and r_{ij} is the distance among atoms. [42]

Molecular dynamics in this specific work is used to investigate seven peptide sequences with various amino acids substitutions. For each peptide ten different possible configurations are examined. The objective is to evaluate the stability of the configurations with the aim to determine the structures that are more likely to be stable, and the aim to make a comparison among peptides and understand the effects of sequence variations.

2.2 SOFTWARE AND HARDWARE

The software used in this work are NAMD for molecular dynamics simulations and VMD as a graphical interface and analysis tool. [ix], [x]

NAMD (NAnoscale Molecular Dynamics) is a software used for molecular modelling, developed for macromolecules simulations. NAMD allows to overcome one of the main obstacles of working with large molecular aggregates, which is the computational power required, by giving the possibility to use more processors in parallel. To launch a simulation, NAMD requires information contained in four main file types:

1. pdb file (Protein Data Bank, an archive of biological molecules structures experimentally characterized), containing atomic coordinates and velocities.
2. psf file (Protein Structure File), containing atoms with the relative charges and masses and information on the protein structure (types of bonds and interactions among atoms).
3. Force Field parameters file, containing the parameters necessary for the mathematical expression of the potential energy of the system. The Force Fields necessitate the use of a topology file, defining the mass, type and charge of atoms in an amino acid residue and a parameter file, including all the constants to assess forces, energies and interactions involving the atoms combinations described in the topology.
4. Configuration file, in which the user specifies all the options that NAMD has to adopt while performing the simulation (for instance duration, temperature, cut-off distance).

The output files that are returned form the program are a trajectory file with the coordinates in time (in dcd format) and a file containing the evolution of energy (bonded and non-bonded atoms energy contributions).

VMD (Visual Molecular Dynamics) is a molecular visualization program for the representation and analysis of biological systems in a 3D graphics. It can be exploited to study the systems trajectories of molecular dynamics simulations and it constitutes a graphical interface that is compatible with many programs, including NAMD. VMD can read standard pdb, psf and dcd files. This software is suitable for large molecular systems thanks to the optimized memory, the extended syntax to organize and select atoms groups and the advanced rendering. Other than the graphical interface, VMD is also provided with a text-based interface built on scripting languages such as Tcl/Tk. The user can therefore generate scripts to perform tasks automatically.

As far as hardware is concerned, simulations are performed on CINECA supercomputer MARCONI100. High Performance Computing allows to speed up the calculations involved in molecular dynamics simulations and to obtain the relative results much faster. The number of GPUs and processors is optimized through a process of trial and error, and the maximum speed is found to be around 50 ns/days, which is an acceptable time and allows simulation of 150 ns to be performed in three days.

Computational modelling is a tool to understand complex supramolecular interactions at the nanoscale level. In this case it allows to determine the structure of molecules at a resolution that overcomes experimental limitations, and to assess the first steps of self-assembly of different peptide sequences within time spans of nanoseconds (starting from organized peptides in solution).

2.3 PEPTIDE SEQUENCES

In the following pages, peptides are numbered according to Table 2 and correspond to their respectively reported amino acid sequence. For each peptide sequence, ten different possible configurations are investigated. The first peptide has a tyrosine residue, the second and third peptide have a modification of phenylalanine in different positions in the sequence. The fourth peptide has a phenylalanine residue instead of tyrosine, whereas the last sequences are characterized by phenylalanine variants (D-phenylalanine, homophenylalanine and β-phenylalanine).

Peptide number	Peptide sequence
Peptide number 1	Ac-YFQQQFK-conh2
Peptide number 2	Ac-YF(I)QQQFK-conh2
Peptide number 3	Ac-YFQQQF(I)K-conh2
Peptide number 4	Ac-FFQQQFK-conh2
Peptide number 5	Ac-(D-F)FQQQFK-conh2
Peptide number 6	Ac-(hF)FQQQFK-conh2
Peptide number 7	Ac-(β-F)FQQQFK-conh2

Table 2 peptide identification code (number) and corresponding amino acid sequence

2.3.1 Amino acids features

Tyrosine is reported in Figure 17a. It is a hydrophobic amino acid, even though it presents a -OH group in the aromatic ring, which is susceptible to the formation of hydrogen bonds. It is a natural occurring amino acid in the L-tyrosine form and it is synthetized from L-phenylalanine in vivo.

Halogenated phenylalanine is represented in Figure 17b, compared to regular phenylalanine in Figure 17c. The aromatic benzene ring has a substitution of a hydrogen atom with one iodine atom, so it is called phenyl-I or iodinated phenylalanine. The halogen atom can form a halogen bond with another electronegative atom thanks to positively charged region and it can further stabilize intermolecular interactions between peptides. Halogens are often introduced as modifications to make the fibrils more tunable.

D-phenylalanine is reported in Figure 17d, compared to its enantiomer L-phenylalanine (N-terminal residue in this case) in Figure 17e. It is possible to note that they are stereoisomers, specular images not superimposable with any rigid motion.

Homophenylalanine is a homologue of phenylalanine. Its chemical 3D structure is represented in Figure 17f and it differs from phenylalanine because the aromatic ring is bonded to a different additive carbon atom (that in turns forms bonds with two hydrogen atoms and a carbon).

β-phenylalanine differs from phenylalanine because it has the amino group and the carboxylic group bonded to two adjacent carbon atoms, so the amino group is bonded to the carbon C_β instead of C_α as in phenylalanine. The chemical 3D structure is reported in Figure 17g.

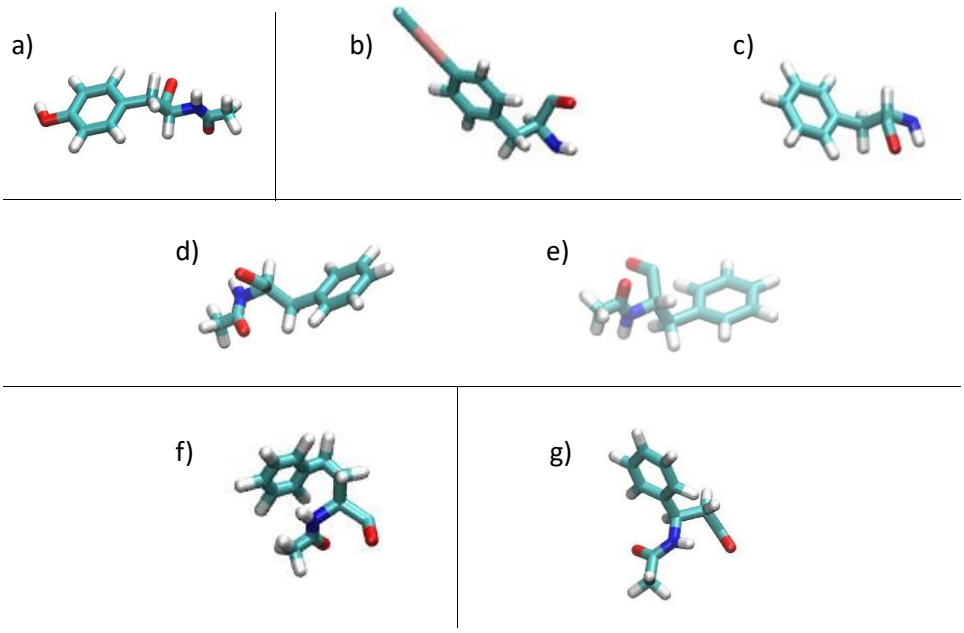

Figure 17 amino acids 3D structure (VMD images) of: tyrosine (a), N-terminal phelyl-I (b) compared to N-terminal phenylalanine (c), D-phenylalanine (d) compared to its enantiomer L-phenylalanine (e), homophenylalanine (f) and β-phenylalanine.

2.3.2 Parametrization of halogenated phenylalanine

The peptide sequences that require a particular attention in their construction are halogenated peptides (peptide number 2 and peptide number 3) containing a phenylalanine residue with an aromatic iodinated benzene ring (phenyl-I).

The parametrization of halogen bonds is challenging because in the CHARMM Force Field the charge of the halogen atom is represented a unitary negative charge, but halogens can form a positively charged region that is involved in the halogen bond. Therefore, they can generate electrostatic interactions. In order to represent the positive charge located opposite to the halogen covalent bond, also called σ-hole, the idea is to add a positive (massless) virtual particle to the halogen.

As far as the halogenated peptides are concerned, the parametrization of halogenated phenylalanine has been realized in the work of Gutiérrez and collaborators [17] using CGenFF (Charmm General Force Field), an extension of the CHARMM Force Field that allows modelling of small molecules that are not proteins (such as ligands).
The topology and parameters files, obtained through the server ParamChem [(xi)] starting form a structure in mol2 format built using Avogadro software, are then adapted to integrate them with CHARMM. The modifications of the halogenated phenylalanine topology include eliminating some unnecessary atoms (such as a methyl group that is added to make the nitrogen atom a N-terminal nitrogen) and adding bonds with the C terminus of the previous amino acid and with the N terminus of the following amino acid in the sequence. To compensate the error introduced by eliminating atoms, their charge is subtracted to the charge of the carbon C_α so that the total charge of the unit is equal to zero. The final topology file is included in the generation of the solvated protein structure file (psf) and the corresponding modified parameters file is included in the configuration file with simulation specifics.

For the simulations of halogenated peptides, parameters of the CGenFF are also included in the configuration file. Some additional modifications are made on this file, in particular the rows relative to atoms (for nucleic acids and carbohydrates) not involved in the simulations are commented.

A similar process has been followed in a previous thesis work for the construction of D-phenylalanine, homophenylalanine and β-phenylalanine. Parameters and topology are included in the file for generation of the psf file and in the configuration file respectively. Simulations in these cases are performed using only CHARMM Force Field.

2.4 PEPTIDES CONFIGURATIONS: AMYLOID-LIKE FIBRILS

According to a previous thesis work [43], the only way to obtain self-assembled fibrils within the simulation time (150 ns in this case) available is to start from a precise organization of peptides in space. Starting from randomly placed peptides, in fact, does not lead to the self-assembly process in hundreds of nanoseconds.

The configurations are chosen on the basis of the work of Eisenberg [44], which deals with amyloid fibrils configurations. Amyloid fibrils are protein aggregates involved in many pathological conditions, such as Alzheimer disease, Parkinson and many diseases classified as amyloidosis. Amyloid fibrils are recognized because they are characterized by a typical cross-β diffraction pattern (in X-rays experiments), formed by parallel layered β-sheets, perpendicular to the long axis of the fibril.

The β-strands within a sheet are bonded through hydrogen bonds. The β-strands can be parallel or antiparallel.

The β-strands of laterally facing β-sheets interactions were observed from the atomic structure of different amyloid fibrils. The β-sheets associate in pairs by interdigitating their lateral residues, creating a tight interface where no water molecules are interposed. The side chains organize by minimizing the steric hindrance, in a way that resembles the teeth of a zipper, which gives the name of the motif of β-sheets pairs as steric zipper. The β-sheets pairs constitute the protofilaments of the amyloid fibril, usually formed by two or three protofilaments. The bonding of β-sheets in a steric zipper can occur in different ways, giving rise to polymorphism. The resulting fibrils have a width ranging between 8 and 20 nm and different amyloid fibrils have a similar appearance to experimental observation thanks to these common characteristics.

Compared to the strands composing a β-sheet that assemble via hydrogen bonds and can easily disassemble, the pairs of β-sheets in a steric zipper are also stabilized by the van der Waals attraction, which is stronger thanks to the close proximity of interdigitating side chains.

It is possible to distinguish four features of β-sheets geometries, leading to ten symmetry classes, represented in Figure 18.

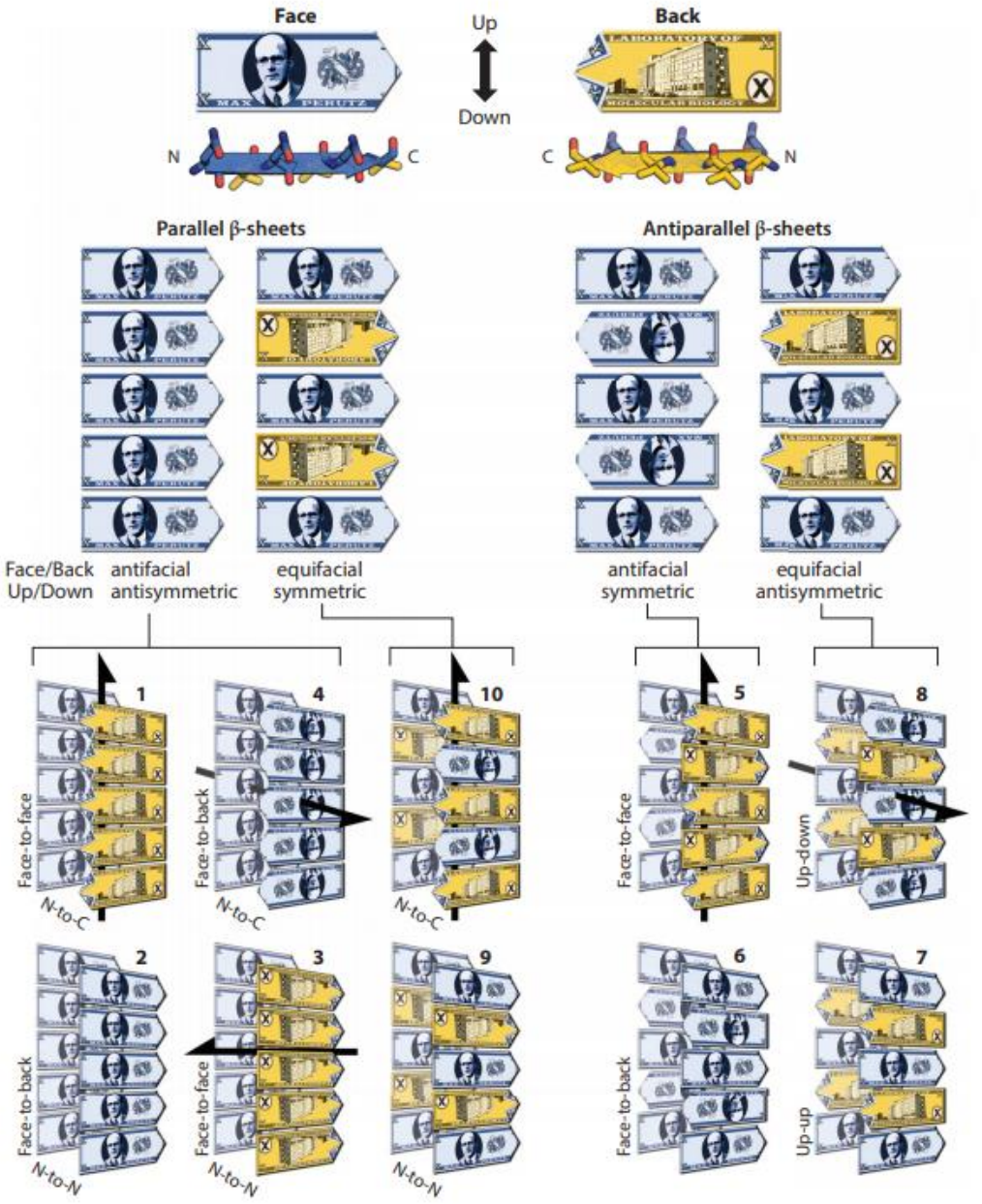

Figure 18 ten symmetry classes of amyloid fibrils: possible configurations, Eisenberg 2017

The four features characterizing a configuration, schematized in Figure 19, that give rise to the ten possible structures for each peptide, are: the β-strands orientation (parallel or antiparallel), the β-strands orientation in the normal direction with respect to the peptide strand axis (antifacial or equifacial), the β-sheets orientation (parallel or antiparallel), the β-sheets lateral packing (face-to-face or face-to-back). The resulting possible combinations of the four features, that originate ten possible configurations, are schematized in Table 3.

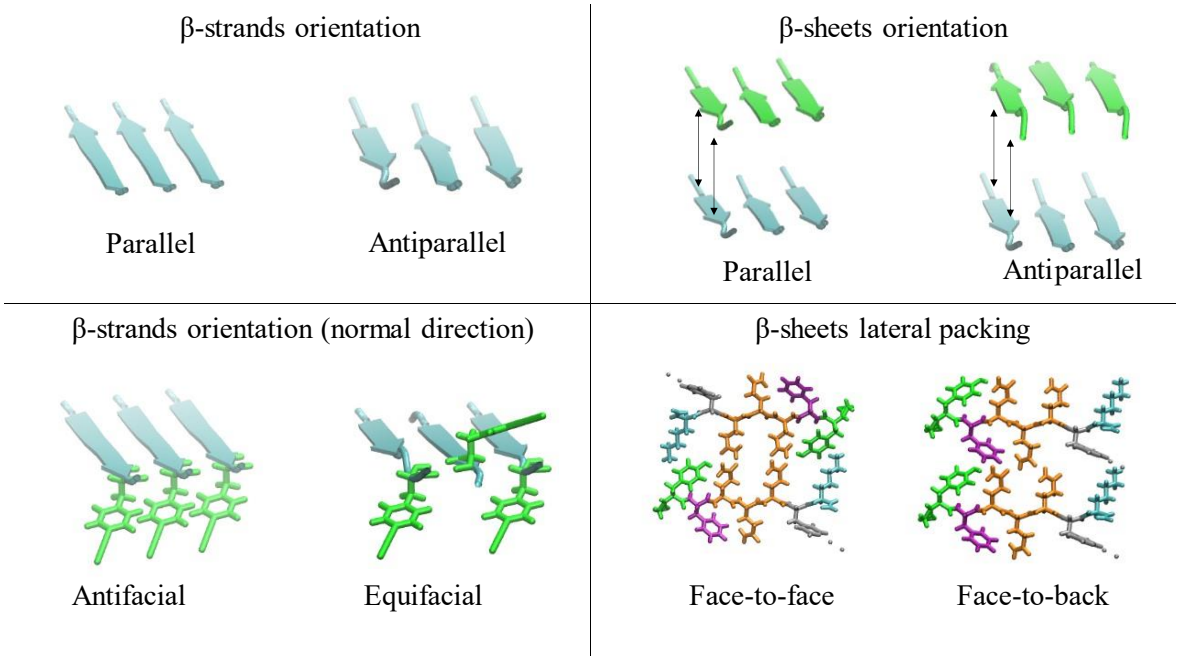

Figure 19 four features that characterize the ten possible configurations

Configuration	β-strand orientation	β-strands orientation (normal)	β-sheets orientation	β-sheets lateral packing
1	Parallel	Antifacial	Antiparallel	Face-to-face
2	Parallel	Antifacial	Parallel	Face-to-back
3	Parallel	Antifacial	Parallel	Face-to-face
4	Parallel	Antifacial	Antiparallel	Face-to-back
5	Antiparallel	Antifacial	Antiparallel	Face-to-face
6	Antiparallel	Antifacial	Parallel	Face-to-back
7	Antiparallel	Equifacial	Parallel	Face-to-back
8	Antiparallel	Equifacial	Antiparallel	Face-to-back
9	Parallel	Equifacial	Parallel	Face-to-back
10	Parallel	Equifacial	Antiparallel	Face-to-face

Table 3 ten possible amyloid configurations

Considering a single β-sheet, the composing β-strands can be parallel or antiparallel depending on the disposition of the N and C termini. Parallel β-stands have the N terminus pointing in the same direction, whereas antiparallel in the opposite direction with an alternance of polarity. The β-strands can also be equifacial or antifacial. Supposing that each

strand has a face of even numbered residues and the opposite face of odd numbered residues, antifacial strands all display the same face of even numbered residues on one side of the sheet and odd numbered on the opposite side. Equifacial strands have alternating odd numbered and even numbered faces by looking at a side of the β-sheet. The same distinction can be better understood by considering a residue that is exposed on one side only of the β-strand, and by observing if it is always exposed on the same side of the sheet (antifacial) or on alternating opposite sides (equifacial).

Considering different β-sheets, they can also be parallel or antiparallel depending on whether the strands of laterally facing β-sheets have the same polarity or opposite alternating polarity. Another feature that characterizes the spatial organization is the lateral packing of residues. Two mating β-sheets exposing the same residues towards one another are face-to-face (and each sheet therefore forms a back-to-back interface with the other β-sheets above and below), otherwise they are face-to-back. In this case there is only one type of interface.

The four features that characterize a configuration are schematized in Figure 20, which reports an example of an 8x8 model. The chosen configuration is configuration 8. In this case it is possible to observe antiparallel β-strands with alternating polarity. The bead representing tyrosine lateral residue shows that the strands are equifacial because it is exposed on opposites sides of two adjacent strands in the same sheet. The β-sheets are antiparallel because laterally facing peptides of two adjacent sheets have opposite polarity. Their lateral packing is face-to-back and this is visible once again from the disposition of tyrosine beads always on the same side of two laterally facing peptides belonging to adjacent β-sheets. Therefore, only one type of interface is present.

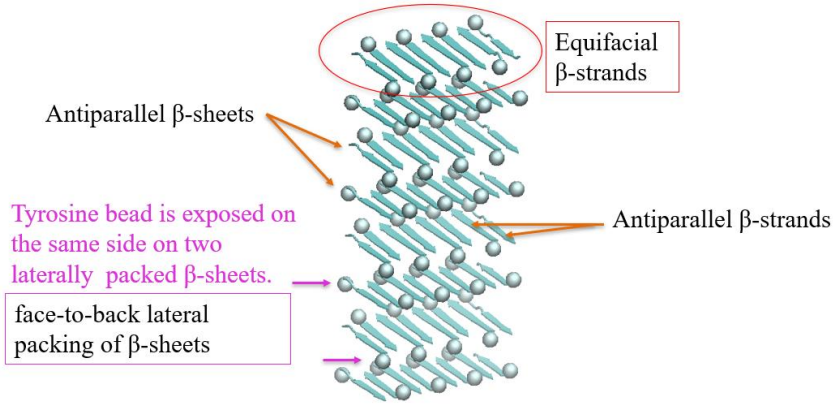

Figure 20 example of configuration 8

2.5 CONFIGURATIONS CONSTRUCTION

Starting from the peptide pdb file, a β-sheet is constructed by creating a total of eight repetitions of the peptide, spaced of a 5 Å distance. With another script the configuration is created through the replications of the obtained β-sheet, translating it by 12 Å and rotating it to create the final 8x8 model consisting of eight β-sheets. The first step allows to create parallel or antiparallel β-strands and equifacial or antifacial β-strands. The second step determines the β-sheets orientations and their lateral packing. The spacing between strands and sheets is chosen according to the work of Eisenberg. [44] The configuration construction process is schematized in Figure 21. The three projections in space of the final 8x8 model are reported as an example in Figure 22.

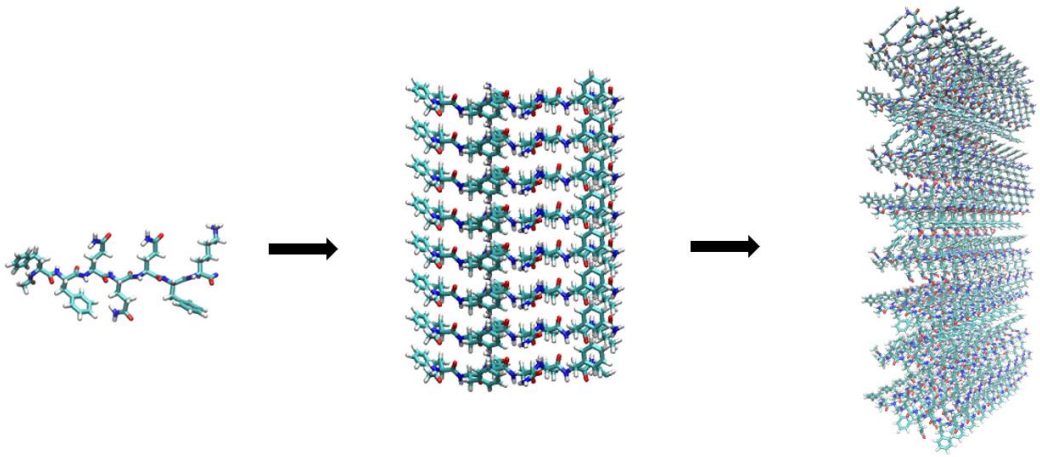

Figure 21 steps of construction of a configuration: from the peptide to the eight β-strands to eight β-sheets

Once the configurations are constructed with the necessary translations and rotations of peptides and β-sheets in space, the structure is solvated, immersed in a water box. The solvent in this case is explicit, which means that water molecules are represented as such, and not as their interactions contributes (which is the case of implicit solvent). This passage gives rise to an atomic system of about 69 000 atoms, that is acceptable in terms of computational cost. The originated system (the corresponding pdb and psf files) is the final system, which reproduces natural environmental conditions of peptides in solution, representing the starting point of minimizations.

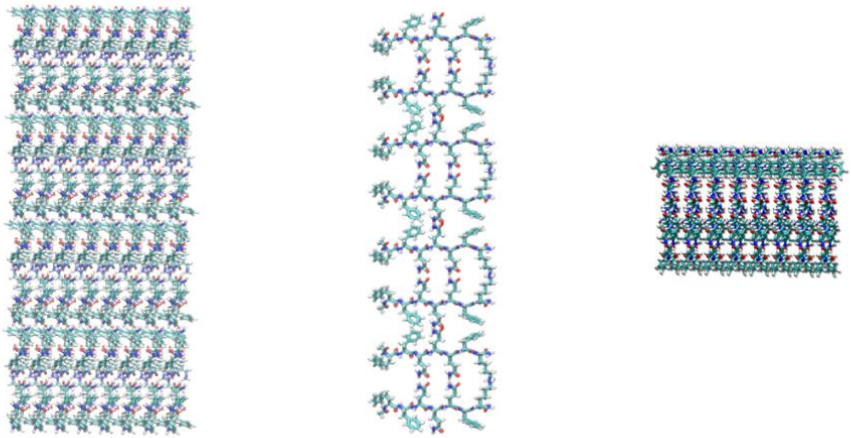

Figure 22 example of projections in the three dimensions of a configuration

2.6 SIMULATIONS PARAMETERS

All simulations are performed using the software NAMD.

The simulations are preceded by a minimization of 50 000 steps, with a timestep of 1 fs/step. This first passage of minimization is necessary for the system to reach a local minimum of potential energy, which is the starting point for the simulation.

The simulation is divided into three consecutive runs (called run01, run02, run03) of 50 ns each, to obtain a total of 150 ns, since a unique simulation of this duration (150 ns) with a timestep of 2 fs/step causes the simulation to interrupt for instability. The division in three shorter simulations with a smaller timestep of 1 fs/step also allows to minimize the computational cost by making each simulation run faster (with a speed of about 50 ns/day). The minimization output constitutes the input for the first run (run01), whose output is used as input for the second simulation (run02) and so it goes for the third. This is possible thanks to restart files, that are periodically written during the simulations, and allow the system to start the next simulation from the final conditions (atoms coordinates, velocities and energies) of the previous.

The main simulation parameters are reported in Table 4.

Duration	50 ns
Timestep	1 fs/step
Output frequency	25 000 steps
Temperature	300 K
Water box dimensions	83x136x67 Å
Minimization steps	2000
Run (number of steps)	50 000 000

Table 4 simulation parameters

The Force Field used for all simulation is CHARMM, in the version CHARMM36.
CHARMM Force Field is mostly used for biological systems (such as peptides, proteins, nucleic acids, carbohydrates) and for inorganic materials in materials design applications. Halogenated peptides also require the inclusion of CGenFF.

2.7 SIMULATIONS ANALYSIS

The simulations analysis consists in both quantitative and qualitative evaluation. As far as quantitative analysis is concerned, different indicators of stability are considered, in order to obtain a ranking from the most stable to least stable configuration. More than one configuration can be stable depending on the experimental conditions of assembly. Therefore, even if the simulations are carried out under equal conditions, a few possible configurations are identified. Another reason why it is hard to assess the absolute most stable configuration is that the quantitative analysis is a combination of different parameters and results are not always in accordance. Moreover, the qualitative observation provides further information that must be taken into account when selecting the best and worst fibril structures.

2.7.1 Quantitative analysis

Different quantitative parameters are considered. A definition and physical explanation of the stability indicators follows. Matlab scripts used to obtain the graphical outputs are reported in the APPENDIX.

Fist, Root Mean Square Deviation (RMSD) is one of the main parameters that are generally used to assess the stability of a molecular system. It is defined as:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{Natoms}(r_i(t1) - r_i(t2))}{Natoms}}$$

Where *Natoms* is the total number of atoms, whose positions at different times are compared, $r_i(t)$ is the atom *i* spatial coodinate at time time *t*.

RMSD provides an indication of how much the atomic system coordinates differ from the initial ones. It allows to evaluate whether some configurations result stable in time, which means that the RMSD graph reaches a plateau. Moreover, it is possible to identify the time after which most configurations result stable by comparing the graphical output for different configurations of the same peptide.

An example of RMSD graph for a configuration of a given peptide (configuration 1 of peptide number 1) is reported in Figure 23. At time 0 ns, the initial RMSD value is equal to zero, and increases in time. The structure becomes stable when the graph reaches a plateau, meaning that the atomic coordinates differ of a certain constant amount from the initial coordinates. It is possible to identify the time after which a configuration becomes stable, in this case after 50 ns the value is constant but shows a final increase.

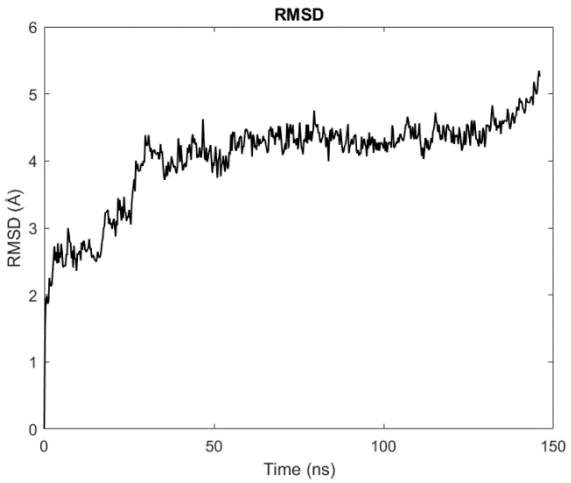

Figure 23 example of RMSD graph

Root Mean Square Fluctuations (RMSF) is another indicator of stability and gives information on which amino acid residues fluctuate more with respect to a temporal mean conformation. The higher the RMSF value, the more a residue fluctuates. Periodicity in the

resulting graph can reflect the disposition of residues in the peptide and within the structure comprising many peptides: external residues tend to be more subjected to fluctuations compared to internal residues.
In this case, the objective is to assess the overall behavior of the configuration, so it is useful to consider a mean value that includes all residues. In so doing, a numeric value for each configuration, that reflects the structure fluctuations, is returned.

Hydrogen bonds number is indicative of the stability: the higher the hydrogen bonds number, the more the structure is stabilized by this type of noncovalent interaction. The continuous formation and disruption of hydrogen bonds is reflected in the noisy graphs, that are difficult to interpret. To make them clearer and be able to distinguish a trend, the mobile mean is calculated. The optimum number of frames on which the mobile mean is calculated is found to be 100 frames, corresponding to 2,5 ns of simulation. The percentage of hydrogen bonds is considered when comparing different configurations of a peptide, to understand which structure maintains the highest percentage of hydrogen bonds. For configurations comparison, the values are expressed in percentage also for β-sheets, hydrophobic contacts and native contacts.

The β-sheets percentage trend reflects how much the original structure is preserved in time: starting from eight β-sheets, the configuration can either maintain or lose the structure. The less the value decreases, the more a structure is considered stable. There is the possibility for a configuration to decrease its β-sheets content, while still assuming another stable configuration. This can be verified with the observation of simulations.

Hydrophobic contacts are another parameter that can inform about configurations stability, which involves hydrophobic residues. The meaning is analogous to hydrogen bonds: the higher the number of hydrophobic contacts, the more stable the structure.

The Stability Index (SI) is a number that is used to classify the configurations from the most stable (with the highest value) to the most unstable (lowest value). It is defined as:

$$SI = 0.5\ x\ \frac{1}{RMSF} + 0.5\ x\ Hbonds\ (backbone)$$

It is directly proportional to the average number of hydrogen bonds and inversely proportional to the mean RMSF of residues. The higher the SI value, the more stable the

structure is. The main limitation of the SI is the fact that it mainly accounts for backbone interactions, so the result of the ranking must be associated with the other analyses.

Native contacts are particularly useful to complement information extracted from the Stability Index. Native contacts are defined as the interactions originally present in the structures at the beginning of the simulation. Any structure that maintains in time a value close to the initial one can be considered stable. Once again, native contacts can see a decrease for different reasons. The structure might get unstable in the simulation time and disassemble or it might originate another nanostructure, stable, but different from the fibril initially hypothesized.

2.7.2 Qualitative observation

The qualitative observation of simulations allows to confirm the quantitative analysis and clarify the behavior of the configurations in time. In case of a stable configuration, the fibrillar initial structure is expected to be maintained throughout the 150 ns of simulation. In case of an unstable configuration, the structure departs from the initial hypothesized configuration. Configurations classified as unstable from the quantitative analysis of this work can have different behaviors: they can either lose the β-sheets fibrillar structure and show the detachment of various peptides (so they are unstable in the strict meaning of the term) or they can assume a different shape from the initial one, that might be just as stable. In the latter scenario, for instance, a configuration that assumes a different shape than the initial fibril (and is therefore considered unstable by the analysis) can originate a stable vesicle or micelle. This is the main reason why quantitative parameters are not sufficient to explain the stability and must be combined with the visual observation in order to draw conclusions on the behaviors of configurations.

3 RESULTS AND DISCUSSION

3.1 PEPTIDE NUMBER 1 (AC-YFQQQFK-CONH2)

3.1.1 Results

For the sake of simplicity, the graphs are reported as an example only for the first peptide. As far as the other peptides are concerned, graphs and data can be found in the APPENDIX.

The first quantitative analysis performed was RMSD (Root Mean Square Deviation). In Figure 24 the RMSD values in time are represented for all the 10 configurations. After about 50 ns most configurations reach a stable plateau. It is possible to distinguish two different behaviors: configurations 2, 3 and 9 reach higher values, take longer to stabilize and do not show a stable plateau as all the other configurations. Among the other structures, 6 (light blue), 7 (pink) and 1 (black) seem to be the most stable.

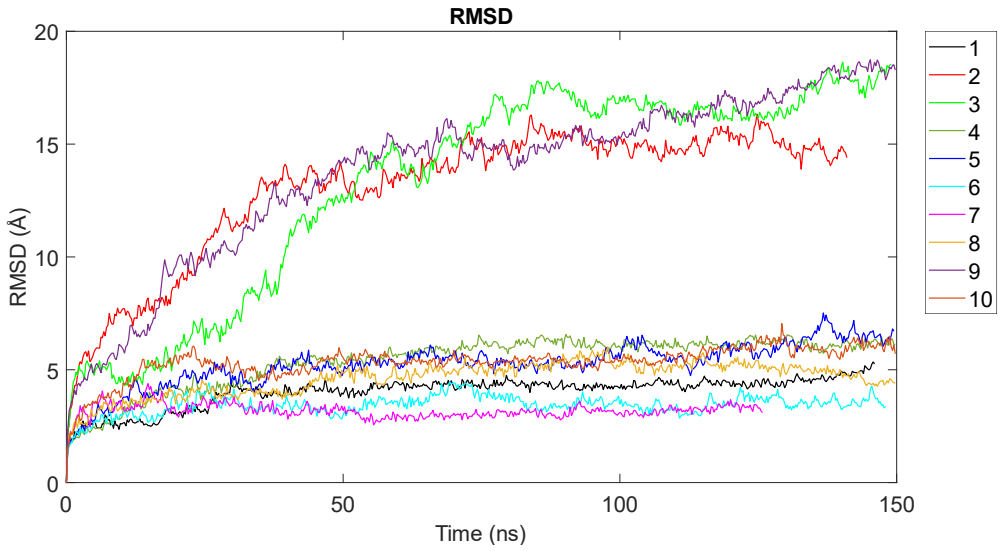

Figure 24 RMSD in time for all configurations of peptide number 1

The second analysis is RMSF (Root Mean Square Fluctuation). Figure 25 shows the RMSF as a function of amino acid residues and gives information on which amino acids fluctuate more with respect to a mean conformation. The figure only reports the best and worst cases. If the interest is in understanding how much a configuration (as a whole) fluctuates, it is preferable to consider a mean of RMSF values on residues. In Figure 26 it is shown a bar

graph of the mean RMSF for all configurations. The mean RMSF reflects the trend of the RMSD with two behaviors: configurations 9, 3 and 2 fluctuate more, whereas the others are more stable and configurations 6, 1 and 7 fluctuate less.

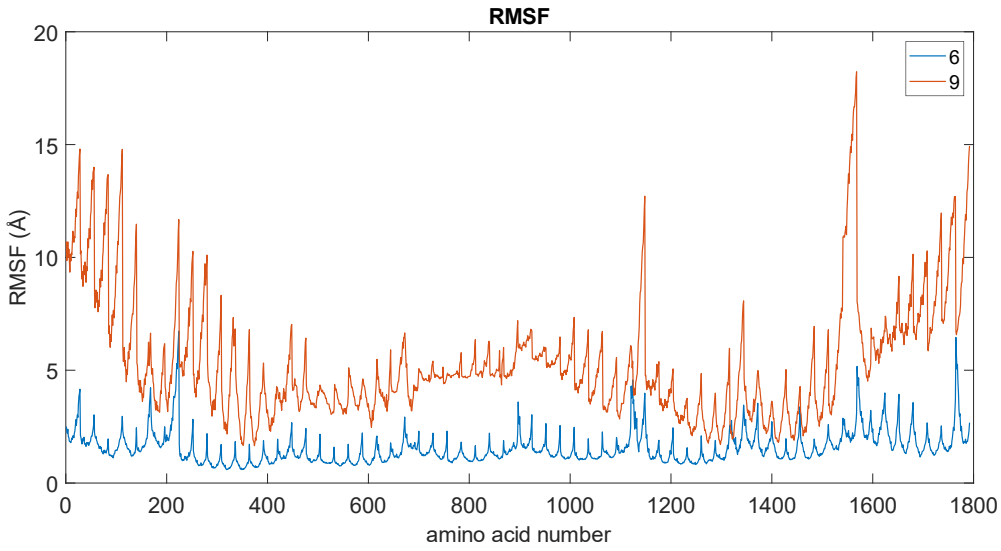

Figure 25 RMSF limit cases: maximum (configuration 9) and minimum (configuration 6)

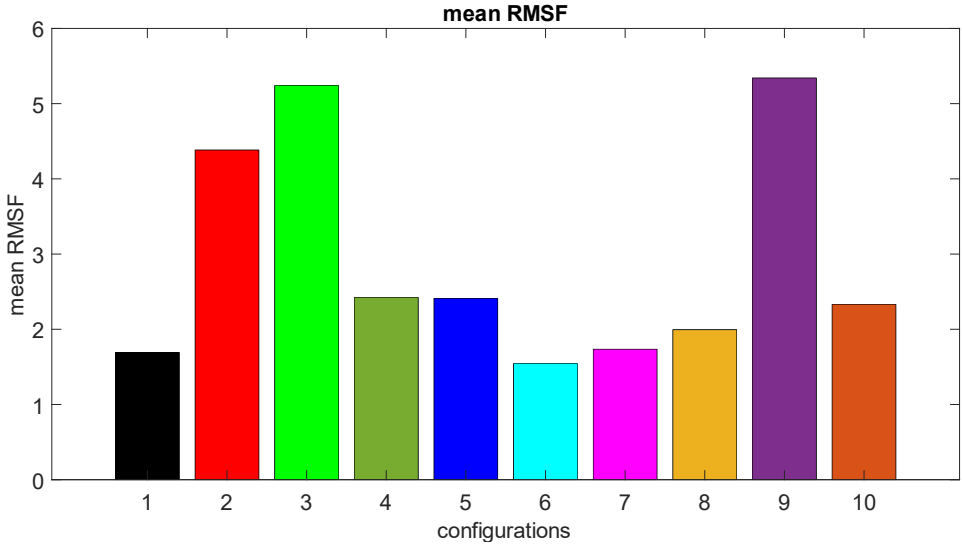

Figure 26 mean RMSF for all configurations

Figure 27 shows hydrogen bonds in time. To neglect the continue formation and disruption of hydrogen bonds, the value is a mobile mean on 100 frames, corresponding to 2,5 ns. Values are expressed in percentage. Configuration 1 maintains the highest percentage with a slight increase in time, configuration 9 is the least stable.

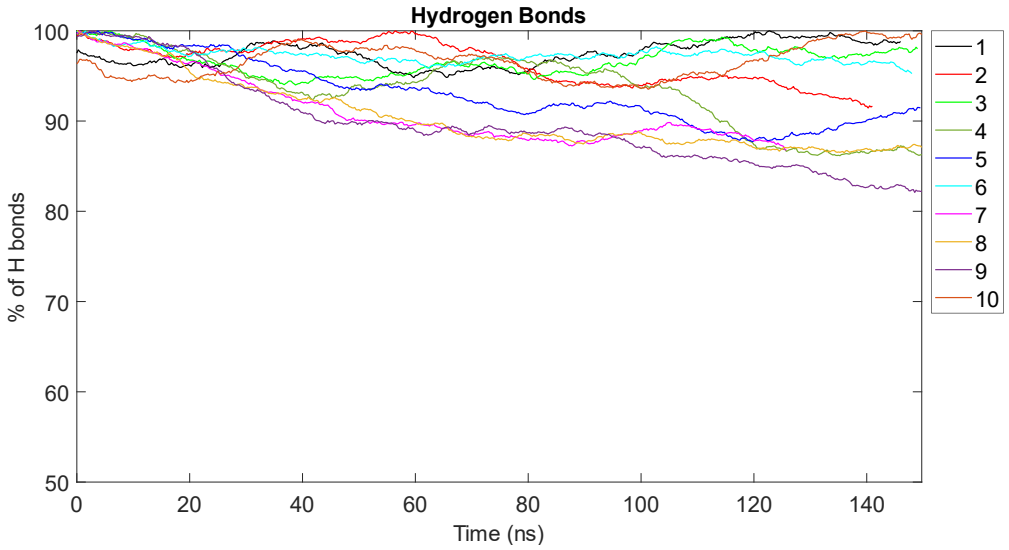

Figure 27 hydrogen bonds percentage in time for all configurations of peptide 1

The percentage of β-sheets is represented in Figure 28 as the mobile mean on 10 frames (0,25 ns). It evidences how structure 9 and 10 are unstable and structure 6 is stable. All other configurations are comparable.

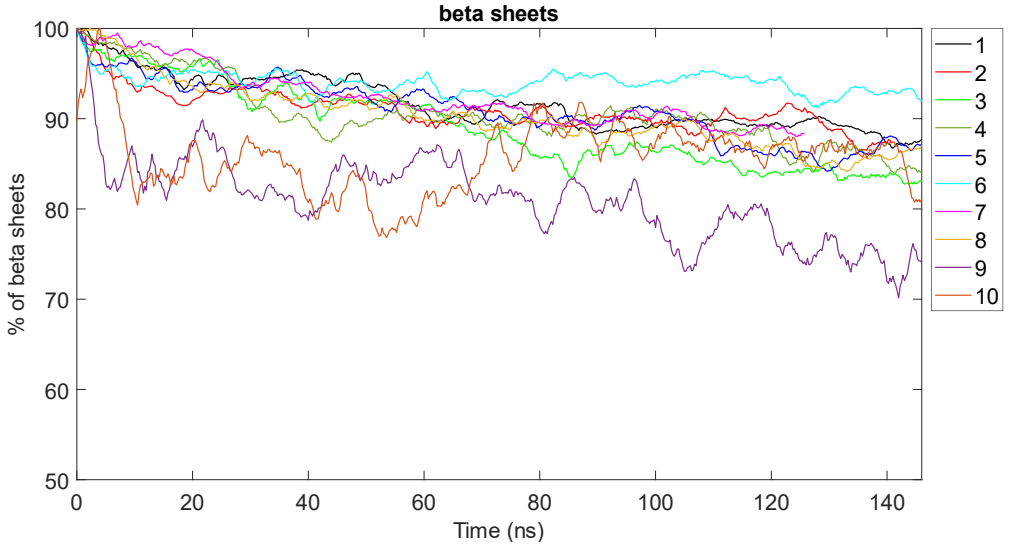

Figure 28 beta sheets percentage in time for all configurations of peptide 1

Hydrophobic contacts in time (mobile mean on 20 frames = 0,5 ns) are represented in Figure 29. They are maintained within a decrease of 10% in the worst case and in general do not evidence great differences among configurations.

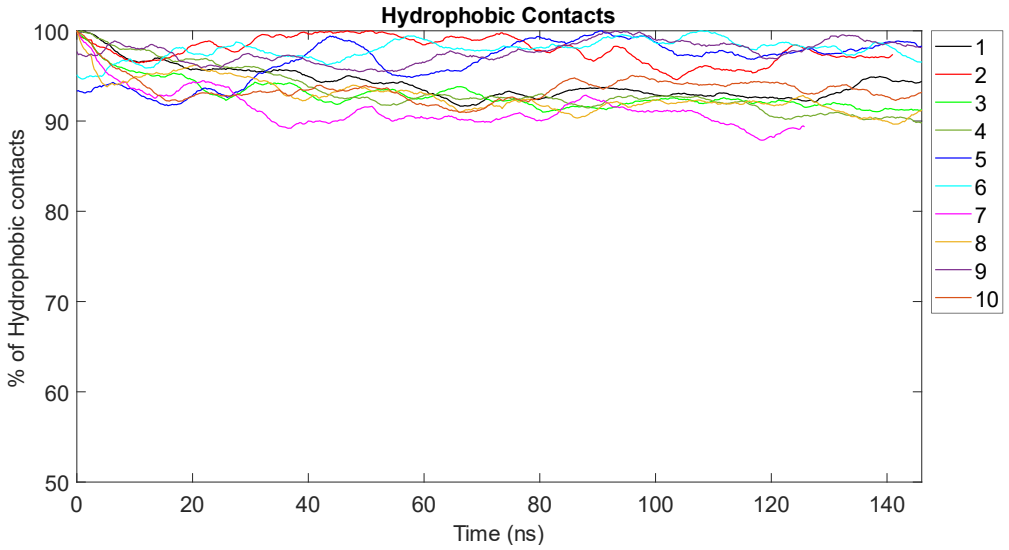

Figure 29 hydrophobic contacts percentage in time for all configurations of peptide 1

The Stability Index, reported in Table 5, sees on the first positions of the stability ranking configurations 5, 6 and 7, which all have in common antiparallel β-sheets, and on the last positions configuration 9 and 10, both with parallel β-sheets. The Stability Index mainly accounts for interactions involving the backbone and neglects lateral residues. This is the reason why it does not favor configuration 1 for example, even if from a visual analysis it seems to have a peculiar spatial organization in a steric zipper, increasing its stability while minimizing steric hindrance.

CONFIGURATION	STABILITY INDEX	CONFIGURATIONS RANKING
1	58.6816	Configuration 6
2	57.0429	Configuration 5
3	55.3939	Configuration 7
4	48.7652	Configuration 1
5	64.5345	Configuration 2
6	65.4438	Configuration 8
7	58.9050	Configuration 3
8	55.9296	Configuration 4
9	47.3761	Configuration 9
10	46.9997	Configuration 10

Table 5 Stability Index for all configurations and ranking from most to least stable (top to bottom)

A possible way to consider this aspect is through native contacts. They are the interactions that are present in nature, in the hypothesized starting amyloid-like configurations, so the number of native contacts gives an indication of how close a final simulated configuration is to the initial one. In this case - reported in Figure 30 - configuration 1 maintains the highest percentage of native contacts, which means that a lot of the original lateral residue interactions at time 0 ns are maintained. Therefore, configuration 1 is likely to be one of the possible stable structures. Configuration 9, instead, loses a higher percentage of native contacts, meaning that the final horseshoe shaped structure is not the one initially hypothesized. This does not exclude that it might be an intermediate to form another stable vesicle-like structure, if the simulation times were prolonged.

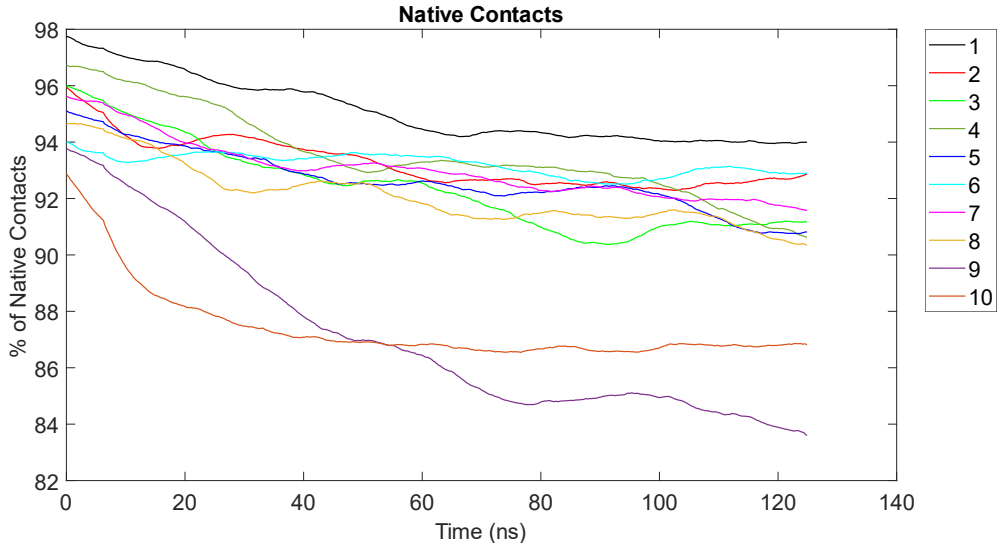

Figure 30 native contacts percentage in time for all configurations of peptide 1

3.1.2 Discussion

Overall, the most stable configurations according to the SI are 5, 6 and 7, which visibly form a stable fibrillar structure (Figure 31a).

The least stable are configuration 9 and 10, represented in Figure 31b and Figure 31c respectively. While it is true that both structures lose a percentage of β-sheets and strands lose their conformation, configuration 9 has a peculiar, curved shape, that can resemble a structure forming a vesicle. It might be possible that, even if it is considered unstable, it is originating a stable vesicle with hydrophobic residue in its core.

Configuration 1, in Figure 31d has an intermediate value of SI, but it is ordered, thanks to the steric zipper created by lateral residues, in particular by the aromatic rings of phenylalanine.

Configuration 3 (Figure 31e) has a U-shaped structure, resembling configuration 9. Residues are represented as follows: tyrosine in green, phenylalanine in purple, glutamine in orange and lysine in light blue. Tyrosine residues form the hydrophobic core and positively charged lysine residues are exposed externally toward the solvent.
The configuration is ordered in its lateral packing and presents a steric zipper. Figure 31e on the right shows two adjacent β-sheets. Polar tyrosine interacts via electrostatic interactions with negatively charged glutamine belonging to the same peptide, giving it a curved shape. Tyrosine residues rings of two lateral β-sheets interact in an ordered way and confer the curved shape to the structure, together with aromatic interactions between phenylalanine residues of different β-sheets. Positively charges lysine is always toward the external part of the structure.

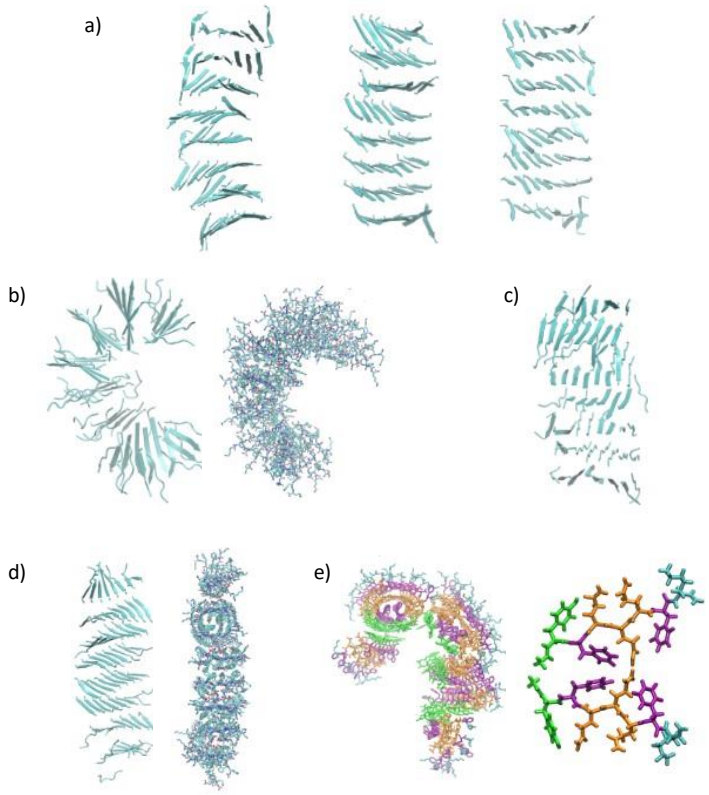

Figure 31 configurations of peptide number 1: configuration 5, 6, 7 from left to right (a), configuration 9 in two visualizations (b), configuration 10 (c), configuration 1 in two visualizations (d), configuration 3 and relative lateral packing (e)

3.2 PEPTIDE NUMBER 2 (AC-YF(I)QQQFK-CONH2)

3.2.1 Results

Peptide number 2 has a similar sequence compared to peptide number 1, but with the introduction of a halogenated residue. In particular, the phenylalanine closer to the N terminus is halogenated phenylalanine. It can give rise to halogen bonds, so it is interesting to evaluate molecular interactions to see how this modification influences the stability of configurations.

RMSD graphs can be distinguished into two behaviors: configurations 9, 2, 3 and 7 have higher values and an increasing trend, whereas all other configurations reach stability sooner and have lower plateau values. Configuration 6 seems to be the most stable, already after 20 ns.

RMSF is in accordance with RMSD, so configurations 9 and 3 are the ones that fluctuate more, meaning they differ a lot from the mean conformation (and this is confirmed by the qualitative observation of simulations). The configurations that fluctuate less are: 7, 10, 1 and 6.

Hydrogen bonds are maintained in the highest percentage by configuration 10, while configuration 2 and 9 lose the highest percentage of bonds, meaning they are less stable. All other configurations have a linear decrease between 5% and 10%.

The β-sheets content percentage for configuration 10 has an initial increase of 25% that settles on +15%. Configuration 6 has a good stability, whereas 9 and 2 seem unstable. All other configurations decrease their β-sheets of the 10-15%.

Hydrophobic contacts characterize configuration 8 as the most unstable, losing 20% of contacts. Configuration 7 also has an unstable trend. Configurations 5, 6 and 9 show a slight increase in hydrophobic contacts, while all other structures have a slight decrease (within 10%).

The Stability Index classifies configurations 6, 5 and 3 as the most stable and 9, 10 and 2 as the least stable.

Native contacts analysis shows in Figure 32 that configuration 1 is the one that maintains the highest percentage, meaning that it is probably the most similar to the initial hypothesized

structure and has a higher probability of being stable. Configurations 2, 9 and 10 show the greatest decrease, indicating that these structures are less likely.

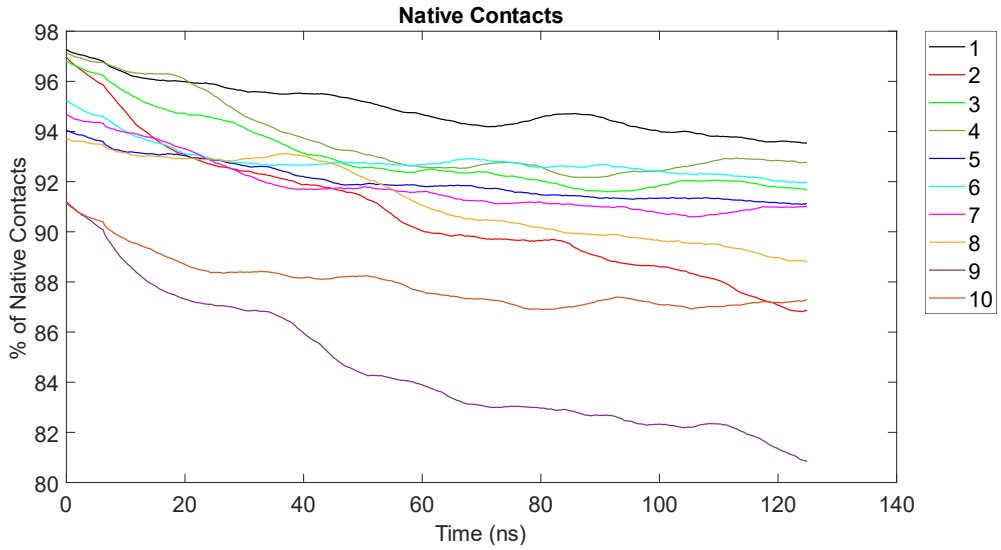

Figure 32 native contacts percentage in time for all configurations of peptide 2

3.2.2 Discussion

The selected configurations, that are considered relevant for the peptide and for the comparison with peptide 1, are described as follows.

Configuration 6 is represented in Figure 33a and its observation confirms the structure is stable and maintains β-sheets integrity, also thanks to the compact lateral packing showed on the right. The postively charged halogen atom of phenylalanine (belonging to the residue in grey) forms a halogen bond with the nitrogen atom from the third glutamine residue of the peptide below.

Configuration 3 in Figure 33b is also stable, but the lateral packing of peptides from adjacent β-sheets is less compact in this case. The main interactions seem to be T-shaped pi-stackings between tyrosine aromatic rings, hydrogen bonds between tyrosine -OH groups, and possibly hydrogen bonds between glutamine residues from peptides of adjacent β-sheets.

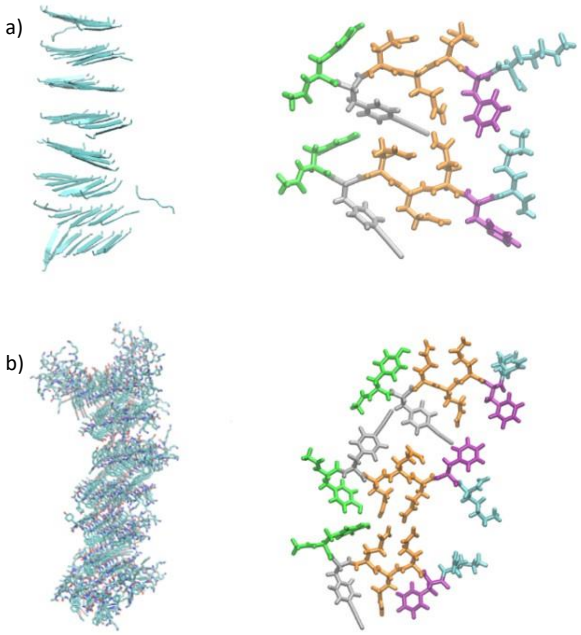

Figure 33 configuration 6 and relative lateral packing (a) configuration 3 and relative lateral packing (b) of peptide 2

Configuration 2, which is ranked as unstable from the Stability Index and shows a decrease in native contacts, tends to assume a curved shape where β-sheets at the extremities are distanced, but at the same time it presents an ordered disposition of lateral residues in the central part of the fibril, similar to a steric zipper. This configuration can be observed in Figure 34a on the right and lateral interactions are reported on the left. Halogen bonds among backbone nitrogen atoms and positive halogen atoms are visible, as well as hydrogen bonds among the hydrogen of one glutamine and the nitrogen of the laterally facing glutamine.

Configurations 9 and 10 are classified as unstable by the Stability Index and the native contacts analysis.

Configuration 9 is represented in Figure 34b. It seems to be guided mainly by hydrophobic interactions among tyrosine residues, that influence the lateral packing. The distance in correspondence of the right edge of the peptides, due to the lack of lateral interactions, gives the whole structure a curved shape. Contrary to peptide 1, configurations 9 in this case does not display a clear disposition of hydrophilic residues toward the external part (and hydrophobic residues in the core) because phenylalanine residues are also exposed, and its shape is less curved.

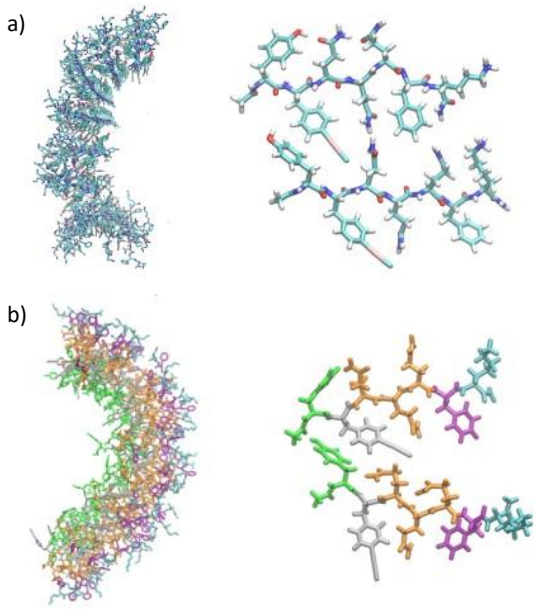

Figure 34 configuration 2 and relative lateral packing (a), configuration 9 and relative lateral packing (b) of peptide 2

Configuration 10 loses some of its integrity and some external β-strands tend to dissociate from the structure, indicating its instability similarly to all the other peptides. This is evidenced in Figure 35a.

Configuration 1, the most stable from the native contacts point of view, has a very compact and ordered lateral packing, visible in Figure 35b, where the face-to-face orientation allows the formation of two halogen bonds between each pair of peptides from different sheets in this interface. They form between the positively charged halogen atom, in correspondence of the σ-hole and an oxygen atom in this case. The back-to-back interface instead is characterized by hydrogen bonds and salt bridges.

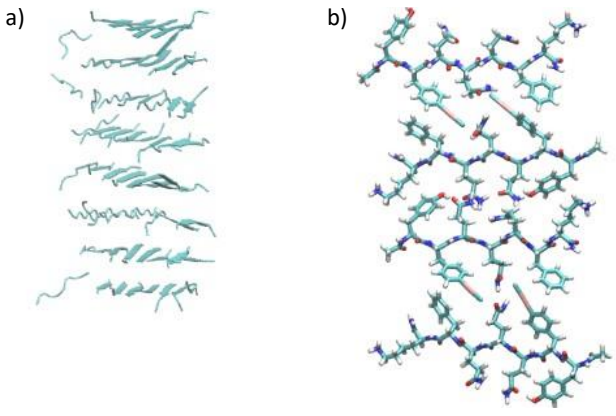

Figure 35 configuration 10 in β-sheets visualization (a) and configuration 1 lateral packing (b) of peptide 2, stabilized by halogen bonds.

3.3 PEPTIDE NUMBER 3 (AC-YFQQQF(I)K-CONH2)

3.3.1 Results

Peptide number 3 only differs from peptide number 2 in the position of the halogenated phenylalanine, which is closer to the C terminus this time. The second amino acid is a regular phenylalanine residue. It is interesting to compare this sequence to the previous two.

RMSD graphs show a completely unstable configuration 5. Configurations 2, 3 and 9 have a similar trend, reaching higher values than the others. Configurations 4 and 6 seem to be the most stable.
RMSF mean values allow to observe that configurations 6, 2, 10 and 7 fluctuate less. The one configuration that differs a lot from the others is configuration 5, reaching the highest value.
Hydrogen bonds characterize configuration 7 as the least stable (with a decrease of 18%). Configuration 2 is the most stable and shows a slight increase of hydrogen bonds, while all other structures have a similar decrease up to 12%.
The β-sheets in time have a similar trend with respect to peptide number 2. Configuration 10 sees an increment followed by a decrease. Configuration 9 loses the highest percentage of β-sheets. Configuration 7 and 3 lose within 10-15% of β-sheets content, while all other configurations are maintained within a loss of 10%.
Hydrophobic contacts are rather constant in the simulation time and do not decrease of more than 10%.

The Stability Index ranks configurations 5, 2 and 1 as the most stable and 9, 10 and 4 as the least stable respectively.
Native contacts analysis, however, reveals that configuration 4 (in dark green) maintains the highest percentage of interactions originally present at time t=0 ns. Configurations 9 and 10 are confirmed to be unstable. Configuration 2 is confirmed to be stable, in accordance with the SI ranking. Native contacts graphs are reported in Figure 36.

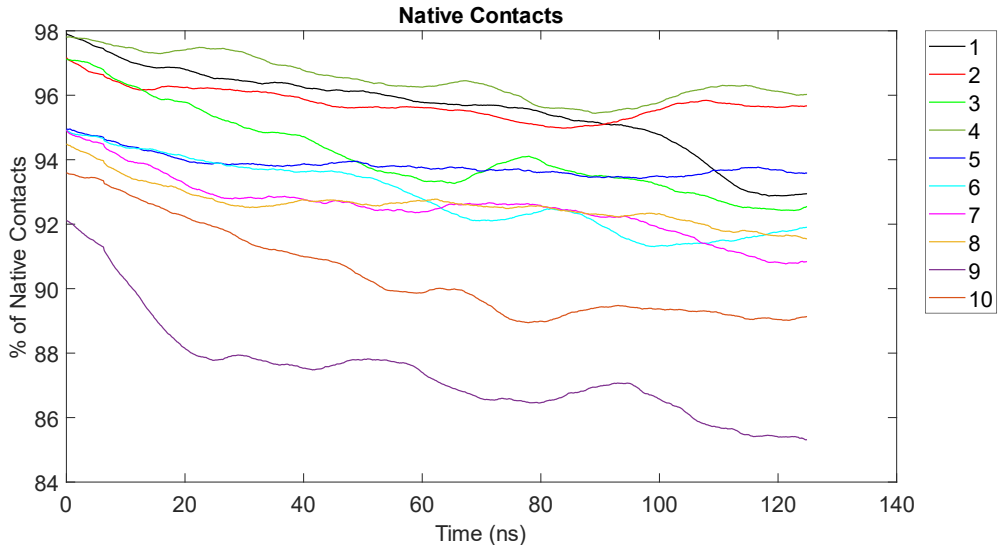

Figure 36 native contacts percentage in time for all configurations of peptide 3

3.3.2 Discussion

The selected configurations, that are considered relevant for the peptide and for the comparison with peptide 1, are described as follows.

Configuration 5, even though it has the highest SI, is unstable, because the structure dissociates in two parts, as it is possible to note in Figure 37a. This result is different compared to peptide 1, so it might be due to the halogenated phenylalanine in this particular position instead of regular phenylalanine.
Configuration 2 is stable and forms a compact fibrillar structure (Figure 37b). Its order is visible in Figure 37b on the left, even in the organization of parallel strands in a single β-sheet, which minimizes the hindrance of halogenated phenylalanine by always exposing it on the same side.

Configuration 1 is shown in Figure 37c and presents an ordered lateral packing of β-sheets, stabilized also by the formation of halogen bonds between the halogen atom of one peptide and the nitrogen belonging to the peptide in the β-sheet laterally located. The face-to-face disposition of β-sheets in the lateral packing originates two different interfaces: one containing two halogen bonds and one where no halogen bonds are present, probably stabilized mainly by hydrogen bonds and salt bridges. Configuration 1 is very ordered, as it

happens in peptide 1 for this particular configuration, that presents an ordered steric zipper structure.

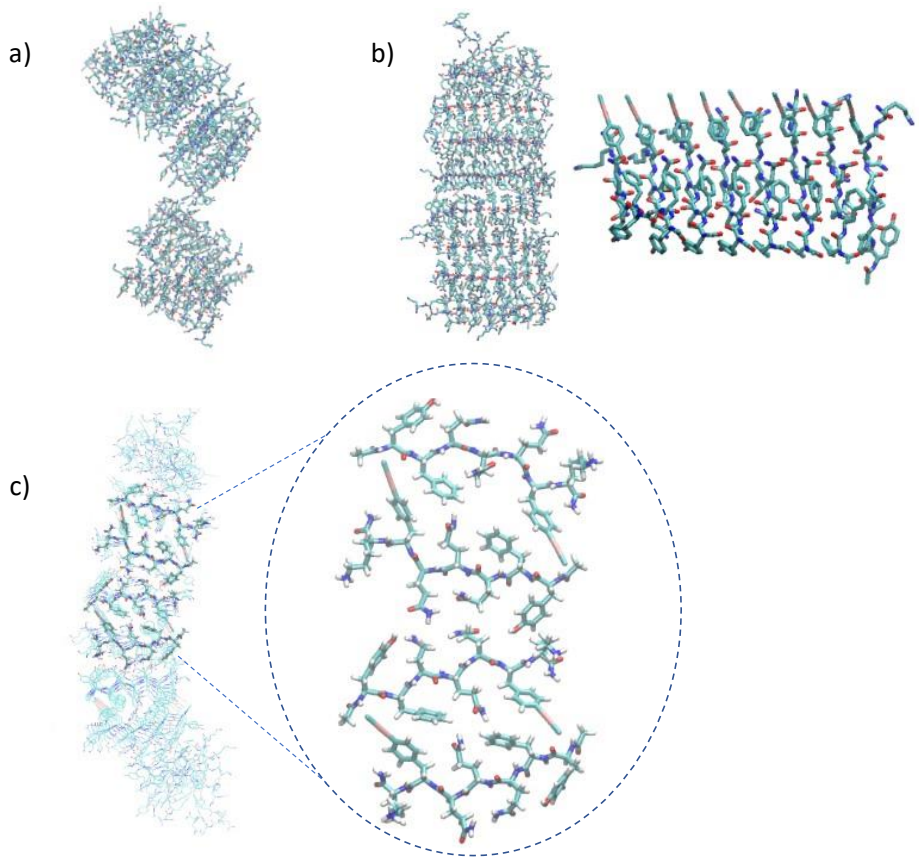

Figure 37 configuration 5 (a), configuration 2 and relative zoom of its ordered parallel β-strands forming a single β-sheet (b), configuration 1 of peptide 3 and zoom of four peptides from adjacent sheets showing halogen bonds (c)

Configuration 9 and configuration 10 are shown in Figure 38a and Figure 38b (in two visualizations) respectively. It is interesting to note that configuration 10, even if classified as one of the least stable in terms of SI, presents a very compact structure with an ordered lateral packing. Compared to the same structure in peptide 1 it retains a much higher percentage of β-sheets, as it emerges in Figure 38b on the right (β-sheets visualization).

The lateral packing of peptides from adjacent β-sheets in configuration 10 is showed in Figure 38c on the left, whereas on the right there is a side view of two peptides belonging to the same sheet.

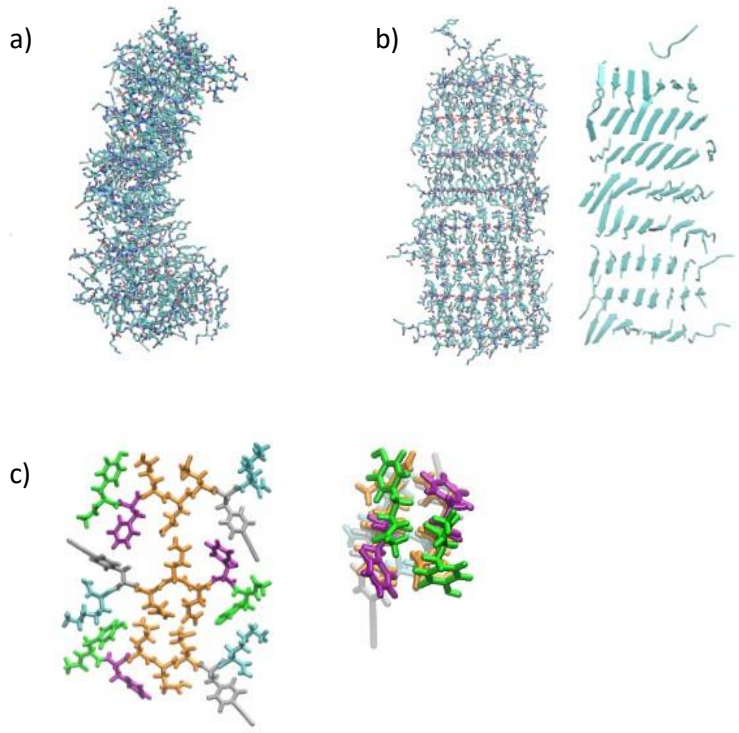

Figure 38 configuration 9 (a), configuration 10 in two visualizations (b), configuration 10 lateral packing of peptides in adjacent β-sheets (c, on the left) and in β-strand (c, on the right)of peptide 3

Configuration 4 in Figure 39a shows a very compact and ordered structure, the halogenated phenylalanine is disposed on the same side on the strands but on the opposite side in adjacent sheets, so this might be one of the structure likely to be stable, as it emerges from native contacts.

Configuration 3, unlike peptide 1, does not tend to form a micellar structure in this sequence, but it assumes an ordered fibrillar conformation very similar to configuration 1. Figure 39b on the left evidences its shape and the disposition of some residues: halogenated phenylalanine is represented in grey, phenylalanine in purple, tyrosine in green and glutamine and lysine both in light blue. There is no clear division of the residue types in particular ways (it is not possible to identify a hydrophobic core and polar outside and in fact the structure is not prone to forming a vesicle). The lateral packing is once again stabilized by halogen bonds. Figure 39b on the right allows to appreciate also aromatic T-shaped pi-stackings between phenylalanine residues, interactions between polar groups of tyrosine, glutamine-glutamine interactions.

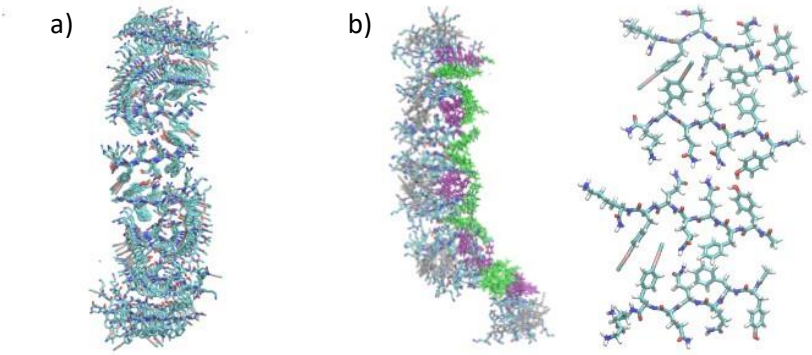

Figure 39 configuration 4 (a), configuration 3 of peptide 3 (b, on the left) and lateral packing of peptides form adjacent sheets (b, on the right)

3.4 PEPTIDE NUMBER 4 (AC-FFQQQFK-CONH2)

3.4.1 Results

Peptide number 4 differs from peptide number 1 only for an amino acid substitution, because it has phenylalanine instead of tyrosine at the N terminus. it is interesting to evaluate the effects. Another interesting comparison is between peptide number 4 and the following peptides (number 5, number 6 and number 7) with the first amino acid modified: D-phenylalanine, homophenylalanine and β-phenylalanine respectively.

RMSD graphs are distinguished into three behaviors. Configuration 3, 9, 2 reach higher RMSD values and are unstable; configuration 10, 5, 4 reach intermediate values but with an increasing trend; all other configurations show a stable plateau. Configuration 7 is particularly stable.

Mean RMSF evidence that configuration 7 fluctuates less, whereas configuration 9 and 3 are the ones that depart more with respect to a mean conformation, and in fact they assume a U-shape that might suggest the possible formation of vesicles or micelles.

Hydrogen bonds for configuration 4 have an interesting increase at time 150 ns. Configuration 9 and 10 are unstable with a decrease of 15-20%, configurations 5, 6 and 7 (that have a high SI) show intermediate values and slightly decreasing trend. Configuration 1 instead has a first increase followed by a slight decrease, but overall it maintains intermediate values.

β-sheets percentage reflects the previous observations. Configuration 4 once again is particularly stable, configurations 9 and 10 are not stable with a decrease of up to 30% (compared to the 5-20% of the other structures). Configuration 7 seems to be stable.
Hydrophobic contacts for all configuration maintain a good percentage, with a loss of 10% at worst. Configurations 5 and 6, which are ranked as the best from the SI, show an increase of 5-10% of hydrophobic contacts.

Stability Index ranks configuration 5, 6, 7 as the best and 9, 10, 4 as the worst.
Native contacts (Figure 40) analysis, however, evidences that configuration 4 (dark green) seems to be very likely, followed by 7, 1, 5 and 6. Configurations 9 and 10 (orange and purple) are the least stable. Configuration 3, that assumes a curved shape, does not seem to be among the most stable, possibly because it might be an intermediate to form another vesicle-like structure.

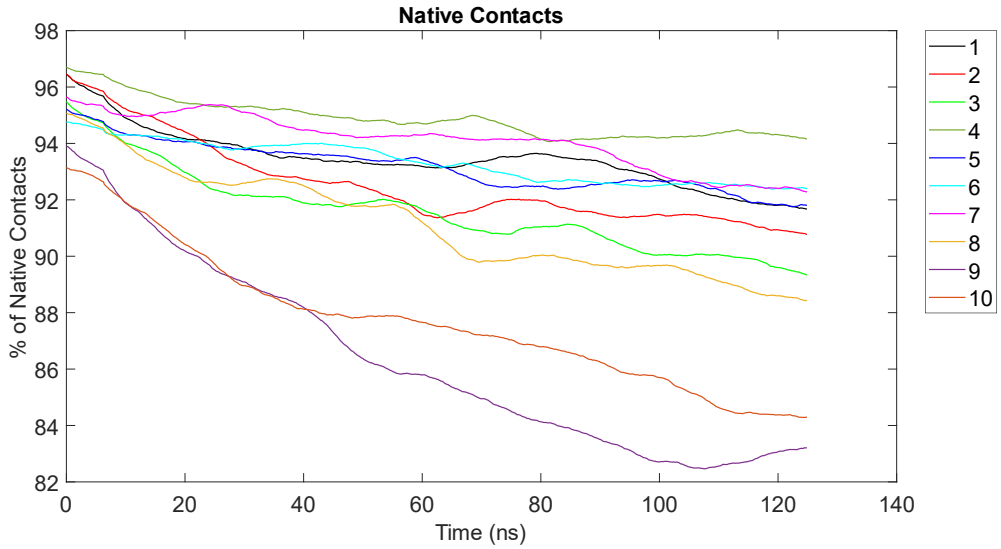

Figure 40 native contacts percentage in time for all configurations of peptide 4

3.4.2 Discussion

Configurations 5, 6 and 7, which are the ones that result most stable from the SI, are represented in Figure 41a. They do not differ a lot from peptide 1, in fact these three configurations are also at the top three configurations in the SI ranking for peptide 1. This result suggests that antiparallel strands configurations are favored in terms of stability, when considering mainly the interactions of the backbone (SI considers hydrogen bonds and

RMSF). The structures are fibrillar, compact and maintain the β-sheets structures. Configuration 5 (on the right) seems to assume a curved conformation of each sheet.

Configuration 9 (Figure 41b, on the left) also loses some of the β-sheets organization but it seems to assume a curved shape that might suggest the formation of a vesicle.
Configuration 10 (Figure 41b, on the right) loses much of the β-sheets integrity because the single strands lose their structure, giving the overall configuration a disordered appearance.

Configuration 9 is not the only structure that assumes this U-shaped conformation, because the same behavior can be observed for configuration 3. What brings together these two configurations is their shape and the residue disposition. As it can be observed in Figure 41c that compares configuration 3 and 9, in both cases hydrophobic residues of phenylalanine (purple) tend to dispose in the inner core, polar glutamine (orange) and positively charged lysine (light blue) tend to dispose towards the external part of the vesicle-forming structure.

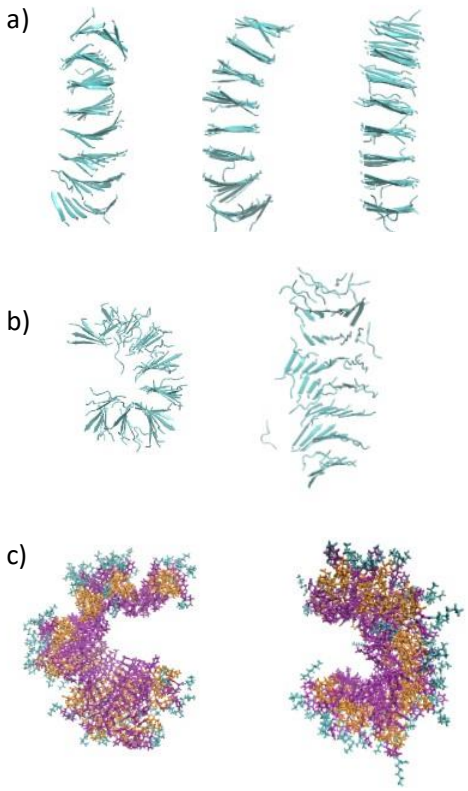

Figure 41 configuration 5, 6 and 7 (figure a, left to right), configuration 9 and configuration 10 (b, left to right), configuration 3 and 9 of peptide 4 (c, left to right)

Configuration 2, that together with configuration 3 and 9 reach high values of RMSD and mean RMSF and might seem unstable, also assumes a folded structure for peptide 4, that seems to be ordered also thanks to aromatic interactions. The β-strands are preserved in their secondary structure. Two visualizations of configuration 2 are reported in Figure 42a.

Lastly, configuration 4 is noteworthy because the native contacts are well preserved, meaning that the hypothesized starting configuration is likely to be close to the natural structure and it seems to be one of the most stable configurations from quantitative analysis. Figure 42b shows configuration 4 in β-sheets visualization (on the right), the ordered fibrillar structure (in the middle). It seems to present a steric zipper, where lateral residues of two peptides belonging to adjacent β-sheets (showed on the right) are disposed to minimize their steric hindrance. Phenylalanine rings can also interact via aromatic interactions, in this case parallel pi-stacking.

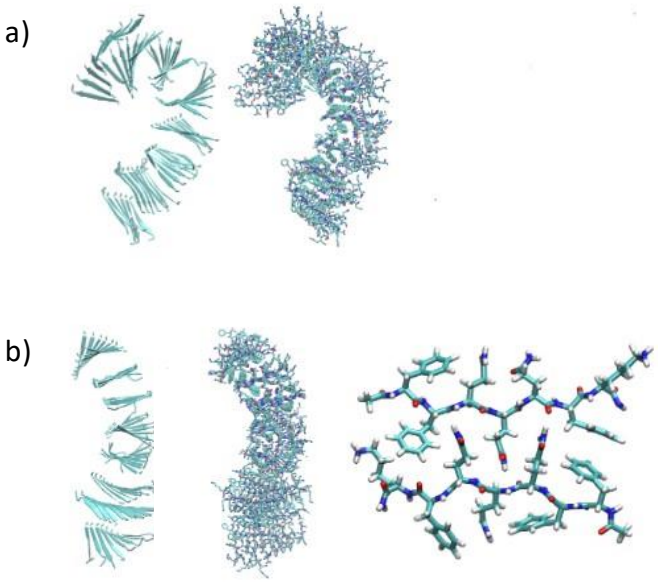

Figure 42 configuration 2 in two visualizations (a), configuration 4 of peptide 4 (b): fibril, β-sheets visualization, isolated peptides from two adjacent β-sheets

3.5 PEPTIDE NUMBER 5 (AC-(D-F)FQQQFK-CONH2)

3.5.1 Results

Peptide number 5 differs from the previous, as it has the enantiomer D-Phenylalanine instead of L-Phenylalanine.

RMSD graphs reveal instability for most structures, except for configurations 7, 8, 3 that reach a stable value.
Mean RMSF analyses evidence that configuration 7 has the lowest value, meaning that, overall, it fluctuates less than all the other configurations. Configuration 3 and 9 have the highest RMSF.
Hydrogen bonds show the stability of configuration 7, that maintains an almost constants number of interactions. Configuration 9, 10, 4 and 1 have a trend that reflects instability.
The β-sheets percentage in time also confirm that configuration 7 is extremely stable and differentiates from all the other structures.
Hydrophobic contacts are well preserved in configuration 7, in accordance with the other parameters, and show a slight increase in configuration 5 and 6.

The configurations ranking from the Stability Index is in accordance with the native contacts results, reported in Figure 43. The structure that result to be more stable are configuration 7 and 8, while the ones that are unstable are configuration 9 and 10. It means that the final configuration 7 and 8 are very close to the initial hypothesized fibrillar structures, characterized by antiparallel strand (as observed in the case of peptide 1). On the other hand, configuration 9 and 10 depart from their initial structure.

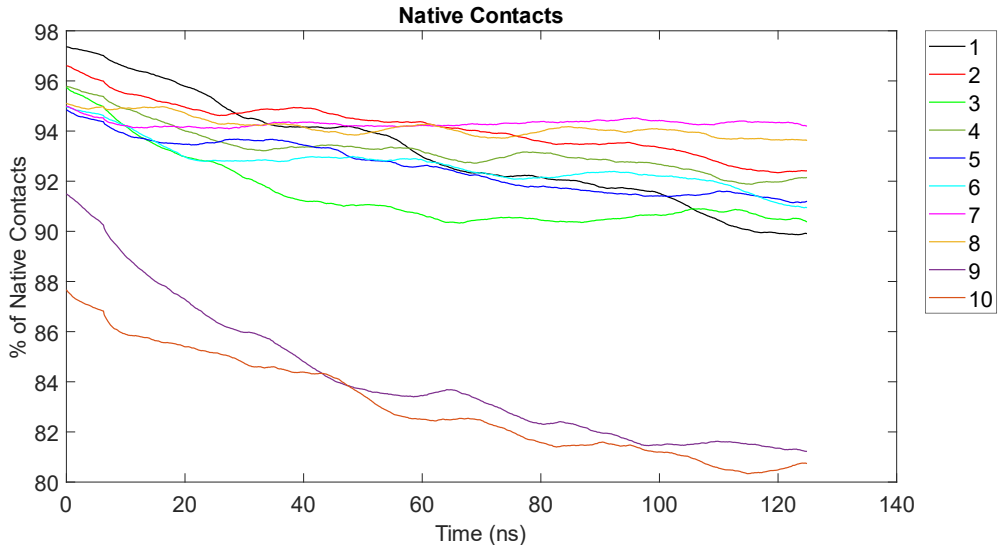

Figure 43 native contacts percentage in time for all configurations of peptide 5

3.5.2 Discussion

Configuration 7 and 8 are shown in Figure 44a (β-sheets visualization and lateral packing of β-sheets) and Figure 44b (lateral packing). They are stable also from the point of view of a qualitative observation. The β-sheets are ordered and maintain their integrity, the resulting structures are compact fibrils. The lateral packing of configuration 7 is very precise and repetitive. It is possible to identify in configuration 7 parallel pi-stacking interactions between a phenylalanine residue and the D-phenylalanine belonging to the laterally facing peptide from another β-sheet. Glutamine-glutamine interactions are also present. The β-sheets are parallel, and this increases the stability with respect to antiparallel β-sheets in configuration 8, even if both configurations are stable. Configuration 8, instead, seems to favor T-shaped pi-stackings, with a broader space between β-sheets with some hollow regions.

Configuration 9 and 10, represented in Figure 44c and Figure 44d respectively, do not differ from the peptide number 2, number 3 (and number 6 reported in the following pages). Configuration 9 always assumes a curved shape with phenylalanine (purple) and D-phenylalanine (grey) residues all exposed on one side and lysine residues (light blue) prevalently exposed on the other. Glutamine is always represented in orange. Compared to peptide number 4, configuration 9 does not form an almost completely vesicle-like folded structure in the simulation time. The D-phenylalanine substitution seems to have an impact

on the tendency to form vesicles, probably because when all hydrophobic residues are the same enantiomers (as in peptide 4) they have a stronger tendency to associate at the inner core. In configuration 10 some peptides detach from the structure, which is therefore unstable.

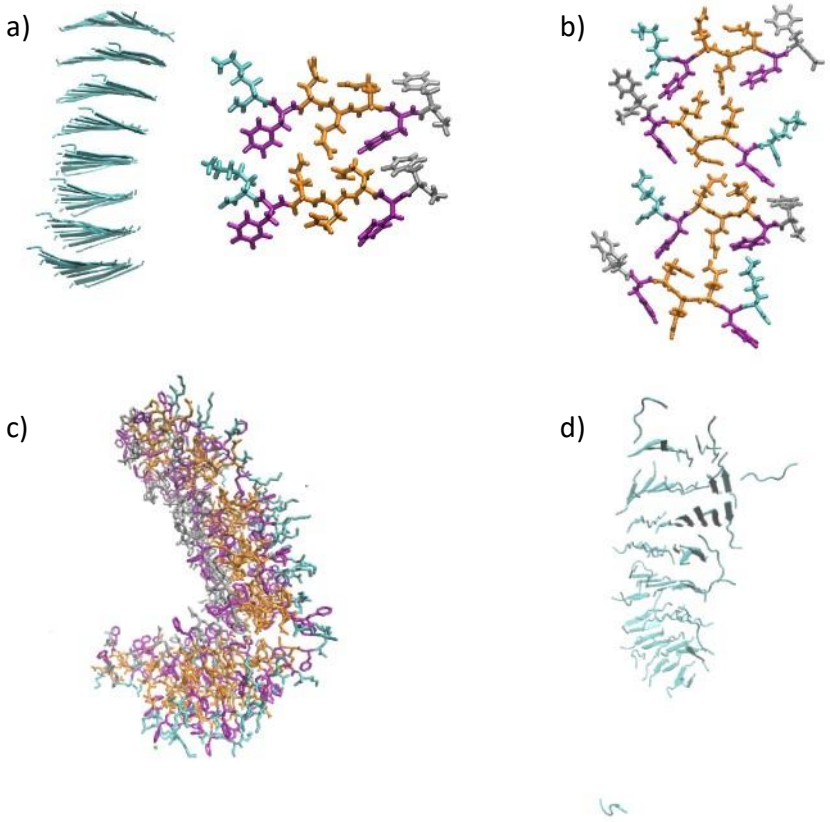

Figure 44 configuration 7 in β-sheets representation (a, on the left) and lateral packing (a, on the right), configuration 8 lateral packing (b), configuration 9 (c) and configuration 10 (d) of peptide 5

3.6 PEPTIDE NUMBER 6 (AC-(*H*F)FQQQFK-CONH2)

3.6.1 Results

Peptide number six differs from the previous in the first amino acid, with the substitution with homophenylalanine.

RMSD graphs reveal that all configurations reach a plateau, except for configuration 9, 2 and 3 that result more unstable.
RMSF analysis show a mean RMSF that results low for configuration 1 and 7. There is not a noticeable difference between configurations in this case, except for configuration 9 with the highest value.
Hydrogen bonds percentage shows that configuration 6 is the most stable, configuration 1 has an increase and configuration 9 is the least stable with the greatest loss of interactions.
The β-sheets analysis evidences that configuration 6 is again the most stable, while configurations 9, 10 the least.
Hydrophobic contacts for 5,6,7 and 1 maintain a higher percentage (almost 100%) compared to the other structures that lose about 10%.

The Stability Index classifies configuration 6 and 7 as stable and configuration 9 and 10 as unstable. Native contacts analysis, reported in Figure 45, confirms this result: 6 and 7 have a value that is almost constant, while 9 and 10 show the greatest decrease. It is also possible to identify another stable structure, configuration 1 (represented in black), that maintains the highest percentage of the initial interactions.

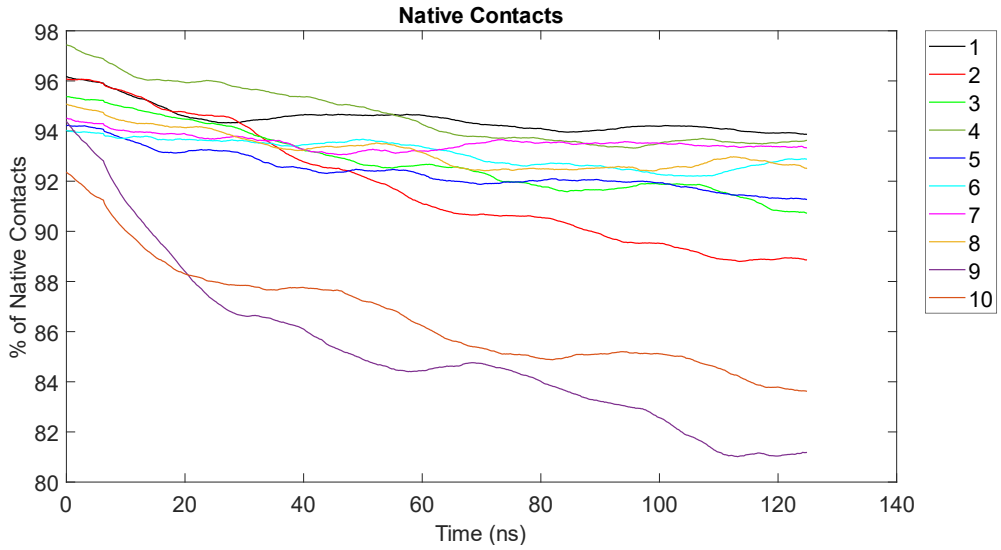

Figure 45 native contacts percentage in time for all configurations of peptide 6

3.6.2 Discussion

Configuration 1 and 7 and the relative lateral packing are shown in Figure 46a and Figure 46b respectively. In the first case it is possible to note a very ordered disposition of glutamine residues in a steric zipper at one interface (the lateral packing of configuration 1 is face-to-face so the other interface is back-to-back and does not display glutamine side chains). Configuration 7 has a face-to-back lateral packing so the same interface is repeated and it presents a very ordered disposition of homophenylalanine and phenylalanine benzene rings.

Configuration 6 is similar to configuration 7 and its lateral packing is shown in Figure 46c, with a slightly less compact structure even though the isolated peptides belong to the inner part of the configuration.

The fibrillar structure of these configurations is clearly more ordered than configuration 9 and 10, in Figure 46d and Figure 46e respectively, where the initial structure is not maintained. However, the residue disposition in configuration 9 and the curved shape might suggest the possibility of formation of a vesicle, that - with prolonged simulation times - might actually result stable. Homophenylalanine is in grey, phenylalanine in purple, glutamine in orange and lysine in light blue in this representation.

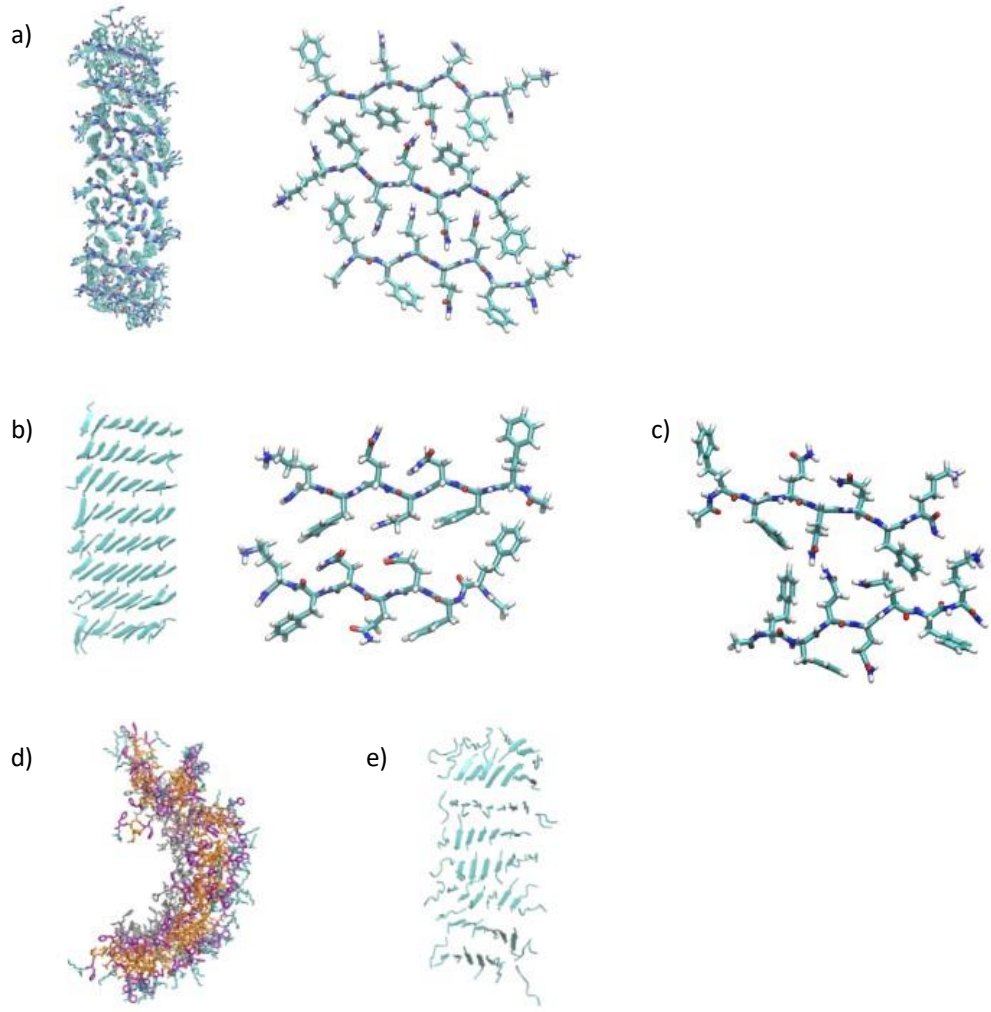

Figure 46 configuration 1 and lateral packing (a), configuration 7 in β-sheets (b, on the left) and relative lateral packing (b, on the right), configuration 6 lateral packing (c), configuration 9 residues (d) and configuration 10 in β-sheets(e)

3.7 PEPTIDE NUMBER 7 (AC-(*β*F)FQQQFK-CONH2)

3.7.1 Results

Peptide number seven is characterized by β-phenylalanine as the first amino acid at the N terminus.

RMSD trends for configurations 6, 7 and 8 are the most stable. Contrary to most other peptides, configuration 1 seems to be unstable in this case.
RMSF values are in accordance with considerations relative to RMSD.
Hydrogen bonds shows a stable trend for configuration 3, and a decreasing unstable behavior for configuration 9.
The β-sheets see an increase of about 30% for configuration 3 and 4 and a decrease in all other cases, with 9, 1 and 10 being the worst.
Hydrophobic contacts show a good trend for configuration 6 with a consistent increase of about 15% of its initial value.

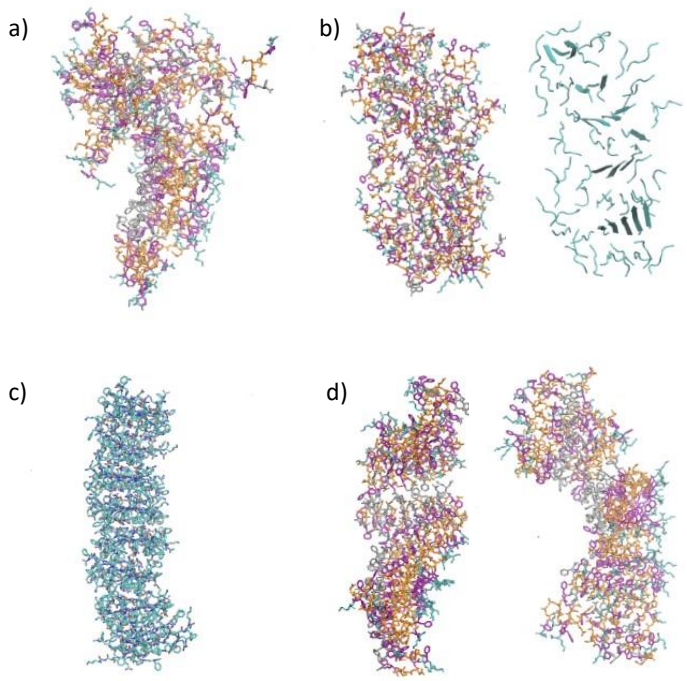

Figure 47 configuration 9 (a), configuration 10 in two visualizations: residues (b, on the left) and β-sheets (b, on the right), configuration 6 (c), two views of configuration 3 (d)

The Stability Index, that classifies configuration 3 and 6 as the most stable, is once again in accordance with the observation of native contacts. In this case it is possible to note in Figure 48 that the light blue line, representing configuration 3, is the one with the most constant trend, as it decreases less than all other configurations. The instability of configurations 9 and 10 is evidenced both in the Stability Index ranking and in the native contacts graphs, where the decrease is noticeable.

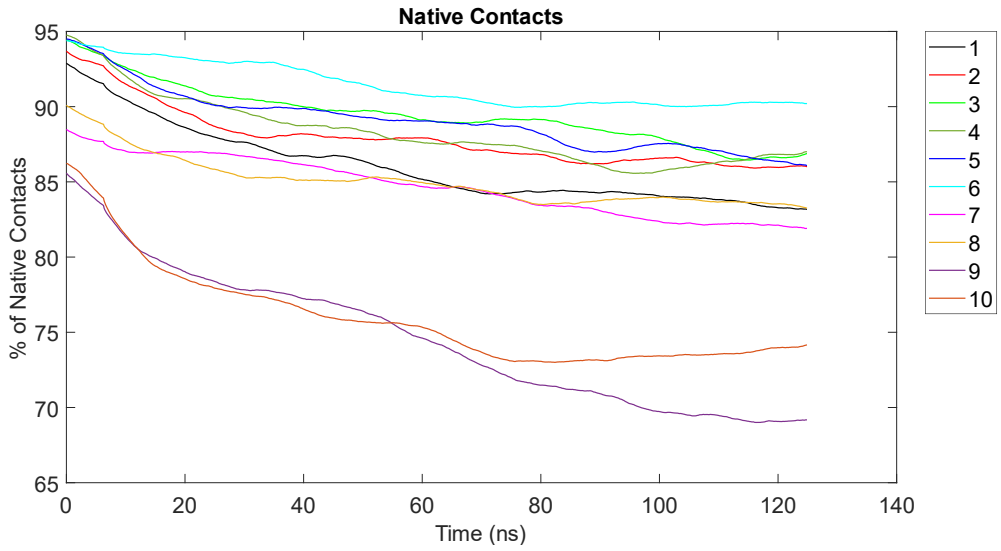

Figure 48 native contacts percentage in time for all configurations of peptide 7

3.7.2 Discussion

Configuration 9 (Figure 47a in the previous page) is considered unstable because it departs from the initial fibrillar configuration, but in the case of this specific peptide it does not seem to form a vesicle-like structure where disposition of residues is ordered. Indeed, the random disposition of hydrophobic and hydrophilic residues does not suggest the possible formation of a hydrophobic core.

Configuration 10 is evidently unstable, because as it is shown in Figure 47b the β-strands - and consequently the β-sheets - completely lose their integrity. The fibril is possibly disrupting and assuming a less ordered and less compact structure in the simulation time.

Configuration 6 is reported in Figure 47c. Configuration 3 is represented for clarity in two different views in Figure 47d. Configuration 3, in particular in the view reported on the left, has a shape that seems to be resembling a spiral, where the residue disposition is ordered. The other view on the right provides a complete picture of the curved shape.

3.8 STABLE AND UNSTABLE CONFIGURATIONS

To summarize the exposed results, the main considerations are schematized as follows.

It is possible to identify, for each peptide, the most stable and least stable configurations from the combined analysis of the previously described quantitative parameters and qualitative assessment. Results are summarized in Table 6.

Peptide	***Sequence***	***Stable configurations***	***Unstable configurations***
Peptide 1	Ac-YFQQQFK-conh2	Configuration 6-5-7-1	Configuration 9-10
Peptide 2	Ac-YF(I)QQQFK-conh2	Configuration 6-5-1	Configuration 9-10
Peptide 3	Ac-YFQQQF(I)K-conh2	Configuration 2-4-1	Configuration 9-10
Peptide 4	Ac-FFQQQFK- conh2	Configuration 5-6-7-4	Configuration 9-10
Peptide 5	Ac-(*D*-F)FQQQFK-conh2	Configuration 7-8	Configuration 9-10
Peptide 6	Ac-(*h*F)FQQQFK-conh2	Configuration 7-6-1	Configuration 9-10
Peptide 7	Ac-(βF)FQQQFK-conh2	Configuration 6-3	Configuration 9-10

Table 6 results of the stable and unstable configurations for each peptide

4 CONCLUSIONS

The concept of stability of a configuration for a given peptide sequence is relative to the analysis of different quantitative parameters, that might provide different results. The most relevant are the Stability Index and native contacts, that - combined with the qualitative observation of simulations - can return information on which structures are more likely to be found in nature for each peptide.

It is possible to conclude that the amino acid substitutions in the investigated sequences affect differently the stability of the configurations. The same configuration can have a different stability when considering different peptide sequences, as a consequence of the various noncovalent intermolecular interactions that play a key role in self-assembly. For instance, halogenated peptides behave differently than the non-halogenated version, thanks to the formation of halogen bonds. Moreover, the peptides sequences containing different stereoisomers (L or D-Phenylalanine) originate different interactions, as well as the peptides with phenylalanine variants (homophenylalanine and β-phenylalanine).

However, through the comparison of quantitative parameters and the visual observation of the behavior of configurations in time, it is possible to make some general assumptions, that can be valid for all the considered peptide sequences.

Configurations 9 and 10, characterized by antiparallel equifacial β-strands result unstable for all peptides. However, configuration 9 in most peptides seems to have a tendency to form vesicles with a precise residue disposition, where hydrophobic residues are towards the inner core and hydrophilic residues of lysine are exposed to the solvent.

Configuration 3, together with configuration 9, also seems to present U-shaped structures with the same residue disposition, suggesting the possible formation of vesicles. This is particularly evident in the first and fourth peptides.

Configuration 2 only has remarkable stability for the third peptide sequence and assumes a curved shape only in the fourth peptide. In all other cases, it has a fibrillar structure with intermediate stability.

Configuration 5, 6 and 7, with parallel β-strands, generally have a good stability in this analysis, meaning that they maintain the initially hypothesized fibrillar structures, with the integrity of β-sheets.

Configuration 1 and 4 often display a very ordered lateral packing of β-sheets, in a steric zipper organization. They originate stable structure and often do not differ, in terms of interactions, from the respective initial configurations.

Future developments might include the experimental characterization of the considered peptide sequences, with the objective to validate the results obtained from simulations and propose hypothesis on the actual structure underlying the observed peptides.

5 APPENDIX

5.1 PEPTIDE NUMBER 2

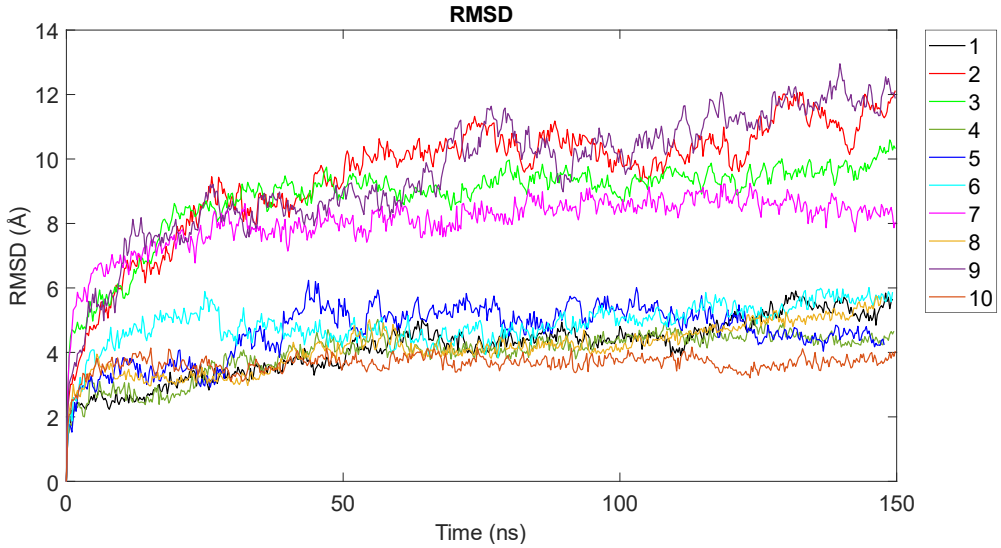

Figure 49 RMSD for all configurations of peptide 2

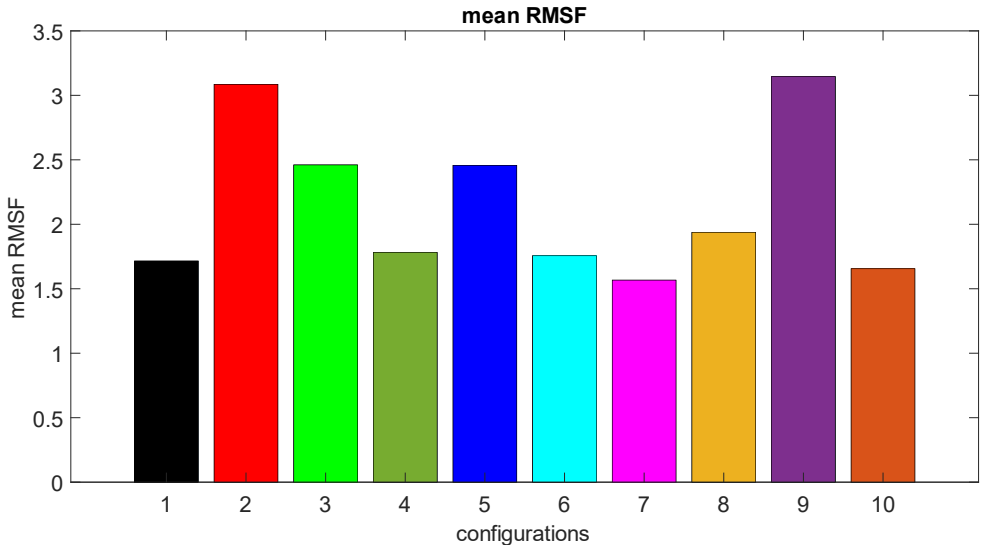

Figure 50 mean RMSF for all configurations of peptide 2

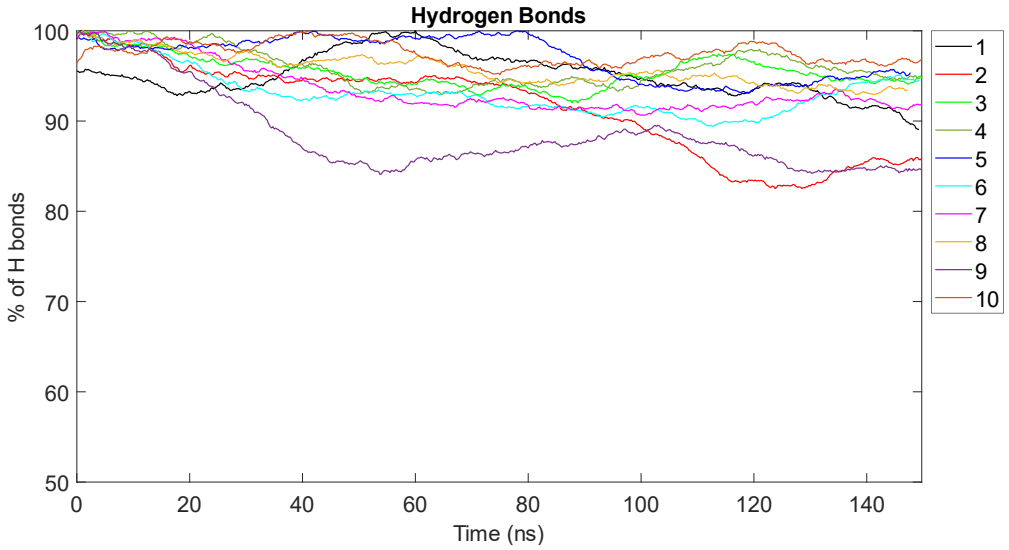

Figure 51 hydrogen bonds percentage in time for all configurations of peptide 2

Configuration	Stability Index	Ranking
1	53.3267	Configuration 6
2	47.9883	Configuration 5
3	55.6382	Configuration 3
4	50.3546	Configuration 1
5	59.1767	Configuration 8
6	61.4531	Configuration 7
7	50.4866	Configuration 4
8	52.4209	Configuration 2
9	38.4877	Configuration 10
10	47.5733	Configuration 9

Table 7 Stability Index for peptide 2 and configurations ranking from most to least stable

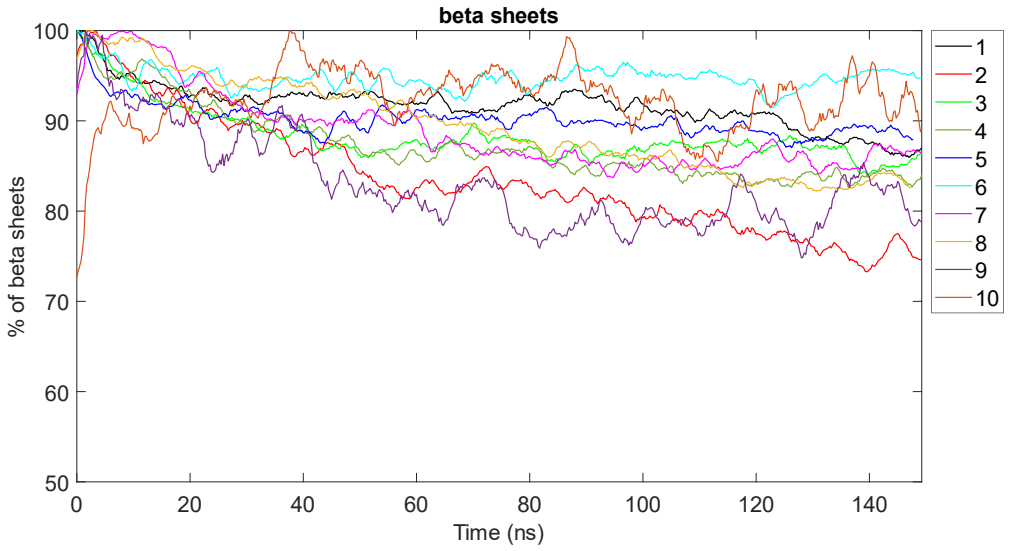

Figure 52 beta sheets percentage in time for all configurations of peptide 2

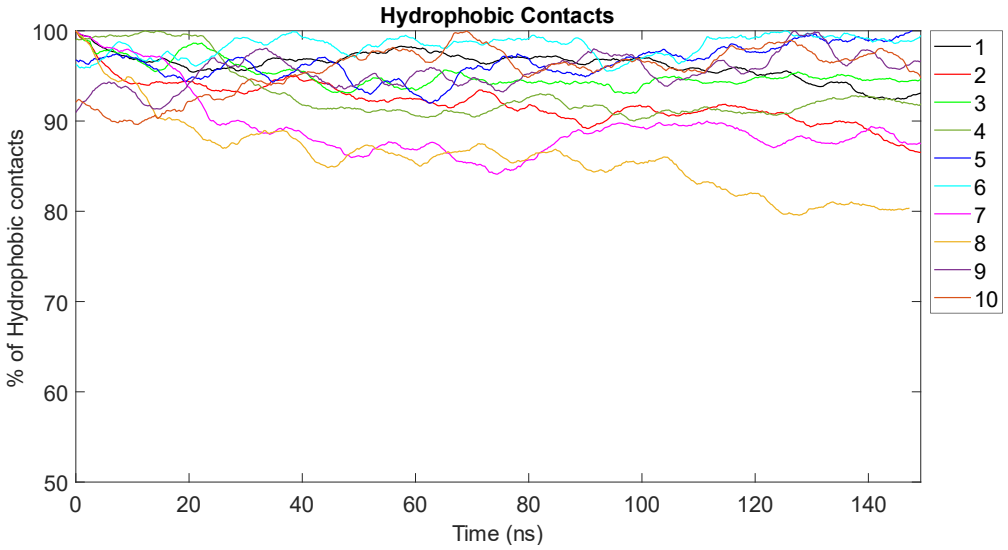

Figure 53 hydrophobic contacts percentage in time for all configurations of peptide 2

5.2 PEPTIDE NUMBER 3

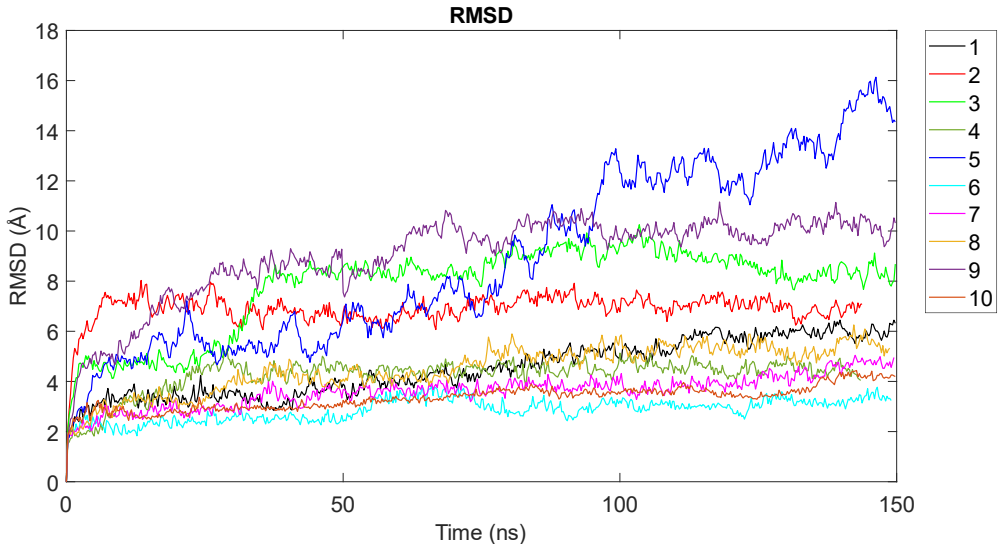

Figure 54 RMSD for all configurations of peptide 3

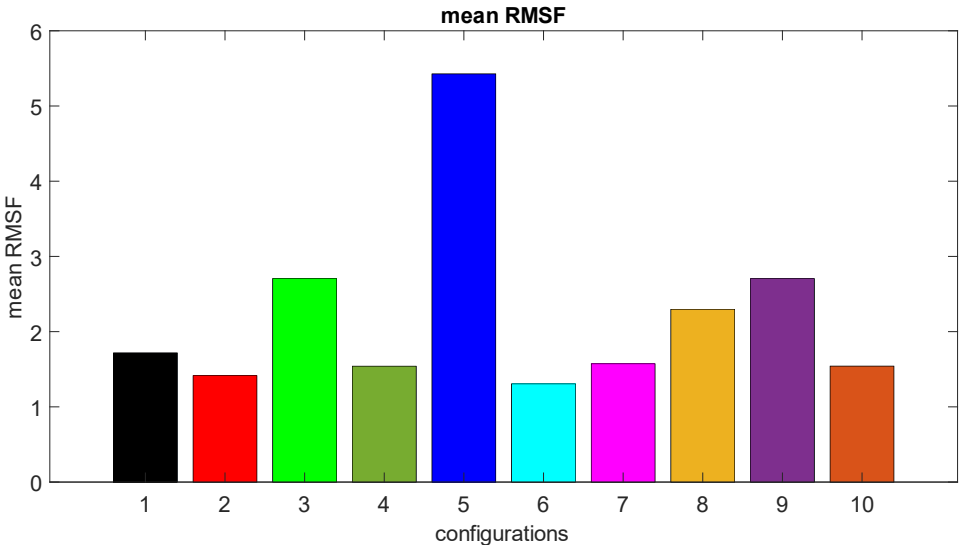

Figure 55 mean RMSF for all configurations of peptide 3

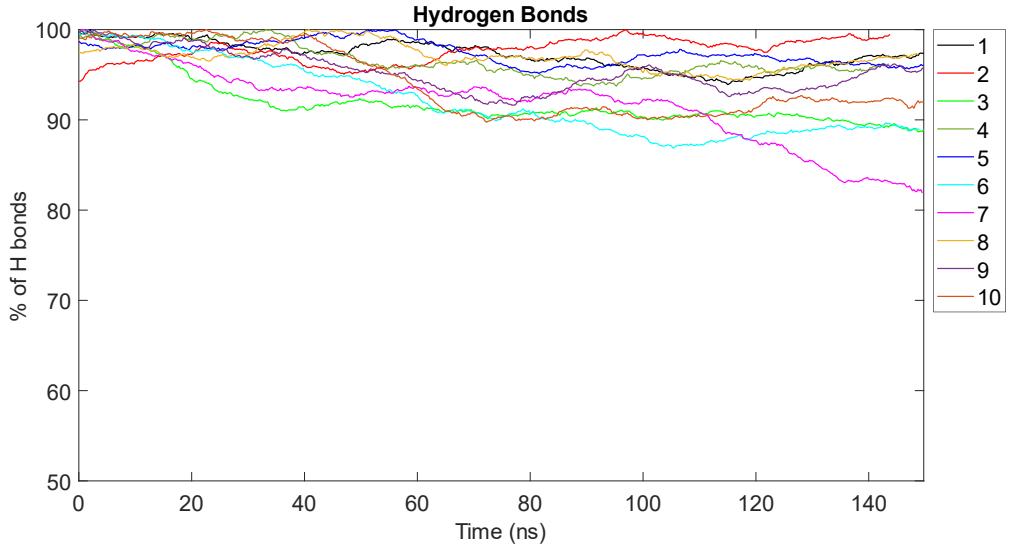

Figure 56 hydrogen bonds percentage in time for all configurations of peptide 3

Configuration	Stability Index	Ranking
1	63.1476	Configuration 5
2	63.9228	Configuration 2
3	54.9460	Configuration 1
4	53.5393	Configuration 6
5	69.2784	Configuration 8
6	61.3528	Configuration 3
7	53.9934	Configuration 7
8	56.9273	Configuration 4
9	44.2399	Configuration 10
10	52.0220	Configuration 9

Table 8 Stability Index for peptide 3 and configurations ranking from most to least stable

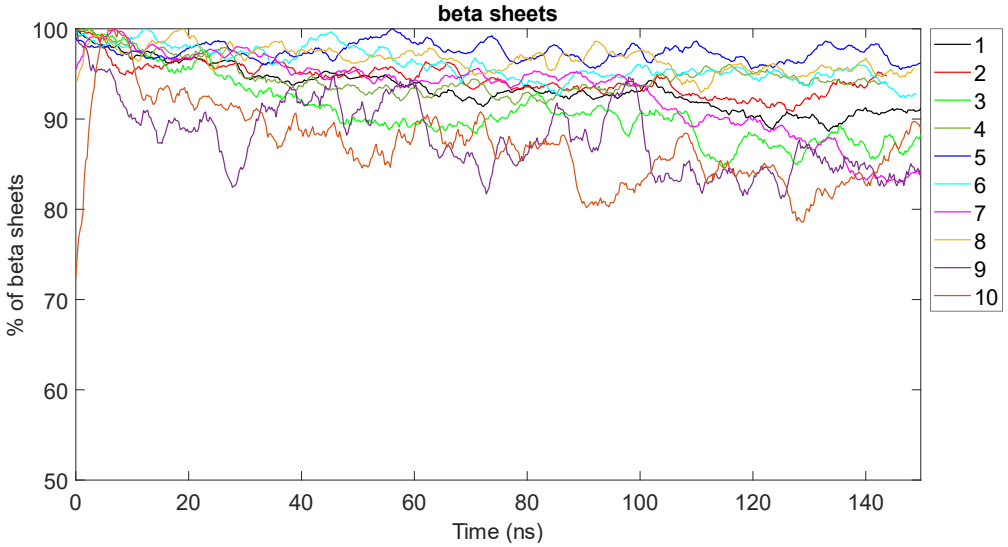

Figure 57 beta sheets percentage in time for all configurations of peptide 3

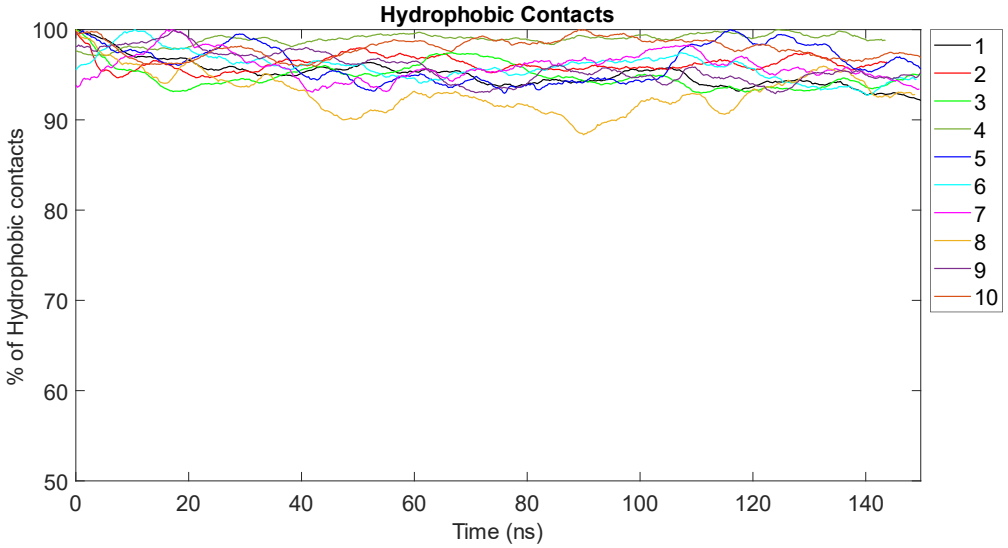

Figure 58 hydrophobic contacts percentage in time for all configurations of peptide 3

5.3 PEPTIDE NUMBER 4

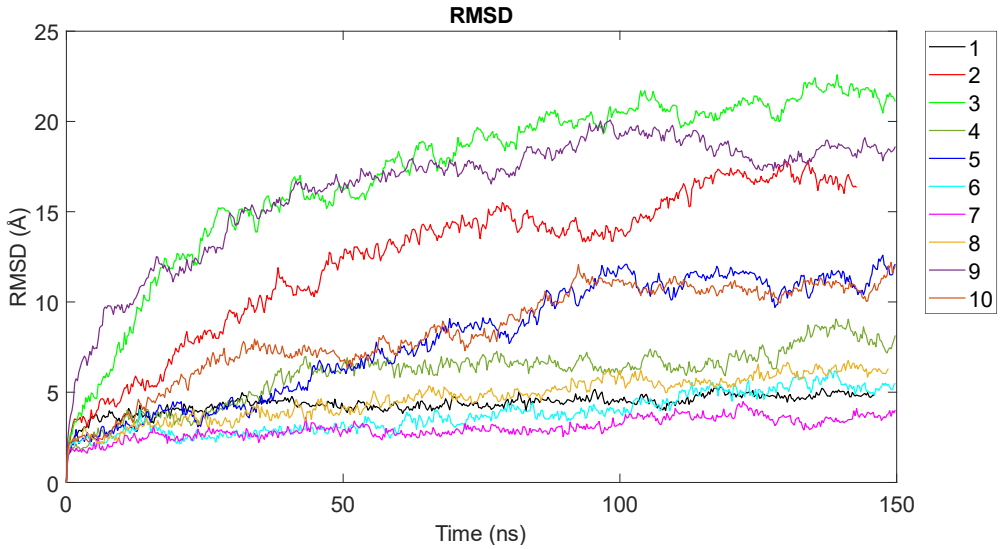

Figure 59 RMSD for all configurations of peptide 4

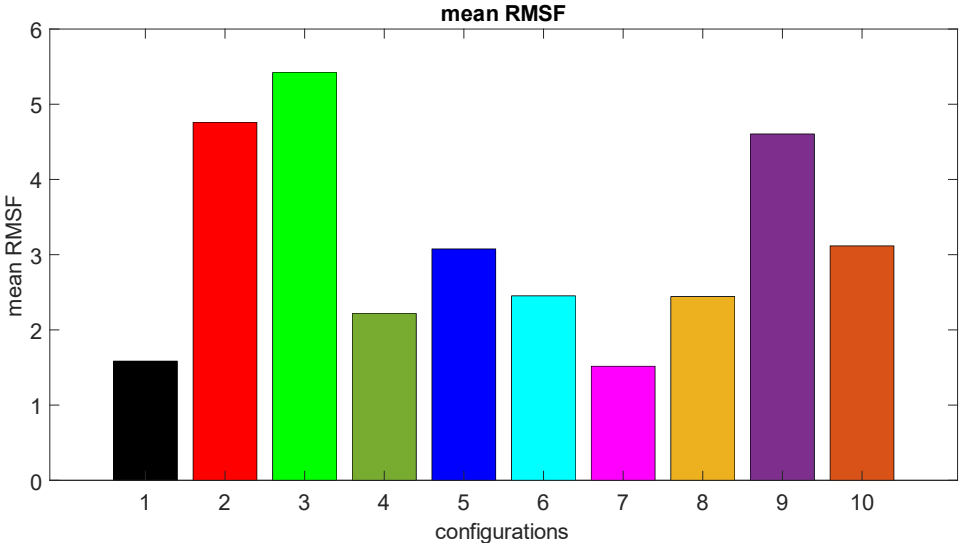

Figure 60 mean RMSF for all configurations of peptide 4

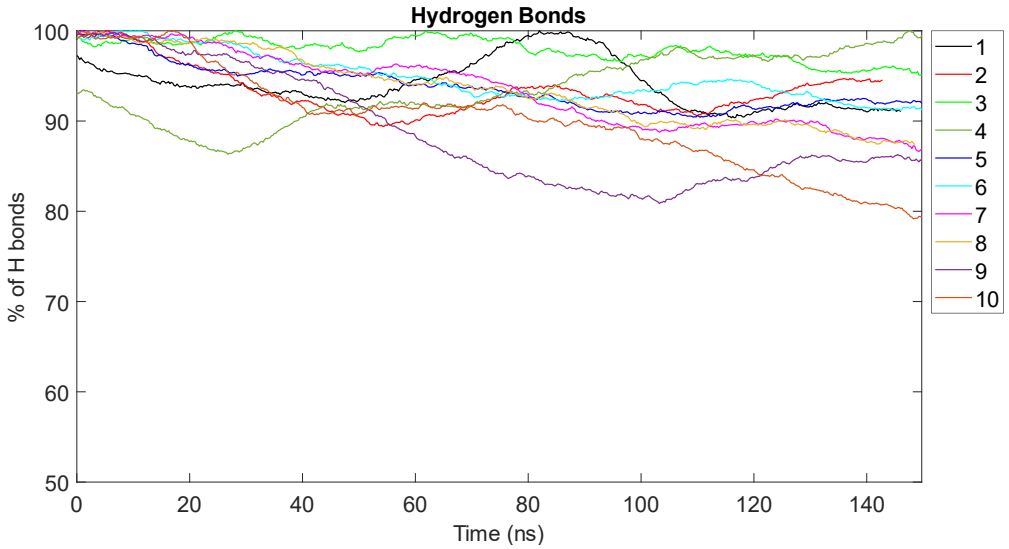

Figure 61 hydrogen bonds percentage in time for all configurations of peptide 4

Configuration	Stability Index	Ranking
1	52.3078	Configuration 5
2	52.7032	Configuration 6
3	55.0272	Configuration 7
4	50.5693	Configuration 8
5	67.0439	Configuration 3
6	65.9527	Configuration 2
7	61.0685	Configuration 1
8	57.8580	Configuration 4
9	44.1361	Configuration 10
10	48.3405	Configuration 9

Table 9 Stability Index for peptide 4 and configurations ranking from most to least stable

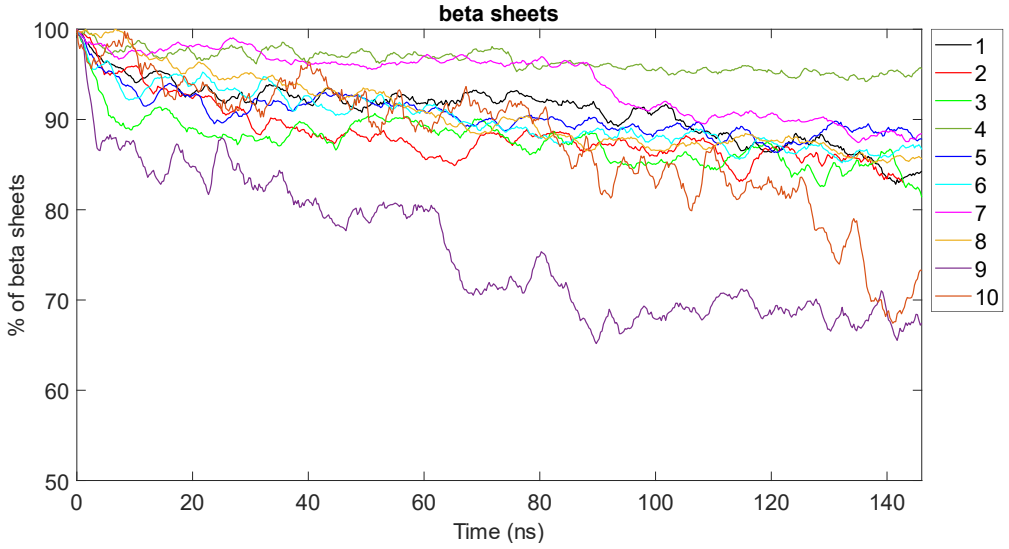

Figure 62 beta sheets percentage in time for all configurations of peptide 4

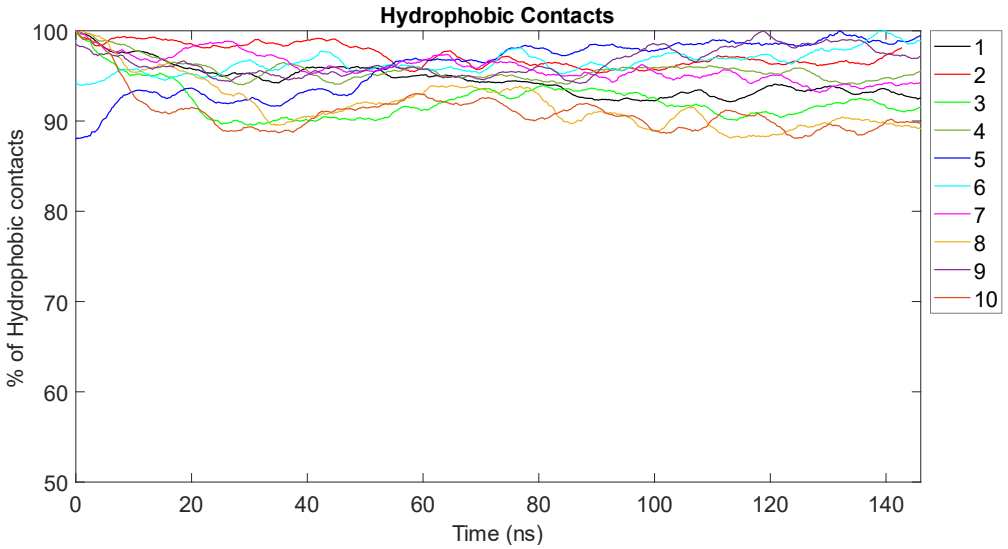

Figure 63 hydrophobic contacts percentage in time for all configurations of peptide 4

5.4 PEPTIDE NUMBER 5

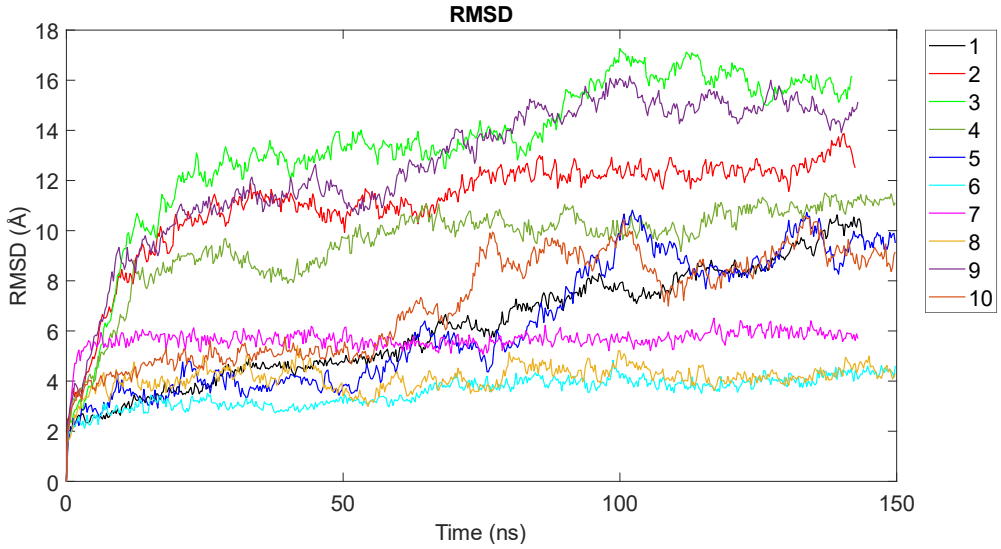

Figure 64 RMSD for all configurations of peptide 5

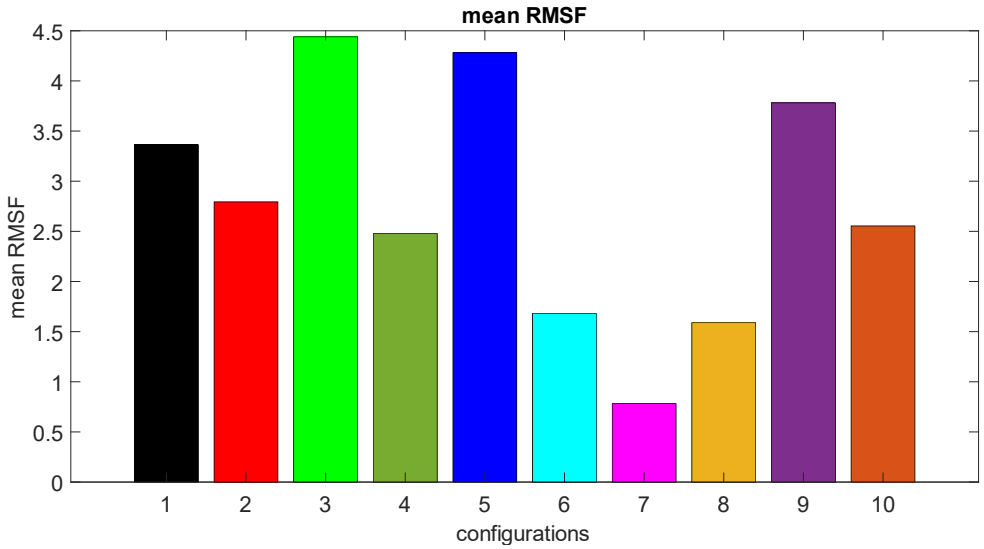

Figure 65 mean RMSF for all configurations of peptide 5

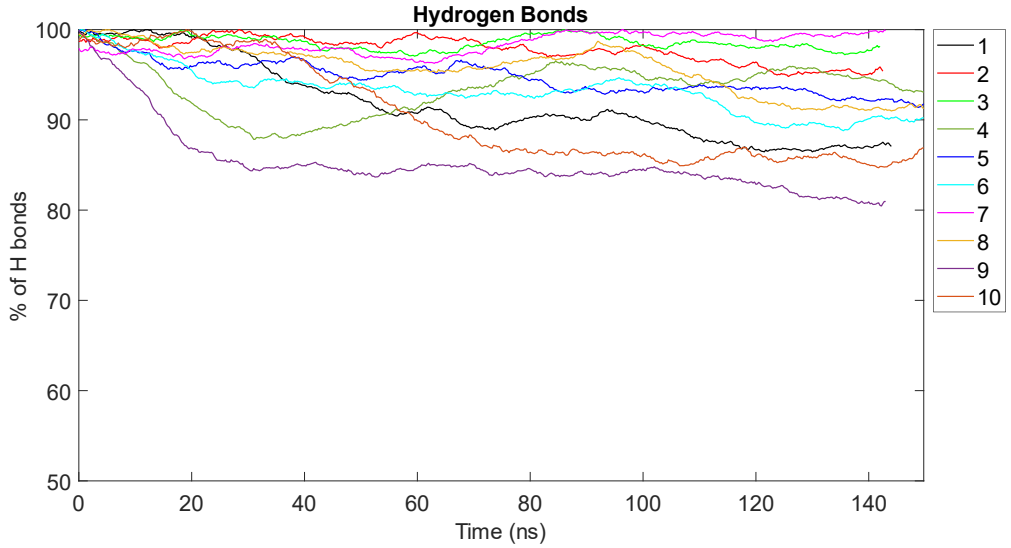

Figure 66 hydrogen bonds percentage in time for all configurations of peptide 5

Configuration	Stability Index	Ranking
1	56.2535	Configuration 7
2	63.1696	Configuration 8
3	62.3186	Configuration 2
4	56.9943	Configuration 3
5	59.9531	Configuration 6
6	61.4975	Configuration 5
7	74.1158	Configuration 4
8	67.5397	Configuration 1
9	40.8105	Configuration 9
10	35.0471	Configuration 10

Table 10 Stability Index for peptide 5 and configurations ranking from most to least stable

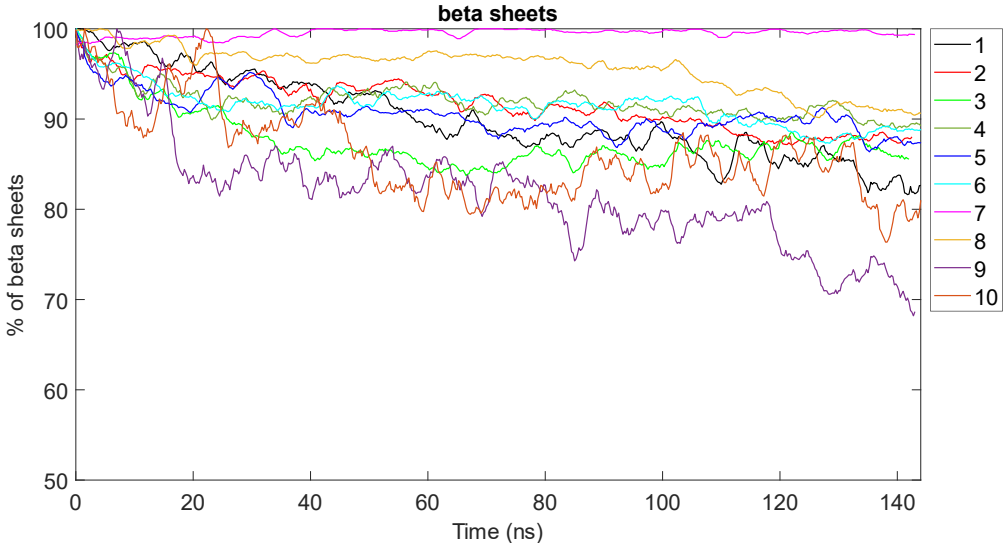

Figure 67 beta sheets percentage in time for all configurations of peptide 5

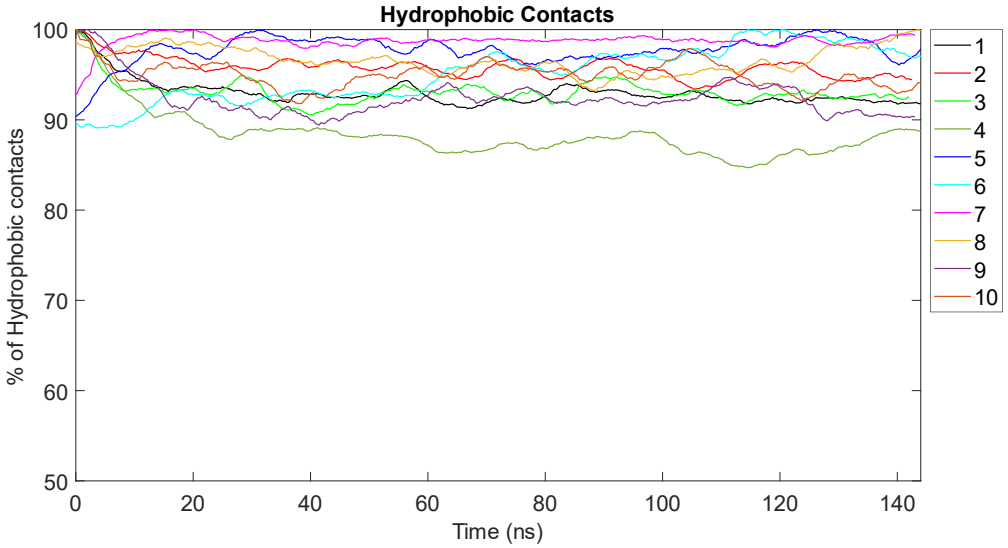

Figure 68 hydrophobic contacts percentage in time for all configurations of peptide 5

5.5 PEPTIDE NUMBER 6

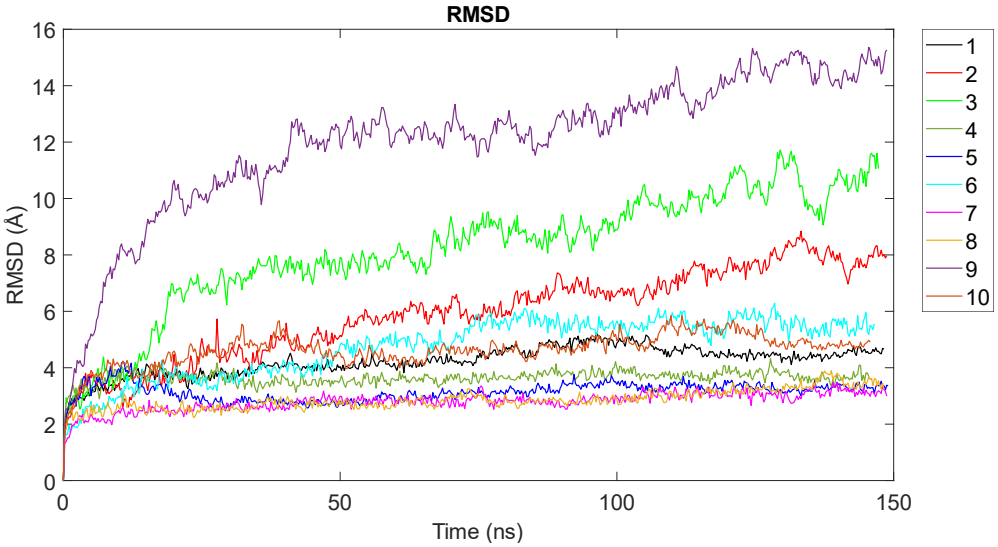

Figure 69 RMSD for all configurations of peptide 6

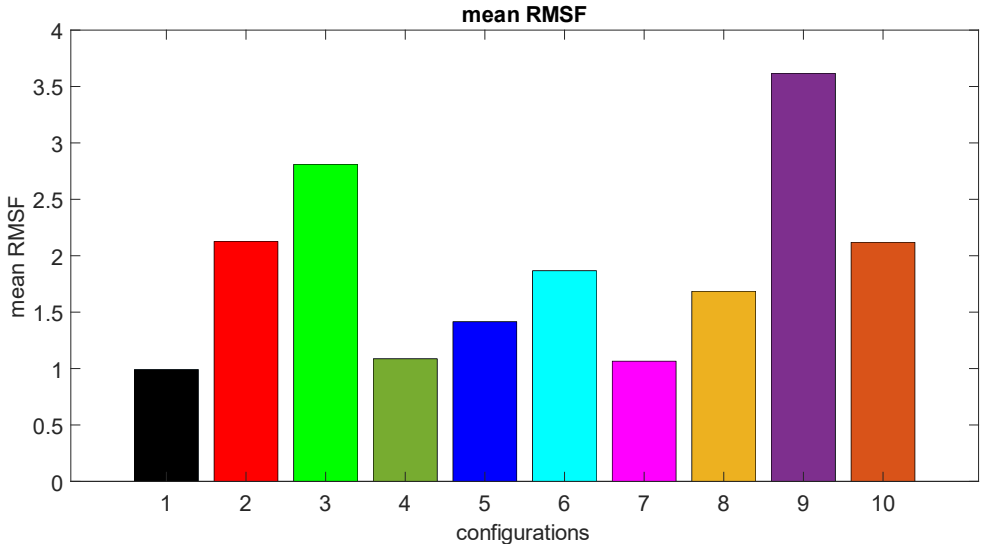

Figure 70 mean RMSF for all configurations of peptide 6

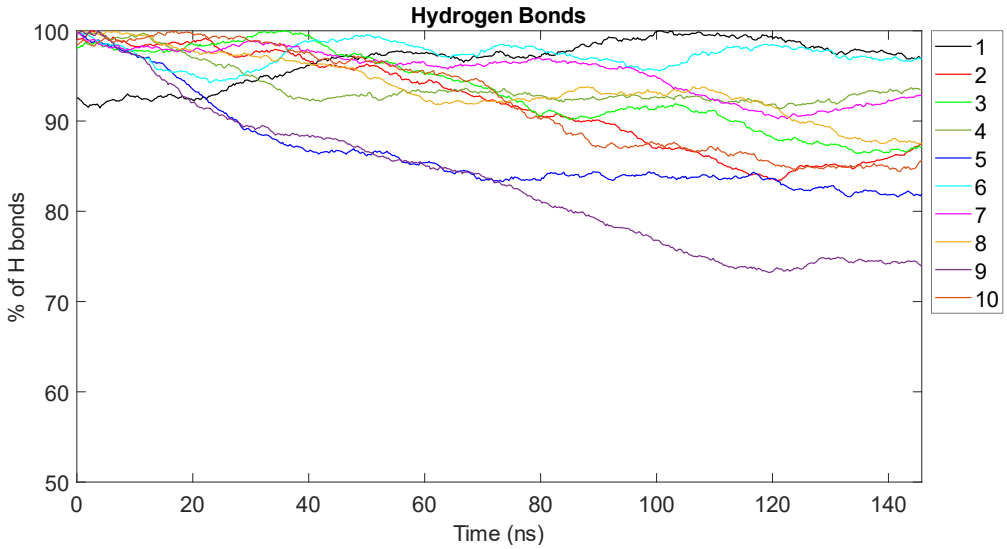

Figure 71 hydrogen bonds percentage in time for all configurations of peptide 6

Configuration	Stability Index	Ranking
1	53.5352	Configuration 6
2	52.2163	Configuration 7
3	51.6036	Configuration 8
4	54.5087	Configuration 4
5	52.1659	Configuration 1
6	64.3196	Configuration 2
7	62.9530	Configuration 5
8	60.6497	Configuration 3
9	39.1736	Configuration 9
10	39.0603	Configuration 10

Table 11 Stability Index for peptide 6 and configurations ranking from most to least stable

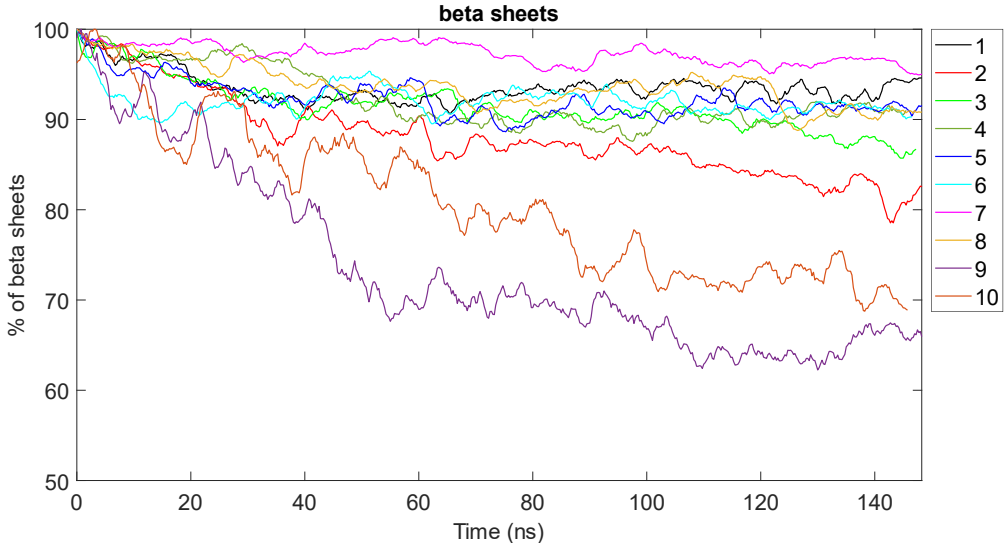

Figure 72 beta sheets percentage in time for all configurations of peptide 6

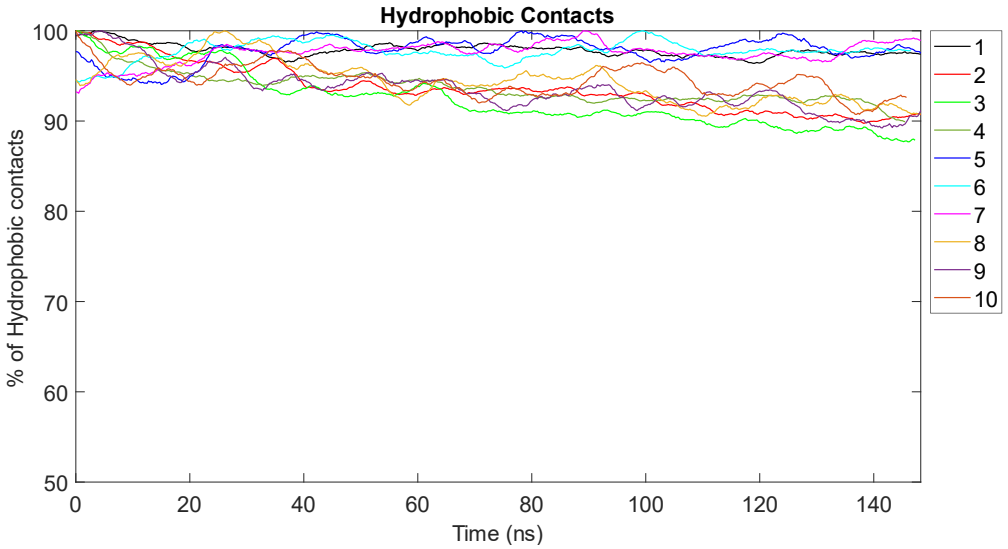

Figure 73 hydrophobic contacts percentage in time for all configurations of peptide 6

5.6 PEPTIDE NUMBER 7

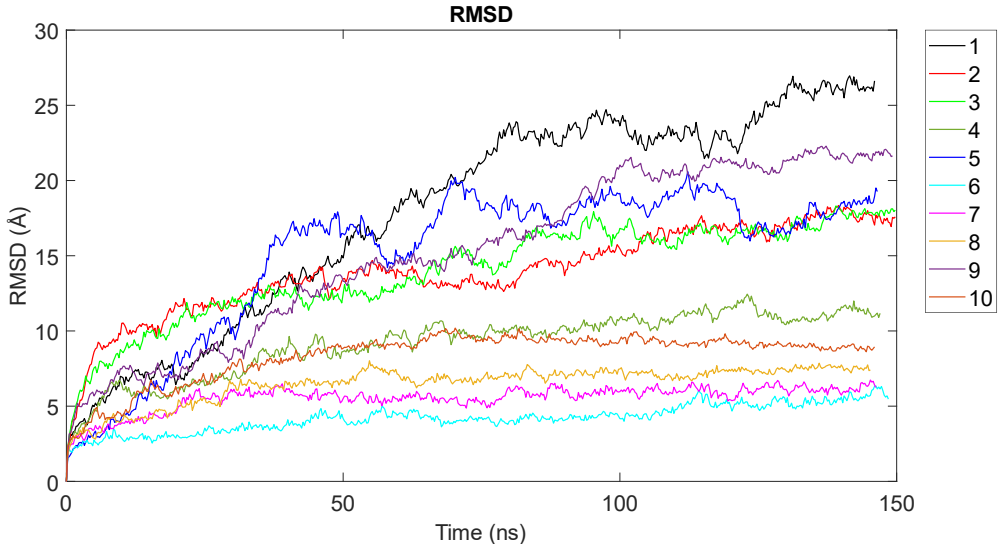

Figure 74 RMSD for all configurations of peptide 7

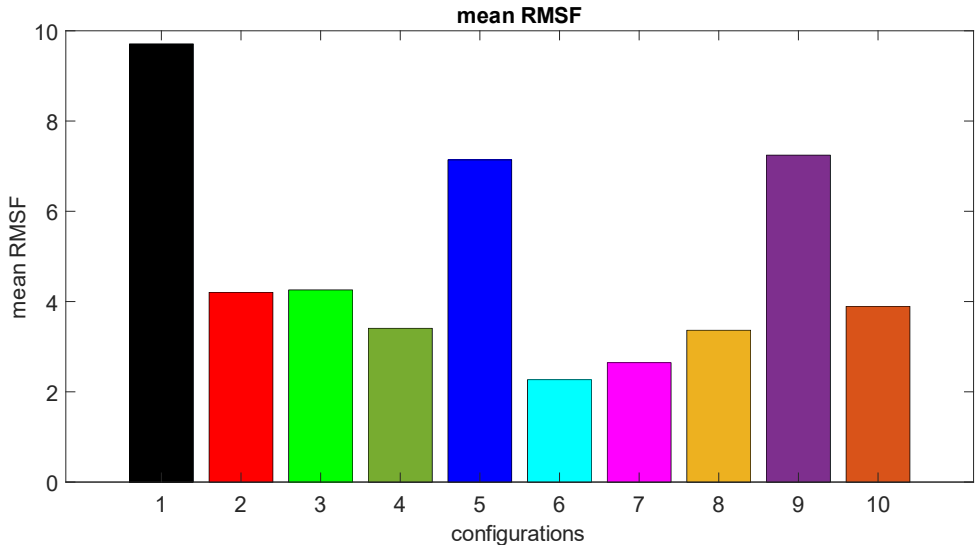

Figure 75 mean RMSF for all configurations of peptide 7

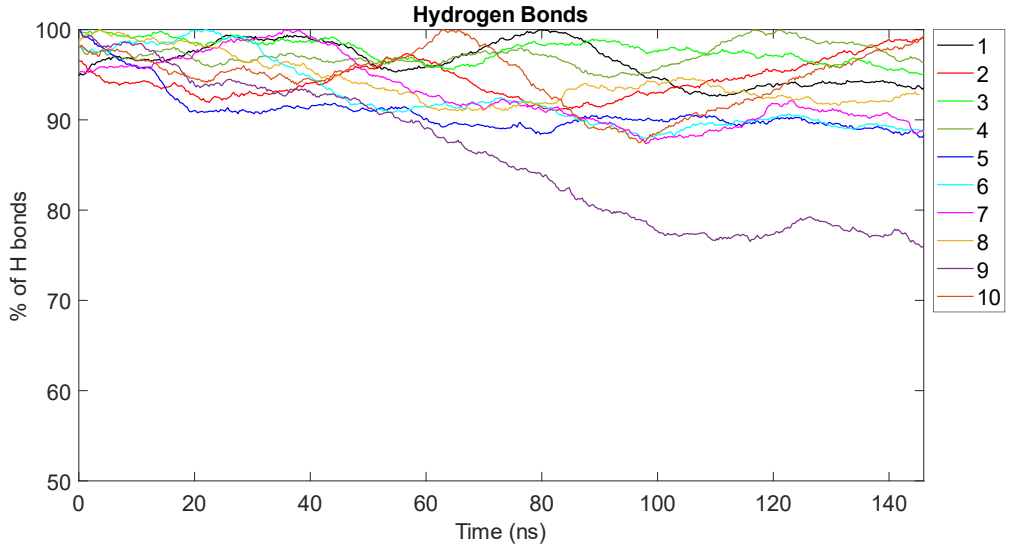

Figure 76 hydrogen bonds percentage in time for all configurations of peptide 7

Configuration	Stability Index	Ranking
1	41.6788	Configuration 3
2	42.6977	Configuration 6
3	44.1487	Configuration 4
4	43.2997	Configuration 8
5	42.3465	Configuration 2
6	43.3927	Configuration 5
7	39.7150	Configuration 1
8	42.8358	Configuration 7
9	23.1947	Configuration 10
10	26.4830	Configuration 9

Table 12 Stability Index for peptide 7 and configurations ranking from most to least stable

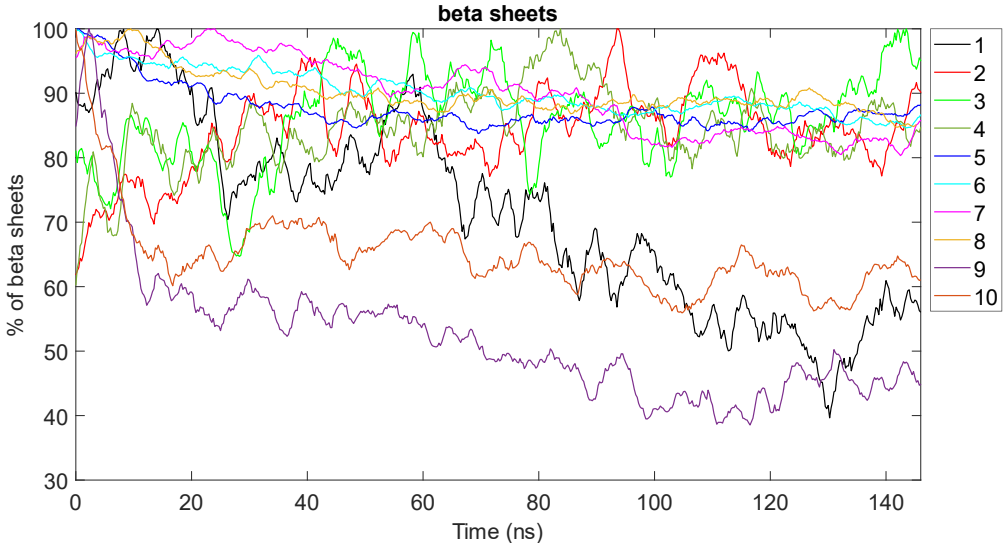

Figure 77 beta sheets percentage in time for all configurations of peptide 7

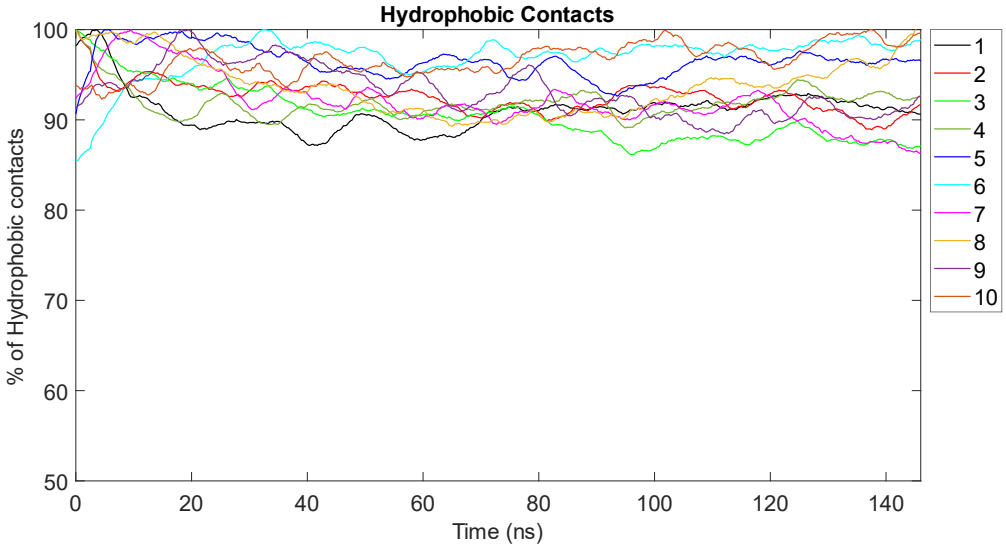

Figure 78 hydrophobic contacts percentage in time for all configurations of peptide 7

5.7 MATLAB SCRIPTS

5.7.1 RMSD plots

```
%RMSD plots
clear all
close all
clc
load rmsd_1.dat
%x1: array of frames
x1=rmsd_1(:,1);
%n1: number of frames
n_1=length(x1);
%identification of the last frame of simulation
last_frame=max(rmsd_1(:,1));
%last time of simulation calculated though the proportion (50 ns correspond
%to 2000 frames as the last time corresponds to the last frame just obtained)
last_time=50*last_frame/2000;
%the x axis is an array of lenght equal to the number of frames with values
%from 0 ns to the last time of simulation
xaxis_1=linspace(0,last_time,n_1);
%the y axis cointains rmsd values corresponding to the relative time points
%of the x axis
y1=rmsd_1(:,2);
figure('DefaultAxesFontSize',18)
%rmsd plot as a function of time
plot(xaxis_1,y1,'-k', 'Linewidth',1)
hold on

load rmsd_2.dat
x2=rmsd_2(:,1);
n_2=length(x2);
last_frame=max(rmsd_2(:,1));
last_time=50*last_frame/2000;
xaxis_2=linspace(0,last_time,n_2);
y2=rmsd_2(:,2);
plot(xaxis_2,y2,'-r', 'Linewidth',1)

load rmsd_3.dat
x3=rmsd_3(:,1);
n_3=length(x3);
last_frame=max(rmsd_3(:,1));
last_time=50*last_frame/2000;
xaxis_3=linspace(0,last_time,n_3);
y3=rmsd_3(:,2);
plot(xaxis_3,y3,'-g', 'Linewidth',1)

load rmsd_4.dat
x4=rmsd_4(:,1);
n_4=length(x4);
last_frame=max(rmsd_4(:,1));
last_time=50*last_frame/2000;
```

```
xaxis_4=linspace(0,last_time,n_4);
y4=rmsd_4(:,2);
plot(xaxis_4,y4,'-','color','#77AC30', 'Linewidth',1)

load rmsd_5.dat
x5=rmsd_5(:,1);
n_5=length(x5);
last_frame=max(rmsd_5(:,1));
last_time=50*last_frame/2000;
xaxis_5=linspace(0,last_time,n_5);
y5=rmsd_5(:,2);
plot(xaxis_5,y5,'-b', 'Linewidth',1)

load rmsd_6.dat
x6=rmsd_6(:,1);
n_6=length(x6);
last_frame=max(rmsd_6(:,1));
last_time=50*last_frame/2000;
xaxis_6=linspace(0,last_time,n_6);
y6=rmsd_6(:,2);
plot(xaxis_6,y6,'-c', 'Linewidth',1)

load rmsd_7.dat
x7=rmsd_7(:,1);
n_7=length(x7);
last_frame=max(rmsd_7(:,1));
last_time=50*last_frame/2000;
xaxis_7=linspace(0,last_time,n_7);
y7=rmsd_7(:,2);
plot(xaxis_7,y7,'-m', 'Linewidth',1)

load rmsd_8.dat
x8=rmsd_8(:,1);
n_8=length(x8);
last_frame=max(rmsd_8(:,1));
last_time=50*last_frame/2000;
xaxis_8=linspace(0,last_time,n_8);
y8=rmsd_8(:,2);
plot(xaxis_8,y8,'-', 'color', '#EDB120', 'Linewidth',1)

load rmsd_9.dat
x9=rmsd_9(:,1);
n_9=length(x9);
last_frame=max(rmsd_9(:,1));
last_time=50*last_frame/2000;
xaxis_9=linspace(0,last_time,n_9);
y9=rmsd_9(:,2);
plot(xaxis_9,y9,'-','color' ,'#7E2F8E', 'Linewidth',1)

load rmsd_10.dat
x10=rmsd_10(:,1);
n_10=length(x10);
last_frame=max(rmsd_10(:,1));
```

```
last_time=50*last_frame/2000;
xaxis_10=linspace(0,last_time,n_10);
y10=rmsd_10(:,2);
plot(xaxis_10,y10,'-','color' ,'#D95319', 'Linewidth',1)

%title and legend
xlabel ('Time (ns)', 'FontSize', 18)
ylabel('RMSD (Å)', 'FontSize', 18)
title('RMSD', 'FontSize', 18)
legend ({'1','2','3', '4','5','6','7','8','9','10'}, 'Location', 'Northeast Outside', 'Fontsize', 18)
```

5.7.2 RMSF plots

```
%RMSF: bar diagram of mean RMSF for all 10 configurations
clear
close all
clc
load rmsf_1.dat
%vector containing rmsf values for different amino acids
vect1 = rmsf_1;
%mean rmsf
mean1 = mean (vect1);
%process iterated for all the configurations
load rmsf_2.dat
vect2 = rmsf_2;
mean2 = mean (vect2);
load rmsf_3.dat
vect3 = rmsf_3;
mean3 = mean (vect3);
load rmsf_4.dat
vect4 = rmsf_4;
mean4 = mean (vect4);
load rmsf_5.dat
vect5 = rmsf_5;
mean5 = mean (vect5);
load rmsf_6.dat
vect6 = rmsf_6;
mean6 = mean (vect6);
load rmsf_7.dat
vect7 = rmsf_7;
mean7 = mean (vect7);
load rmsf_8.dat
vect8 = rmsf_8;
mean8 = mean (vect8);
load rmsf_9.dat
vect9 = rmsf_9;
mean9 = mean (vect9);
load rmsf_10.dat
vect10 = rmsf_10;
mean10 = mean (vect10);
%data: vector of mean values
data= [mean1, mean2, mean3, mean4, mean5, mean6, mean7, mean8, mean9, mean10];
```

```
x=[1 2 3 4 5 6 7 8 9 10];
%bar plot having configuration numbers on the x axis and respective
%mean rmsf value on the y axis
figure('DefaultAxesFontSize',18)
bar(x,data)
hold on
%colors assignment, labels and title
b=bar(1,data(1))
b.FaceColor = 'k'
b=bar(2,data(2))
b.FaceColor = 'r'
b=bar(3,data(3))
b.FaceColor = 'g'
b=bar(4,data(4))
b.FaceColor = '#77AC30'
b=bar(5,data(5))
b.FaceColor = 'b'
b=bar(6,data(6))
b.FaceColor = 'c'
b=bar(7,data(7))
b.FaceColor = 'm'
b=bar(8,data(8))
b.FaceColor = '#EDB120'
b=bar(9,data(9))
b.FaceColor = '#7E2F8E'
b=bar(10,data(10))
b.FaceColor = '#D95319'
xlabel('configurations', 'FontSize', 18)
ylabel('mean RMSF', 'FontSize', 18)
title ('mean RMSF', 'FontSize', 18)
```

5.7.3 Beta sheets, hydrogen bonds, hydrophobic contacts

The scripts used for the calculation of beta sheets, hydrogen bonds and hydrophobic contacts follow the same scheme. The difference is the optimum number of frames for the mobile mean, which is found to be 10, 100 and 20 respectively.

```
%beta sheets: mobile mean on 10 frames (0,25 ns) of the number of beta sheets,
%expressed in percentage.
load beta_1.dat
%x: vector of frames
x = beta_1 (:,1);
n=length(x);
last_frame=max(x);
%conversion from frames to time (ns)
last_time=50*last_frame/2000;
%xaxis: vector of times
xaxis=linspace(0,last_time,n);
%values: beta sheets absolute value
values = beta_1 (:,2);
%mobile mean on 10 frames (100 in case of Hbonds, 20 in case of HC)
media_mobile_10frame=movmean(values, 10);
```

```
%beta sheets expressed in percentage
maximum_value=max(media_mobile_10frame);
y=media_mobile_10frame./maximum_value;
perc=y.*100;
%plot of beta sheets mobile mean in percentage vs time
figure('DefaultAxesFontSize',18)
plot(xaxis,perc,'-k', 'Linewidth', 1)
hold on
xlabel('Time (ns)', 'FontSize', 18)
ylabel('% of beta sheets', 'FontSize', 18)
title ('beta sheets', 'FontSize', 18)
%axis range
axis ([min(xaxis) max(xaxis) 50 100])

load  beta_2.dat
x = beta_2 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_2 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-r','Linewidth', 1)

load  beta_3.dat
x = beta_3 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_3 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-g','Linewidth', 1)

load  beta_4.dat
x = beta_4 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_4 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-','color','#77AC30','Linewidth', 1)
```

```
load  beta_5.dat
x = beta_5 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_5 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-b','Linewidth', 1)

load  beta_6.dat
x = beta_6 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_6 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-c','Linewidth', 1)

load  beta_7.dat
x = beta_7 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_7 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-m','Linewidth', 1)

load  beta_8.dat
x = beta_8 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_8 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-', 'color', '#EDB120','Linewidth', 1)

load  beta_9.dat
```

```
x = beta_9 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_9 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-', 'color' ,'#7E2F8E','Linewidth', 1)

load  beta_10.dat
x = beta_10 (:,1);
n=length(x);
last_frame=max(x);
last_time=50*last_frame/2000;
xaxis=linspace(0,last_time,n);
values = beta_10 (:,2);
media_mobile_100frame=movmean(values, 10);
maximum_value=max(media_mobile_100frame);
y=media_mobile_100frame./maximum_value;
perc=y.*100;
plot(xaxis,perc,'-','color' ,'#D95319','Linewidth', 1)
legend ({'1','2','3', '4','5','6','7','8','9','10'}, 'Location', 'Northeast Outside', 'Fontsize', 18)
```

5.7.4 Native contacts

```
%Native Contacts data are imported as a matrix from a .txt file
%frames are in the first column from row 1 to 1000
frames=contacts1(1:1000,1);
n=length(frames);
%to obtain the real number of frames this number is multiplied by 5
%because the analysis was carried out on 1 frame every 5 frames
real_frames=frames.*5;
%conversion from frames to time (50 ns correspond to 2000 frames)
t_max=real_frames(end)*50/2000;
t_min=0;
%construction of x axis containing time points in ns
times=linspace(t_min,t_max, n);
%native contacts are in the third column of the matrix from row 1 to 1000
nc=contacts1(1:1000,3);
%mobile mean on 100 frames
nc=movmean(nc, 100);
figure
plot (times, nc, '-k', 'Linewidth', 1')
hold on

%the same steps are repeated for all the other configurations
frames=contacts2(1:1000,1);
n=length(frames);
real_frames=frames.*5;
```

```
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts2(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-r','Linewidth', 1)

frames=contacts3(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts3(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-g','Linewidth', 1)

frames=contacts4(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts4(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-','color','#77AC30','Linewidth', 1)

frames=contacts5(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts5(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-b','Linewidth', 1)

frames=contacts6(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts6(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-c','Linewidth', 1)

frames=contacts7(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
```

```
nc=contacts7(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-m','Linewidth', 1)

frames=contacts8(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts8(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-','color', '#EDB120','Linewidth', 1)

frames=contacts9(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts9(1:1000,3);
nc=movmean(nc, 100);
plot (times, nc, '-', 'color','#7E2F8E','Linewidth', 1)

frames=contacts10(1:1000,1);
n=length(frames);
real_frames=frames.*5;
t_max=real_frames(end)*50/2000;
t_min=0;
times=linspace(t_min,t_max, n);
nc=contacts10(1:1000,3);
nc=movmean(nc, 100);
figure('DefaultAxesFontSize',18)
plot (times, nc, '-','color','#D95319','Linewidth', 1)
xlabel('Time (ns)' ,'FontSize', 18)
ylabel('% of Native Contacts','FontSize', 18)
title ('Native Contacts', 'FontSize', 18)
legend ({'1','2','3', '4','5','6','7','8','9','10'}, 'Location', 'Northeast Outside', 'Fontsize', 18)
```

5.7.5 Stability Index

```
clear all
close all
clc
load rmsf_1.dat
load hbonds_1.dat
%mean_RMSF is the mean value on the residues
mean_RMSF=mean(rmsf_1);
%mean_hbonds is the mean hydrogen bonds on frames (or simulation times)
%excluding the first 2000 frames (corresponding to 50 ns), where frame 2000
%is at the row 201 of the vector h_bonds_1
mean_hbonds=mean(hbonds_1(201:end,2));
```

```
Stability_Index_1 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_2.dat
load hbonds_2.dat
mean_RMSF=mean(rmsf_2);
mean_hbonds=mean(hbonds_2(201:end,2));
Stability_Index_2 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_3.dat
load hbonds_3.dat
mean_RMSF=mean(rmsf_3);
mean_hbonds=mean(hbonds_3(201:end,2));
Stability_Index_3 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_4.dat
load hbonds_4.dat
mean_RMSF=mean(rmsf_4);
mean_hbonds=mean(hbonds_4(201:end,2));
Stability_Index_4 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_5.dat
load hbonds_5.dat
mean_RMSF=mean(rmsf_5);
mean_hbonds=mean(hbonds_5(201:end,2));
Stability_Index_5 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_6.dat
load hbonds_6.dat
mean_RMSF=mean(rmsf_6);
mean_hbonds=mean(hbonds_6(201:end,2));
Stability_Index_6 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_7.dat
load hbonds_7.dat
mean_RMSF=mean(rmsf_7);
mean_hbonds=mean(hbonds_7(201:end,2));
Stability_Index_7 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_8.dat
load hbonds_8.dat
mean_RMSF=mean(rmsf_8);
mean_hbonds=mean(hbonds_8(201:end,2));
Stability_Index_8 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_9.dat
load hbonds_9.dat
mean_RMSF=mean(rmsf_9);
mean_hbonds=mean(hbonds_9(201:end,2));
Stability_Index_9 = 0.5*(1/mean_RMSF+mean_hbonds);

load rmsf_10.dat
load hbonds_10.dat
mean_RMSF=mean(rmsf_10);
```

```
mean_hbonds=mean(hbonds_10(201:end,2));
Stability_Index_10 = 0.5*(1/mean_RMSF+mean_hbonds);

Stability_Index = [
    Stability_Index_1
    Stability_Index_2
    Stability_Index_3
    Stability_Index_4
    Stability_Index_5
    Stability_Index_6
    Stability_Index_7
    Stability_Index_8
    Stability_Index_9
    Stability_Index_10]

Ranking= sort (Stability_Index, 'descend')
```

BIBLIOGRAPHY AND SITOGRAPHY

1. Savyasachi, A. J. *et al.* Supramolecular Chemistry: A Toolkit for Soft Functional Materials and Organic Particles. *Chem* **3**, 764–811 (2017).
2. Zhou, J., Li, J., Du, X. & Xu, B. Supramolecular biofunctional materials. *Biomaterials* **129**, 1–27 (2017).
3. Steed, J. W., Turner, D. R. & Wallace, K. *Core Concepts in Supramolecular Chemistry and Nanochemistry*. (2007).
4. Zhang, S. Discovery and design of self-assembling peptides. *Interface Focus* **7**, (2017).
5. Beesley, J. L. & Woolfson, D. N. The de novo design of α-helical peptides for supramolecular self-assembly. *Curr. Opin. Biotechnol.* **58**, 175–182 (2019).
6. Panda, J. J. & Chauhan, V. S. Short peptide based self-assembled nanostructures: Implications in drug delivery and tissue engineering. *Polym. Chem.* **5**, 4418–4436 (2014).
7. Zhou, P. *et al.* Amino acid conformations control the morphological and chiral features of the self-assembled peptide nanostructures: Young investigators perspective. *J. Colloid Interface Sci.* **548**, 244–254 (2019).
8. Berg, J., Tymoczko, J. & Stryer, L. Chemical Bonds in Biochemistry. in *Biochemistry* 8–14 (2002).
9. Zhang, N., Ruan, J., Duan, G., Gao, S. & Zhang, T. The interstrand amino acid pairs play a significant role in determining the parallel or antiparallel orientation of β-strands. *Biochem. Biophys. Res. Commun.* **386**, 537–543 (2009).
10. Bosshard, H. R., Marti, D. N. & Jelesarov, I. Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *J. Mol. Recognit.* **17**, 1–16 (2004).
11. Kumar, S. & Nussinov, R. Close-range electrostatic interactions in proteins. *ChemBioChem* **3**, 604–617 (2002).
12. Pylaeva, S., Brehm, M. & Sebastiani, D. Salt Bridge in Aqueous Solution: Strong Structural Motifs but Weak Enthalpic Effect. *Sci. Rep.* **8**, 1–7 (2018).
13. Singh, P. *et al.* Self-assembling behaviour of a modified aromatic amino acid in competitive medium. *Soft Matter* **16**, 6599–6607 (2020).
14. Erica L. Bakota, O. S. Self-Assembling Multidomain Peptide Fibers with Aromatic Cores. *Biomacromolecules., 2013 may 13* **14**, 1370–1378 (2013).
15. Zhuang, W. R. *et al.* Applications of π-π stacking interactions in the design of drug-delivery systems. *J. Control. Release* **294**, 311–326 (2019).
16. Gautieri, A. *et al.* Molecular dynamics investigation of halogenated amyloidogenic

peptides. *J. Mol. Model.* **25**, (2019).

17. Ignacio Soteras Gutiérrez, Fang-Yu Lin, Kenno Vanommeslaeghe, Justin A. Lemkul, Kira A. Armacost, Charles L. Brooks III, and A. D. M. J. Parametrization of Halogen Bonds in the CHARMM General Force Field: Improved treatment of ligand-protein interactions. *Bioorg Med Chem. 2016 Oct.* (2016) doi:10.1016/j.bmc.2016.06.034.Parametrization.

18. Allouche, A. Software News and Updates Gabedit — A Graphical User Interface for Computational Chemistry Softwares. *J. Comput. Chem.* **32**, 174–182 (2012).

19. Huxford, T. X-Ray Crystallography. *Brenner's Encycl. Genet. Second Ed.* 366–368 (2013) doi:10.1016/B978-0-12-374984-0.01657-0.

20. Wider, G. Structure determination of biological macromolecules in solution using nuclear magnetic resonance spectroscopy. *Biotechniques* **29**, 1278–1294 (2000).

21. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **1**, 2876–2890 (2007).

22. Gore, [4]R.F. FischettiD.J. RodiD.B. Wide-Angle X-Ray Solution Scatteringas a Probe of Ligand-InducedConformational Changes in Proteins. *Chem. Biol.* **11**, 431–1443 (2004).

23. Frederix, P. W. J. M., Patmanidis, I. & Marrink, S. J. Molecular simulations of self-assembling bio-inspired supramolecular systems and their connection to experiments. *Chem. Soc. Rev.* **47**, 3470–3489 (2018).

24. Frederix, P. W. J. M. *et al.* Exploring the sequence space for (tri-)peptide self-assembly to design and discover new hydrogels. *Nat. Chem.* **7**, 30–37 (2015).

25. Fu, I. W., Markegard, C. B., Chu, B. K. & Nguyen, H. D. Role of hydrophobicity on self-assembly by peptide amphiphiles via molecular dynamics simulations. *Langmuir* **30**, 7745–7754 (2014).

26. Hu, X. *et al.* Recent advances in short peptide self-assembly: from rational design to novel applications. *Curr. Opin. Colloid Interface Sci.* **45**, 1–13 (2020).

27. Lu, J. & Wang, X. Biomimetic Self-Assembling Peptide Hydrogels for Tissue Engineering Applications. in *Biomimetic Medical Materials: From Nanotechnology to 3D Bioprinting* (ed. Noh, I.) 297–312 (Springer Singapore, 2018). doi:10.1007/978-981-13-0445-3_18.

28. Bellis, S. L. Advantages of RGD peptides for directing cell association with biomaterials. *Biomaterials* **32**, 4205–4210 (2011).

29. Tashiro, K. *et al.* A synthetic peptide containing the IKVAV sequence from the A chain of laminin mediates cell attachment, migration, and neurite outgrowth. *J. Biol. Chem.* **264**, 16174–16182 (1989).

30. Sun, Y. *et al.* Functional Self-Assembling Peptide Nanofiber Hydrogels Designed for Nerve Degeneration. *ACS Appl. Mater. Interfaces* **8**, 2348–2359 (2016).

31. Horii, A., Wang, X., Gelain, F. & Zhang, S. Biological designer self-assembling

peptide nanofiber scaffolds significantly enhance osteoblast proliferation, differentiation and 3-D migration. *PLoS One* **2**, 1–9 (2007).

32. Ravichandran, R., Griffith, M. & Phopase, J. Applications of self-assembling peptide scaffolds in regenerative medicine: The way to the clinic. *J. Mater. Chem. B* **2**, 8466–8478 (2014).

33. Li, Y., Wang, F. & Cui, H. Peptide-based supramolecular hydrogels for delivery of biologics. *Bioeng. Transl. Med.* **1**, 306–322 (2016).

34. McGrath, A. M., Novikova, L. N., Novikov, L. N. & Wiberg, M. BD™ PuraMatrix™ peptide hydrogel seeded with Schwann cells for peripheral nerve regeneration. *Brain Res. Bull.* **83**, 207–213 (2010).

35. Mitchison, T. J. *et al.* Self-Assembly and Mineralization of Peptide-Amphiphile Nanofibers. **294**, 1684–1689 (2001).

36. He, H., Wang, H., Zhou, N., Yang, D. & Xu, B. Branched peptides for enzymatic supramolecular hydrogelation. *Chem. Commun.* **54**, 86–89 (2017).

37. He, H. *et al.* Enzymatic Cleavage of Branched Peptides for Targeting Mitochondria. *J. Am. Chem. Soc.* **140**, 1215–1218 (2018).

38. Cao, M. *et al.* Enzyme-Triggered Morphological Transition of Peptide Nanostructures for Tumor-Targeted Drug Delivery and Enhanced Cancer Therapy. *ACS Appl. Mater. Interfaces* **11**, 16357–16366 (2019).

39. Sun, L., Zheng, C. & Webster, T. J. Self-assembled peptide nanomaterials for biomedical applications: Promises and pitfalls. *Int. J. Nanomedicine* **12**, 73–86 (2017).

40. Yang, Z., Xu, H. & Zhao, X. Designer Self-Assembling Peptide Hydrogels to Engineer 3D Cell Microenvironments for Cell Constructs Formation and Precise Oncology Remodeling in Ovarian Cancer. *Adv. Sci.* (2020) doi:10.1002/advs.201903718.

41. Hospital, A., Goñi, J. R., Orozco, M. & Gelpi, J. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinforma. Chem.* **8**, 37–47 (2015).

42. Gautieri, A. & Redaelli, A. (Politecnico di Milano). Molecular Modelling Lecture 2. 22–33 (2018).

43. Ghezzi, J. (Politecnico di Milano). Design and characterization of new self-assembling peptides through Molecular Dynamics simulations. (2020).

44. Eisenberg, D. S. & Sawaya, M. R. Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.* **86**, 69–95 (2017).

i. https://worldwide.espacenet.com/patent/
ii. https://www.3dmatrix.com/
iii. http://www.pepgel.com/pgmatrix.php

iv. https://www.sigmaaldrich.com/life-science/stem-cell-biology/3d-stem-cell-culture/hydromatrix-peptide.html
v. https://www.curodont.us/
vi. https://www.sigmaaldrich.com/catalog/product/sial/dlw354250?lang=it&region=IT
vii. https://www.biogelx.com/technology-synthetic-peptide-hydrogel-2/technology-benefit/
viii. https://www.3dmatrix.com/products/purastat/
ix. http://www.ks.uiuc.edu/Research/namd/
x. https://www.ks.uiuc.edu/Research/vmd/
xi. https://cgenff.umaryland.edu/