Contents

Part 2

Introduction

Part 1

The term "proteinopathy" was coined to indicate a class of disorders related to protein misfolding such as Alzheimer's (AD) and Parkinson's (PD) diseases. Nowadays, dementias, AD and type II diabetes are considered within the top 20 causes of death worldwide by World Health Organization (WHO). Misfolding not only induces the loss of proteins' biological functions, but also promotes a thermodynamically favoured process called "aggregation", that generally ends with the formation of protein fibrils, i.e. an ordered thread-like structure made of misfolded proteins. The formation of non-native protein species as well as the formation of fibrils deposits within or outside cells could compromise the native biological functions of cells, tissues and organs. Several endogenous and exogenous factors inducing protein structural misfolding have been identified, even if a complete understanding of the main players, interactions and forces leading to misfolding and fibrillation have been partially characterized. Currently, the pathological conformation of many proteins has been elucidated by cryo-EM and solid-state NMR, which actually present many experimental limitations regarding sample preparation and analysis. Despite UV Resonance Raman (UVRR) spectroscopy does not offer high-structural resolution measurements, it provides valuable insights on the native-like structure of the sample, thus overcoming the limitations of cryo-EM and ssNMR. To name few advantages, UVRR spectroscopy offers the possibility to work in diluted conditions, in native-like aqueous conditions and physiological temperature, without requiring a chemical manipulation of the sample. More importantly, UVRR spectroscopy opens the possibility to selectively enhance the Raman cross section of peculiar protein chromophore vibrational modes depending on the radiation excitation energy chosen.

The aim of the this Ph.D. thesis is to show the usefulness of UV Resonance Raman (UVRR) spectroscopy for the structural investigation of protein fibrils and also of protein-ligand interactions, linking the behaviour of several spectroscopic biomarkers to the structural modification of proteins during both phenomena. In particular:

- we demonstrated the ability of UVRR spectroscopy to get important insights on the structural modification of proteins upon fibrillation and upon interaction with ligands, identifying and studying the behaviour of few spectral fingerprints sensitive to both phenomena. We investigated the fibrillation of hen egg white lysozyme (HEWL) and human insulin and their interaction with an antioxidant, which is resveratrol. HEWL and human insulin have been chosen due to the large characterization reported in literature and considered as protein model systems. On the other hand, resveratrol, which is one of the simplest hydrophobic natural compounds, has been selected since its beneficial effect on both HEWL and insulin fibrillation has been already reported.
In this section, we provide a solid spectroscopic approach based on UVRR spectroscopy and complemented by other classical spectroscopic techniques, with the aim to translate this multi-technique approach towards the investigation of a more interesting class of proteins, i.e. the intrinsically disordered proteins. Interestingly, experimental results show that UVRR spectroscopy can replace the intrinsic fluorescence spectroscopy for the investigation of Trp-lacking proteins.

- in the second section we characterized a novel mechanism of fibrillation of *E.Coli* dihydrofolate reductase (DHFR), induced by the presence of spermine. In fact, differently from common DHFR inhibitors such as trimethoprim, the interaction with spermine induces only a partial inhibition of the enzymatic activity, with the formation of a high energy conformer. This structure remains stable for few days, then starts to precipitate into insoluble fibrils. The mechanism of interaction between *E.Coli* DHFR and spermine has been discussed in this section, providing a clear example of a fibrillation-inducing ligand. In fact, despite many compounds show a

beneficial effect, polyamines are actually able to promote the fibrillation of several proteins. The firstly reported case of *E.Coli* DHFR fibrillation at neutral pH has been reported here and characterized by a multi-technique approach based on UVRR spectroscopy.

- in the last section, the solid UVRR-based protocol tested with model systems, and the identification of peculiar UVRR spectral biomarkers sensitive to protein fibrillation and to interaction with ligands have been exploited dealing with a well-known intrinsically disordered protein (IDP), which is α-synuclein (aS). Nowadays, due to the plethora of published papers, aS is considered a Trp-lacking IDP model system and its proteinaceous deposits are considered the pathological hallmark of Parkinson's disease.

Hence, UVRR spectroscopy has been exploited using two different energies corresponding to 228 and 244 nm-excitation wavelengths, for probing different chromophores with the aim to extensively investigate the structure and the arrangement of the full-length wild-type and of the C-terminus truncated aS, both in the monomeric and in the fibrillar phases. In particular, we studied the structure of the (1-120) aS, which differs from the full-length protein for the removal of the three tyrosine residues located in the excluded portion of C-terminus. We characterized not only the secondary structure of both aS fibrils, complementing the UVRR results with FTIR and nanoFTIR data, but we also evaluated the morphological differences between full-length and C-truncated aS fibrils. The removal of part of C-terminus induces a major tendency of aS to form clusters and bundling of protofibrils.

Additionally, we proposed an alternative method of UVRR spectra analysis based on the use of an external protein reference. In particular, we interpreted the intensity variation of Tyr residues Raman peaks of aS depending on their solvent-exposure extent, and comparing these modifications to the Tyr residues intensity variation of a reference protein. Since the reference protein structure has been widely characterized in literature and the solvent-exposure of its Tyr residues is known in both the monomeric and fibrillar states, we could roughly estimate how many Tyr residues are solvent-protected in the case of both aS fibrils and monomers. We obtained results in line with the recent literature, indicating that Tyr39 remains

protected in the hydrophobic core of the fibrils, while it remains partially exposed in the case of (1-120) aS fibrils. Differently, in the monomeric phase, Tyr39 seems to remain solvent-exposed in both cases.

The structural arrangement of aS fibrils is of crucial importance for the development of targeting drugs and of specific radioligand which could be employed for the in vivo detection of aS aggregated. To this purpose we investigated the interaction between aS fibrils and Thioflavin T, considered as the golden standard for amyloid staining, which has been used to mimic the interaction between aS fibrils and model diagnostic ligand. Tyr39 has retained one of the main player in the interaction with antiaggregating ligands, due to its aromatic character, which is actually fundamental for maintaining aS biological function. Thus, by combining UVRR, FTIR and nanoFTIR spectroscopy, we deduced that Tyr39 seems not directly involved in the interaction with ThT, but in a region spatially close to a putative binding site.

Part 2

The second part of my Ph.D. thesis focuses on the characterization of the novel SARS-CoV-2 virus. This section collects the main results of my research activity on SARS-CoV-2 virus, obtained in collaboration with Elettra Sincrotrone Trieste, Ulisse Biomed S.p.A. (Area Science Park, Trieste), Institute of Human Virology (Baltimore, US) and Biomedical Campus of Rome. In fact, from March 2020, the experimental research activity concerning the Ph.D. original project has been limited, aiming to avoid a further viral spread of the emerging SARS-CoV-2 viral infections. However, in these difficult months, I had the pleasure to work with virologists, molecular biologists and epidemiologists, finally publishing four important papers, two of them as the first author.

The aim of this collaboration focused on the characterization of SARS-CoV-2 genome in order to monitor the emergence of new variants and to evaluate the impact of lockdown strategies and PCR-testing campaign during the pandemic.

Nowadays, these aspects are of fundamental importance as demonstrated by the emergence of new variants with alarming clinical significance. In particular:

- The first section of the Part 2 focuses on the characterization of newly emerging SARS-CoV-2 strains in different geographical areas and during the first three months of the pandemic. We identified several recurrent mutations hotspots that could increase SARS-CoV-2 viral adaptation and fitness, playing a key role in viral spread. In particular, we characterized for the first time the well-known recurrent mutation in the RNA-dependent RNA polymerase (RdRp) and the important D614G mutation in the spike protein, both identifying the emergence of the novel G-clade. The importance of these findings is about the fact that RdRp is usually considered as a target of many antiviral drugs since it is usually the most conserved portion of the viral genome. A RdRp mutation could impair the viral replication and could induce a process called lethal mutagenesis, where the virus becomes more and more attenuated, or could disfavour antiviral drugs binding. Additionally, the D614G mutation in spike protein could have an impact on viral transmission, escaping host immunity and modulating its penetrability in cells. Our speculation has been confirmed by experimental results recently published on *Cell* journal by Korber et al. (2020).

- In the second section of Part 2, we analysed SARS-CoV-2 case fatality rate (CFR) in several countries and during the first months of the pandemic. In particular, we discussed which parameters possibly affect CFR, indicating that the diagnostic test capability and the lockdown strategies adopted have a strong impact on CFR. Normalizing the CFR by a factor which takes into account the ratio between the amount of tests made and the positive cases diagnosed, we highlighted that CFR strongly depends on both the ability to individuate positive cases conducting an aggressive diagnostic test campaign, and on the adoption of strict lockdown strategies, aiming to reduce the viral spread. In particular, countries adopting less rigid lockdown strategies and/or making a less broad diagnostic test campaign are marked by a higher CFR, confirming that both

aspects are crucial for containing pandemic. In the same work we continue the characterization of SARS-CoV-2 mutation pattern, identifying the emergence of other viral mutations in Europe and US.

The present thesis is formed by two parts. *Part 1* is organized as follows:

- **Chapter 1**: protein folding and protein fibrillation are deeply discussed, elucidating the linkage between the onset of degenerative diseases and the presence of proteinaceous deposits inside or outside cells;
- **Chapter 2**: the basics of the quantum theory of Raman, UV Resonance Raman, FITR, nanoFTIR spectroscopy, AFM have been reported, dedicating particular attention in listing the advantages and the limitations of each techniques and on the description of the instrument setup;
- **Chapter 3**: here we report the investigation of HEWL and insulin fibrillation and interaction with resveratrol, testing a UVRR-based multi-technique approach, indicating the most sensitive reporters of both phenomena;
- **Chapter 4**: the novel DHFR fibrillation at neutral pH induced by spermine has been presented for the first time, characterizing the system with UVRR spectroscopy, complemented by preliminary biochemical study conducted by dr. Pucciarelli and dr. de Fronzo from Camerino University;
- **Chapter 5**: the structural modification of full-length wild-type and C-truncated α-synuclein has been presented and discussed in light of the results obtained by a UVRR-based multi-technique approach already tested in Chapter 3. Moreover, we propose an alternative method of UVRR analysis based on the use on an "external standard". Finally, the interaction between both full-length wild-type and C-truncated aS fibrils and Thioflavin T has been presented;
- **Conclusions**: this chapter discusses the main conclusions of the presented Ph.D thesis.

Part 2 of this Ph.D. thesis focuses on the study and on the mutation pattern evolution of SARS-CoV-2 virus, which has caused the emergence of the ongoing

pandemic. These studies were designed in collaboration with Institute of Human Virology (Baltimore, US), Ulisse BioMed S.p.A (Trieste), Elettra Sincrotrone Trieste, Biomedical Campus of Rome (Rome).

Thus, Part 2 contains a brief introduction on the SARS-CoV-2 genome architecture, followed by:

- **Chapter 1**: here, the characterization of newly emergent SARS-CoV-2 variants in the first 4 months of the COVID-19 pandemic and in different countries is reported, indicating that SARS-CoV-2 virus is prone to accumulate mutations with possible alarming clinical significance, potentially altering its drug resistance mechanisms and increasing its tendency to elude human immunity system;
- **Chapter 2**: this chapter reports the novel approach used for the estimation of the case fatality rate during the ongoing pandemic, introducing a novel parameter which takes into account the impact of lockdown restrictions and the PCR testing campaign capability in selected countries.

Part 1

Chapter 1

The unresolved problem of protein misfolding

1.1 Protein folding overview

Current advances in medicine and technology try to make longer and improve the quality of our life. Huge efforts have been currently made to understand why several diseases leading to death are not caused by external pathogens such as viruses or bacteria, but their onset is linked to an incorrect protein folding[1]. Proteins are essential for every organism since they are involved in a wide range of biological aspects and phenomena such as bio-recognition, DNA replication, structural function and enzymatic activity. Thus, once their function is somehow impaired, the effects could be disastrous. Most of the structural and functional properties of each protein are codified in DNA. The genetic information is translated by cellular structures called ribosomes from "nucleic acid language", encoded by adenine, cytosine, thymine and guanine, to "protein language", encoded by amino acids, which are 20 bio-bricks composing proteins.

Amino acids are formed by a carboxylic and an amine group and are identified by a "R-group", which is the functional group giving them different chemical and biological properties. In fact, some of them are hydrophilic, some others are

hydrophobic, positively or negatively charged. Amino acids are linked by covalent bonds, forming a polypeptide called protein *primary structure* which can be modelled as a linear chain. Other weaker interactions such as hydrogen bonds stabilize the *primary structure* into protein *secondary structure*, characterized by the arrangement of primary structure into α-helices and linear sheets, called β-sheets as shown in Fig. 1.1. The formation of α-helices and sheets is tightly dependent on protein primary sequence composition. Noteworthy, the alpha carbon (C_α) in the center of each amino acid forms two rotatable bonds with the amine and carboxyl groups. These dihedral (torsion) **angles** are called **Phi** and **Psi** (in Greek letters, φ and ψ). Depending on those values and given the primary sequence, specific energy-based calculations are able to predict which protein secondary structure is the most probable to occur. This calculation could be visualized by the Ramachandran plot, a graph reporting the values of φ and ψ on the (x,y) axis (Fig. 1.2). Based on the values of the dihedral angles, we can easily visualized the energetically allowed regions for the formation of α-helix and β-sheet structures. A factor that has to be considered is the steric hindrance of side chains, since it furtherly limits the allowed regions.

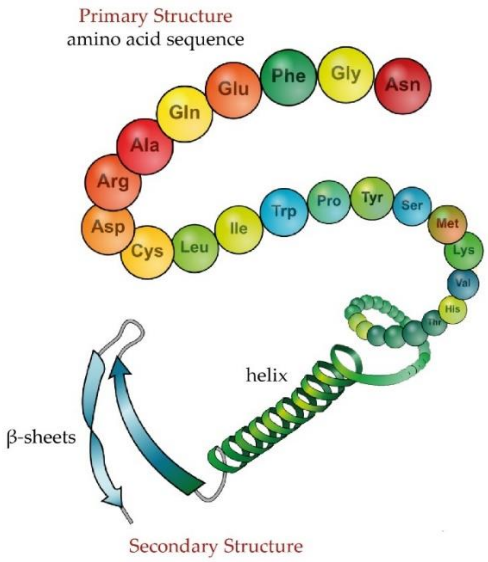


Fig. 1.1: Primary and secondary structure of a protein. Primary structure is formed by the amino acids sequence and the secondary structure is the 3D arrangement of the primary structure into hydrogen-bonded structure such as α-helix and β-sheets. The figure has been readapted from[2].

α-helix is particularly stable because of its higher number of hydrogen bonds that keep the structure stably folded. It is characterized by an optimum value of 3.6 amino acids per turn and the number of turns may vary from 3 to 4[3]. Regions richer in alanine, glutamic acid, leucine and methionine usually tend to form helices. Differently, β-sheets are often found in globular proteins and they are less hydrogen bonded than helices. They are formed by the interaction between stretched residues segments and they can be aligned in parallel or antiparallel structure[3]. In the antiparallel disposition of strands, N-termini are in close contact with C-termini and this arrangement produces stronger H-bonds respect to the parallel arrangement, thus favouring the macrodipole interactions. Two consecutive secondary structure are linked by β-turns, which display a H-bond between the CO and the NH group. Generally, proteins are able to further be arranged into an additional three-dimensional stable structure by a compact folding of secondary structure, forming the so-called *tertiary structure*[3]. Compared to primary and secondary structures, the tertiary structure is stabilized by both electrostatic and hydrophobic interactions. These forces permit the formation of a quite compact and packed structure where hydrophobic amino acids remain solvent-protected in the inner core of the protein. Due to this arrangement, molecular groups or atoms that can be far away from each other in the aminoacidic sequence, can be found in close contact and interact with each other.

These weaker forces let a biomolecule able to perform also large motions, being essential for protein functionalities[3].

The maintenance of this native tertiary structure is influenced by solvent pH, temperature, pressure and other factors. The dynamics and the static equilibrium of proteins are governed by a delicate balance of stabilizing and destabilizing forces and are also influenced by steric hindrance. On one hand, this landscape of interactions and phenomena increases protein complexity and consequently their capability to perform multiple functions and activities but, on the other hand, complexity contributes also to increase structural errors including misfolding which lead to protein dysfunctions.

The structural and dynamical modifications upon external variation have been explored and are currently modelled as an energy landscape characterized by isoenergetic minima separated by energy barriers. Proteins tends to reach energetic minima by modifying their structure, namely by fluctuating across this energy landscape.

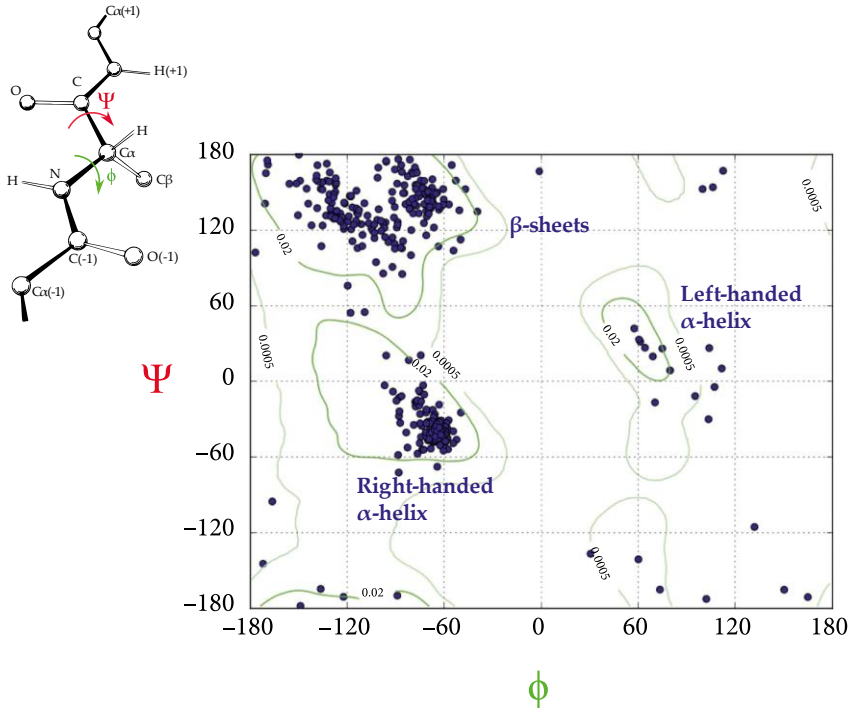


Fig. 1.2: The Ramachandran plot is a representation of the torsional angles (ϕ and ψ) of the residues contained in a protein, which are shown in the upper part of the picture. Ramachandran plot graphically shows which combination of angles are possible: ϕ and ψ define the geometry of the residue attachment to its two adjacent residues, thus determining the conformation of the residues and the peptide. Figure readapted from[4].

Each conformation can be stabilized by a pool of different interactions such as disulphide bridges, hydrogen bonds, electrostatic, Van der Walls interactions and hydrophobic forces. In particular, disulphide bridges mostly stabilize protein secondary and tertiary structures, while hydrogen bonds are mainly involved in the stabilization of the secondary structure. Hydrophobic interactions are essential during protein folding, keeping most of the non-polar residues as solvent inaccessible, whereas electrostatic interactions are mostly sensitive to the distribution of charges along the amino acidic sequence with respect to pH and

solvent type. The inner core of a protein is strongly packed and kept stable mainly by Van der Waals interactions; however, due to these low energy-interaction, a protein is still able to perform amplitude motions.

On the basis of protein tertiary structure, it is possible to mainly define three different class of proteins: globular proteins, intrinsically disordered proteins (IDPs) and fibrous proteins. Globular proteins are water-soluble proteins characterized by different secondary structures motifs and are usually folded in a spherical shape. This class of proteins usually performs enzymatic or regulatory activities. On the contrary, IDPs are flexible and can adapt its aminoacidic sequence into multiple secondary structures depending on their activity. Finally, fibrous proteins such as collagen, are basically β-sheets-like water-insoluble biomolecules, which provide structural support to tissues.

1.2 Energy landscape

Protein configurations are not energetically heterogeneous, and proteins tend to adapt their structure in order to minimize the energy costs. Hence, the energy of each configuration is strongly dependent not only from protein sequence, but also from the solvent[3]. Hence each protein conformation, induced by the interplay between the effects provoked by environmental factors and by multiple stabilizing interactions, represents an energetic state of a complex protein energy landscape, characterized by the presence of energetic minima separated by free-energy barriers. Thus, the energy landscape describes the potential energy of a protein as a function of its conformations (see Fig. 1.3). The minima are linked to specific protein conformational states CS such as native, unfolded, partially folded, molten globule and so on. The energy landscape may be extremely rugged with hills and valleys. Fluctuations between energy minima are pivotal for protein activity, since protein motions are the effect of the transitions among different sub-states. Those fluctuations have non-zero probability to occur, but they are relevantly influenced by external conditions such as pH, solvent, temperature, pressure and so on. Each energy minimum is formed by several sub-valleys (CSi), therefore implying a certain

structural and dynamical hierarchy[3]. Each sub-minimum is linked to a minor structural or kinetic modification. These sub-valleys may also contribute to the fluctuations in the positions of individual atoms.

By changing their structural conformation, proteins naturally could explore this energy landscape, trying to reach the best conformational state, i.e. an energy minimum. In this view, the configurational entropy decreases in a funnel-like manner, where the vertical axis represents the free-energy[3]. As the free-energy increases, the number of conformations grows, implying an increasing in conformational entropy, so in the surface area. Thus, the bottom of the folding funnel is characterized by high free-energy and by the presence of unfolded conformations, while local minima can kinetically trap proteins in misfolded or intermediate conformations and the deepest ones usually trap proteins in native or aggregated conditions[5,6].

Nonetheless, several proteins do not follow a unique funnel, but some of them could fold into different but energetically favoured structures. In light of these consideration, another class of fundamental proteins called chaperones, which are extraordinarily flexible and dynamic, are able to control the quality of protein folding. Interestingly, they assist proteins to reach their native fold, preventing protein aggregation and mediating targeted unfolding and disassembly[7].

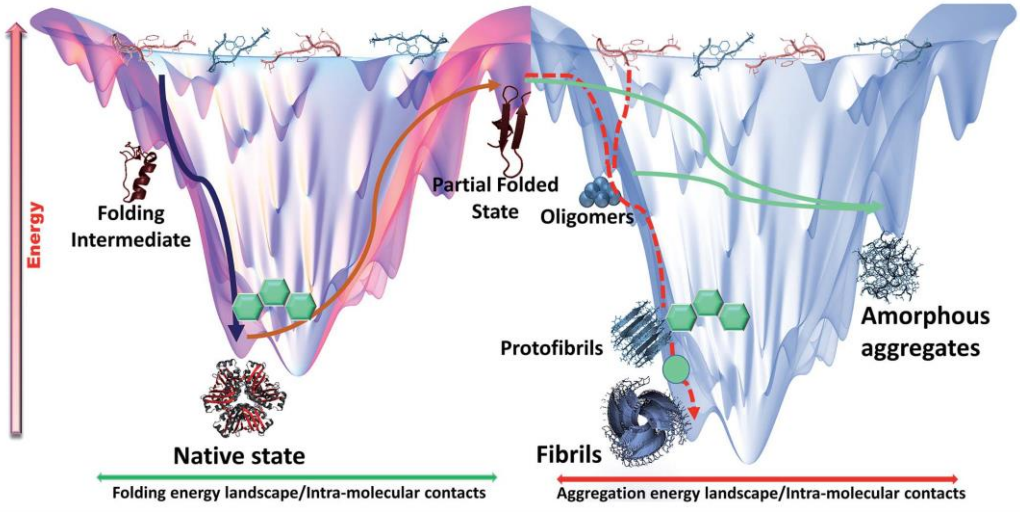


Fig. 1.3: The folding and aggregation energy landscapes. The funnel on the left describes the folding pathway of a protein into its native state. Blue arrows indicate the consequent minimization of the protein free energy, indicating

that protein folding is characterized by the presence of transient folded intermediates. On the right, the pale blue funnel describes the aggregation landscape dominated by inter-molecular interactions. Destabilization of native states or partially folded states could induces their entrapment in a high energy partially folded state, that could be minimize their energy by forming amorphous (green line) or amyloid states (dashed red line). In order to reach the amyloid state, proteins undergo a structural rearrangements into intermediate species such as oligomers and protofibrils. The figure is readapted from[8].

1.3 Protein structure-function paradigm and its limitation

The central dogma of structural biology is that the three-dimensional protein structure determines its function[9–12]. This paradigm has been retained valid over many years, especially because it actually well explained lot of biological processes such as enzymatic activity of certain proteins[9,13,14]. Enzymes are proteins able to reduce the activation energy of a chemical reaction, i.e. to catalyse it, participating to the process without consuming or altering irreversibly their structure[9]. Specifically, enzymes catalyse most chemical reactions in cells, accelerating the rates of such reactions by over a million-fold, due to their large and flexible structure. However, to perform their biological functions, enzymes need to change their structural conformations several times in order to allow the binding of the substrate in its active site, converting it to the final product of the reaction and finally releasing it . Hence, it exists an underlying relation between protein structural conformation and their biological functions. Haldane and Pauling presented how this paradigm works and is essential to explain the enzymatic activity, treating the active site as a "static site" where the reaction could take place by lowering its activation energy[9,10,15].

However, recent experiments showed that actually enzymatic active sites are dynamic and mobile on the microsecond to second time scale[16]. Additionally, the discovery of large disordered regions in at least 20% of eukaryotic proteins which displays multiple biological functions question the reliability of the structure-function paradigm[16]. Thus, this dogma has to be furtherly updated in light of the recent findings[11,16,17]. In fact, despite the linkage between protein folding and energy landscape as a key to take into account for the determination of protein functions is

well-established, the mechanism that lead disordered proteins or partially disordered ones to regulate transduction processes while exhibiting enzymatic activity, without an arranged 3D structure, remains mostly unclear[11,17].

1.4 Misfolding and amyloidosis

Misfolded proteins are the result of their incorrect folding, of a wrong energy-minimizing funnel, that can be induced by multiple factors such as proteins' intrinsic instability or/and to unfavourable external conditions and interactions with other molecules[1,18,19]. Misfolded proteins are usually toxic for living organism and able to catalyse the transition of other native proteins into toxic ones, through a cross-seeding process[20–23]. Cells dispose of many mechanism for preventing the incorrect protein folding or for getting rid of misfolded proteins such as the intervention of chaperones, ubiquitin proteasome and autophagy systems[24–26].

Despite the presence of this cellular regulatory and protection systems, several proteins still misfold and change irreversibly their secondary structure, evolving into β-sheets-rich protein and depositing inside or outside cells as toxic formations. The presence of these proteinaceous deposits, i.e. *amyloids*, has been correlated to the onset of related degenerative diseases and are the result of an incorrect protein folding due to a plethora of exogenous and endogenous reasons[27].

1.5 History of amyloids

First described in the XXVII century as lardaceous and waxy liver, spongy and "white-stone" spleen, the term "amyloid" was coined by the German scientist Rudolph Virchow in 1854[27,28]. Working with brain tissues, Virchow started to study cerebral corpora amylacea that have an abnormal macroscopic appearance. He found that these deposits stained pale blue when treated with iodine and, consequently, violet upon the addition of sulfuric acid. Similar behaviour had been observed for starch (*amylum* in Latin and *amylon* in Greek), so he concluded the abnormality was given by the presence of polysaccharides in these deposits and for

this reason he coined the term "amyloid". Few years later, Friederik and Kekulé showed that amyloids contain a relevant content of nitrogen, that questioned Virchow speculation[29]. Actually, they demonstrated amyloids are mainly formed by proteins and glycosaminoglycans.

In order to understand whether amyloids were amorphous or ordinated, scientists started to study amyloids using light microscopy and employing also dyes such as Congo Red and Thioflavin T. Curiously, polarized light microscopy demonstrated that amyloid accumulations in tissues showed an interesting spectroscopic property when stained with Congo Red: they showed a red-green birefringence respect to the axis of the deposits. This evidence revealed that, actually, these proteinaceous deposits were somehow "structurally ordered". In 1901, Prof. Alzheimer's reported the first documented case of post-mortem observation of Congo Red-positively stained amyloid plaques in one of his patient brain tissue[30,31]. In 1936, the first X-Ray fibre diffraction studies performed on *in vitro* denatured proteins, revealed that lots of proteins may underwent aggregation. Later, Cohen and Calkins performed electron microscopy studies on human tissues-extracted amyloid and *in vitro* ones, discovering they were actually formed by bundles of fibrils[32]. In 1968, Geddes and colleagues deeply characterized the structure of these fibrils using X-ray diffraction, discovering they were formed by a repeating motif alongside their axis, which they called cross-β[33].

Tissues-isolated proteins fibrils are structurally and morphologically similar to those obtained in vitro, as X-ray diffraction studies revealed[34]. Interestingly, all amyloidogenic peptides presented an ordered secondary structure, mostly characterized by the presence of β-sheets oriented perpendicular to fibrils axis.

Proteins sequencing revealed that each amyloid-related disorder was associated to the presence of specific proteins deposits[27,35]. In some cases, the presence of these deposits was found in a specific organ such as pancreas or brain. Differently, in systemic amyloidosis, amyloid fibrils are deposited in multiple organs[35]. A different class of highly amyloidogenic proteins was discovered in 1983 by Stanley Prusiner[36]. He defined these proteins as "prions" since they are infective. Few years later, a prion diseases called bovine spongiform encephalopathy (BSE) caused an epidemic

affecting cattle and several years later, a new variant of prion-disorder affected humans (Creutzfeld-Jacob diseases), which was strongly related to the exposure of the BSE pathogen.

Today, more than 50 diseases were correlated to the self-assembly of specific proteins and peptides into fibrils, depending both on the site of amyloid deposition and on the protein sequence portion involved[18,27,37–39].

In Table 1.1, several amyloid-related disorders are summarized together with the disease-related precursor protein.

Crosstalk between amyloid diseases is also reported. As an example, fragments of a-synuclein were found in plaques of patients suffering from Alzheimer's disease.

Table 1.1: Protein precursors associated with amyloid disease. A list of protein precursors is presented here, complemented by their native structure and the related disease.

Disease	Aggregating specimen	Native structure
Alzheimer's disease	Aβ, tau proteins	Natively unfolded and α-helical
Creutzfeld-Jacob disease	Prion proteins	Intrinsically disordered
Amyotrophic lateral sclerosis	Superoxide dismutase 1	β-sheets
Parkinson's disease, Dementia	α-synuclein	Intrinsecally disordered
Hungtington's disease	Hungtingtin fragments	the poly-Q region is largely unstructured
Familiar amyloitic polyneuropathy	Transthyretin mutants	β-sheets
AL amyloidosis	Immunoglobulin light chains	β-sheets
Haemodyalisis-related amyloidosis	β2-microglobulin	β-sheets
AA amyloidosis	Serum amyloid A1	α-helical and unknown fold
Lysozyme amyloidosis	Lysozyme mutants	α-helical and β-sheets
Apolipoprotein A1 amyloidosis	Apo-A-1-fragments	Intrinsecally disordered
Type II diabetes	Amylin	Intrinsecally disordered
Atrial amyloidosis	Atrial natriuretic factor	Intrinsecally disordered
Injection-localized amyloidosis	Insulin	α-helical
Cataract	γ-crystallins	β-sheets

1.6 Structural characteristics of amyloidogenic proteins

Nowadays, the study of amyloids has become a central topic in science since only neurodegenerative diseases caused by proteins disfunction and aggregation are listed in the top 20 causes of death worldwide by WHO, together with type II diabetes, cardiovascular disorders and cancer. Amyloidosis comprises a large group of disorders characterized by the extracellular presence of misfolded and aggregated fibrillar proteins. Several diseases are due to the formation of intracellular amyloidogenic deposits as well. These deposits can be concentrated in specific organ areas or can be present in multiple organs, contributing to systemic disorders such as light-chain amyloidosis (AL)[35]. As reported in Table 1, aggregation-prone proteins may have a native structure rich in α-helix, β-sheets or can be intrinsically disordered, depending on their particular functionalities. Similar to globular native states, amyloid structures are closely packed and highly ordered whereas, differently from native states, they possess a generic architecture that is rich in β-sheet structure[1].

Amyloid fibrils architecture of different proteins seems to be remarkably similar at the nanometre length scale: they typically consist of multiple protofilaments twisted around each other. The different morphology of fibrils is directly related to the number of protofilaments and to the type of interactions that held together. In fact, other protofilaments arrangement have been observed such as cylinders, flat ribbons and pseudo-crystalline sheets. It has also been observed that the morphology of fibrils can change over time and in response to changing environmental conditions[40,41].

X-ray diffraction studies indicated that the core of each protofilament adopts a cross-β structure, in which a ladder of β-strands, separated by a 4.8A spacing, form hydrogen bonded β-sheets that run along the fibril axis[1]. Due to the array of β-sheets, cross-β architecture could provide an enhanced stability to fibrils, forming also dry "steric zippers", namely sort of solvent-protected tight interfaces formed by hydrophobic side chains[1]. An example of the fibrils structure, of the arrangements of protofilaments and of the β-sheets staking characterizing cross-b structure is reported in Fig.1.4. Despite their presence results in negative effects in organisms,

amyloid state are characterized by a thermodynamically more stable conformation than the functional native states of many proteins. In fact, the conversion from soluble proteins to the amyloid form involves the formation of intermolecular contacts[1]. However, the energy landscape of fibril formation is far more complex than that for globular proteins, potentially involving multiple intermediates, resulting in diverse amyloid end products with different cross-β structures[42]. Great efforts have been made for correlating the different fibrils morphologies and deposits composition to the early onset of the pathology. Noteworthy, amyloid deposits do not contain only a specific precursor protein fibril, but also glycoprotein (SAP), proteoglycans such as heparan sulphate proteoglycans and proteins such as Apolipoproteins[28,35].

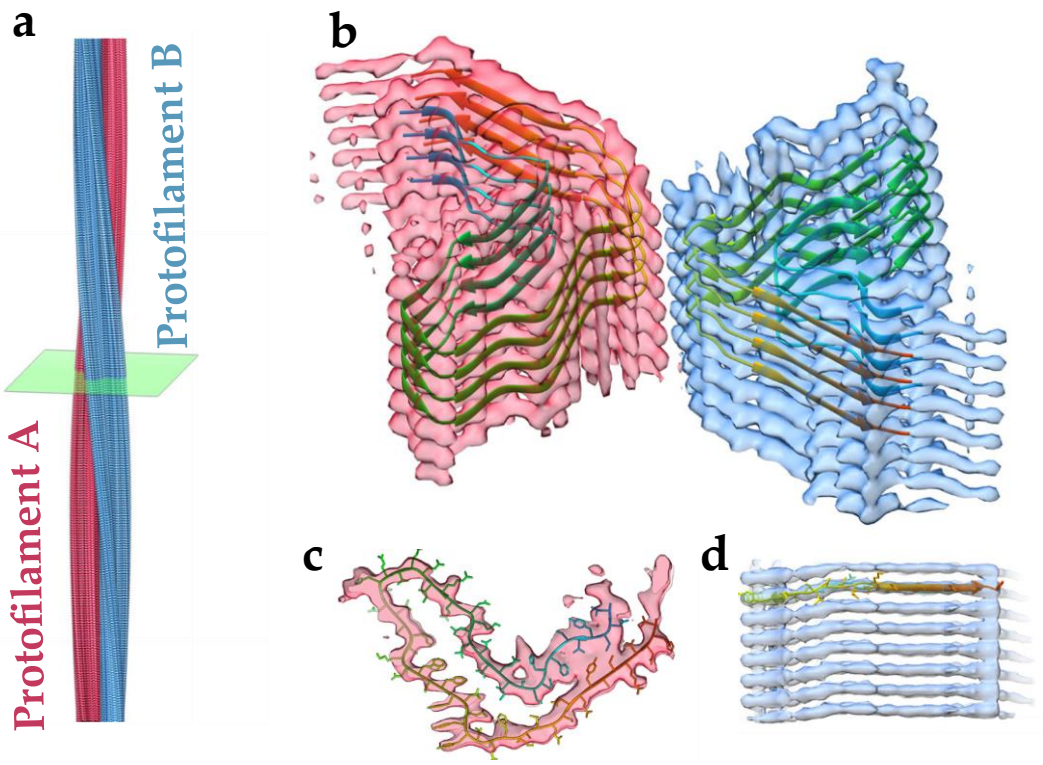


Fig.1.4: An example of amyloid fibrils. (a) 3D structure of an amyloid fibril formed by two protofilaments (A, in red and B, in light blue). (b) Transversal section of the fibrils reported in (a). Light red and light blue portions represent protofilaments A and B, formed by a ladder of β-sheets aligned perpendicularly to the fibril axis. (c) Cross-section of a protofilament showing the arrangement of β-sheets and (d) side view of a protofilament where each ladder stack vertically and equally spaced. Picture readapted from[43].

1.7 Amyloid fibrils aggregation mechanisms and pathogenesis

The exact mechanism that converts proteins into amyloid fibrils *in vivo* is not fully understood. Usually this structural conversion is a pathological process, but the occurrence of functional amyloidogenic deposits are also reported[44].

Proteins can evolve from a functional soluble state to the amyloid state going through steps involving the formation of multiple precursor species such as oligomers as well as proto-fibrillar structures[1,27].

Fibrils formation pathway depends on multiple factors and not only proteins or peptides forming in-vivo deposits are prone to aggregate. Under favouring environmental conditions such as those that reduce the tertiary structure stability, non-amyloidogenic proteins can be stimulated in forming fibrils similar to those extracted from tissues .

Thus, fibrils forming capability is a protein intrinsic property, mostly depending on the polypeptide backbone[19,45]. In fact, proteins with a diverse native structure, i.e. different amino acid sequence, could evolve in similar fibrillar structure under denaturant conditions. Diverse amino acidic sequence contributes to modulate fibrils formation capability which is mainly due to the presence of hydrogen bonds between NH and CO groups of the polypeptide backbone.

The mechanism of fibrils formation is usually schematized as a nucleation-dependent polymerization[46–49]. This process is mainly characterized by two step: nucleation and elongation. Proteins lose their native structure in favour of a b-sheet rich one in the nucleation phase and, consequently, we can assist to the formation of seeding nuclei and small aggregates such as oligomers[35]. The formation of nuclei is an energetically unfavoured step, thus pre-formed nuclei are prone to attach other soluble proteins or small aggregates to them, enlarging their dimension in order to recover the loss of energy due to the nucleation[1].

A rapid polymerization occurs when monomers concentration overcome a critical concentration[50]. At this point, when a critical concentration of nuclei or aggregated species has been reached, the elongation of fibrils from pre-formed nuclei takes place

following an exponential growth and reaching a plateaux when mature fibrils were formed[35].

Schematically, the aggregation process can be summarized in three major phases: the structural transformation of proteins, then the formation of oligomers and protofibrils and, finally, the formation of aggregates[35]. Thus, this process follows a sigmoidal reaction time course as reported in Fig. 1.5. However, few exceptions have been observed and the sigmoidal fibrils formation is not always observed. In fact, secondary processes might interfere such as fibrils fragmentation, which increases the number of fibrils ends that can recruit other soluble protein molecules, reducing the length of lag time (i.e. the requested time for the formation of nuclei) and facilitating the proliferation of fibrillar species[1]. In fact, the presence of truncated proteins in amyloid deposits is well-documented. Interestingly, in several cases it has been demonstrated that fragmentation occurs selecting fragments which have a higher propensity in self-aggregation respect to the full-length protein[51,52].

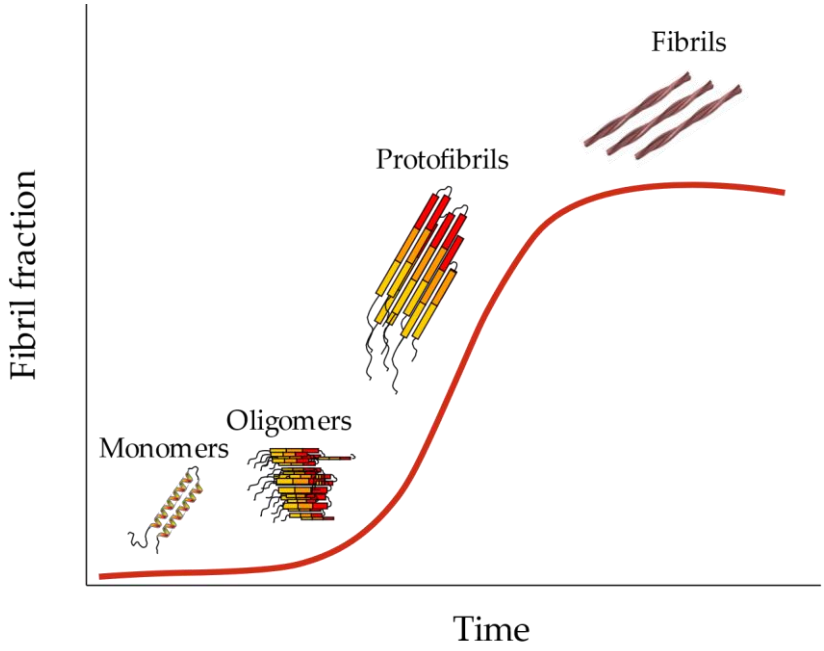


Fig. 1.5: Kinetics of fibrils formation. The red line indicates the formation of amyloid fibrils. The process begins from a solution of monomeric proteins, then due to endogenous or exogenous conditions, the formation of non-native specimens (i.e. oligomers) take places during the so-called lag phase. Once these nuclei have been formed, the process of aggregation proceeds with a fast kinetics. After lag phase, the formation and the elongation of protofibrils induces the formation of mature fibrils.

An example is α-synuclein, where C-terminal truncated proteins show an highly amyloidogenic propensity to self-aggregate[52]. Additionally, pre-existing fibrils may catalysed the formation of oligomers along their surfaces, strongly influencing the length of lag time[1].

Nonetheless fragmentation and fibrils surface catalysis could take place, other factors make faster the aggregation pathway[53,54]. In fact, several proteins are characterized by an intrinsic native instability such as transthyretin or the whole class of intrinsically disordered proteins, or by a higher concentration in biological tissues or liquids such as for β2-microglobulin[35]. Surely, precursor protein gene mutation and primary sequence variations such as sequence expansion, can be directly related to the early onset of a disease as well as to the duration. These two phenomena could result into a different fibril folding. As an example, SNCA gene modification may induce one of the most representative nonsynonymous amino acidic mutation of a-synuclein, which can lead to an early onset of the Parkinson's diseases, to a reduced disorder duration or to the emergence of additional symptoms (cognitive impairment, hallucinations and so on)[55]. Additionally, as reported by ssNMR, mutated a-synuclein present a different core architecture[56,57], a different *in vitro* cytotoxicity and a different rate of plaque deposition *in vivo*. Same evidences have been obtained dealing with Aβ[58]. Other disorders are associated with amyloidogenic sequence expansion, as shown for polyQ in the case of Huntington disease[59], of poly-GlyAla in C9orf72 which is the most common genetic form of ALS and frontotemporal dementia[60,61]. In polyQ related diseases, the more is the expansion of the repeating unit, the worst are the symptoms[59]. Moreover, in patients suffering from Huntington's disease the age of the disease onset has been correlated with the number of CAG unit repeats[62]. Additionally, the modification of the amyloidogenic precursor sequence by truncation or post-translational modifications (phosphorylation…) may modulate the extent of amyloidogenicity of a precursor protein[55]. Moreover, it has been demonstrated that the presence of small molecules or metal ions can induce a rapid fibrils-formation[55].

As reported in literature, oxidative stress plays an important role in cell exposure to early aggregates[63,64]. In fact, cell treatment with antioxidants or phenolic compounds

results in an increase protection against aggregate toxicity[64]. However, protein aggregation is generally accompanied by and increased in reactive oxygen species in the cell and in vitro[63,65]. Intracellular oxidative stress may derive from the destabilization of cell membrane as well as their ion pumps, and from the disfunctions of mitochondria[64]. Ageing is surely another key factor that expose cells to a major oxidative stress, letting them more vulnerable to amyloid interaction[66,67]. The permeabilization of cell membranes, membrane lipid peroxidation and the ion pumps dysfunction have a clear involvement in amyloid-related disorders[55,64,68]. However how the observed structural polymorphism of amyloid fibrils affects disease onset, progression and presentation is still unclear[69,70]. Non-infectious amyloid disease progression is usually a slow. Differently, prion diseases are characterized by a reduced incubation time and death can be observed also after few days of incubation[27].

Understanding how each process may interfere with the aggregation pathway and how they combine to determine the early onset, or the severity of the disease is vital for treatment purposes. In any case, the presence of intra- or extra-cellular toxic aggregates can impair a number of cell functions such as synaptic transmission and plasticity, cell membrane permeability, mitochondria functioning, cell signaling, eventually promoting cell death by apoptosis or by necrosis[55,71–76].

Even if the presence of amyloid deposits are associated to the onset of pathological events, usually leading to a general impairment of synaptic transmission and plasticity, of cell membrane permeability, mitochondria functioning, cell signaling and so on, much remains to be understood about the specific mechanism underlying these events.

1.8 Cells cytotoxicity

Nowadays, the identification of toxic species generated during protein aggregation and how they are involved in the cellular dysfunction and death remain still priority and challenging. Pre-fibrillar or fibrillar products may impair several cellular mechanisms such as autophagy, they facilitate the production of reactive

oxygen species or the sequestration of other proteins or molecules, the disruption of cellular membranes[24,64,66,68,77]. Several experiments were performed in order to explore these phenomena and, lots of them assess oligomers as the most toxic species formed able to disrupt cellular membranes. However, not all the oligomers are cytotoxic as evidence by Fusco for α-synuclein[68,78,79]. The mechanism with which toxic oligomers disrupt membranes has been studied and it seems that the origin of oligomers toxicity may arise from their misfolded structure. In fact, differently to monomers and to fibrils, oligomers displays many hydrophobic residues on their surfaces which have a higher affinity in binding to membrane[68,80–82]. This inappropriate binding is followed by a relevant perturbation of the chemical equilibrium of membranes, provoking a calcium efflux from cytoplasm towards the extracellular space leading to cell death[68,78]. However, a high-resolution structure of the toxic oligomers may help in understanding the molecular mechanism undergoing their cytotoxicity. Unfortunately, oligomers are unstable and heterogeneous and determining their structure-function relationship is still a challenge. Surely, not only toxic oligomers may lead to the onset of the disease but also fibrils may perturb lipid membranes and their deposits may interacts with membranes, sequestering lipids and other molecules[83–85]. Different fibrils morphologies as well as different aggregation products may have a different impact on cells and on the onset, duration and progression of the diseases[55,57,69]. Amyloid formation is a dynamic process, that is directly dependent on environmental pH, on the overproduction of proteins, on the presence of diverse metabolites or small molecules and so on. The study of fibrils structure may help in understanding the presence of certain amyloids deposits, such as those formed by PMEL, are functional for organisms even sharing a similar fibrils architecture[44].

1.9 Amyloidosis diagnosis and treatment

The diagnosis of amyloidosis has evolved over time. In the past, biopsy of an affected organ was the less invasive and sensitive way[28,86]. Nowadays a way is to perform SAP scintigraphy, which exploits radiolabelled SAP protein which

specifically localizes to amyloid deposits *in vivo*, in proportion to the extent of amyloid present[87]. However, SAP scintigraphy is not always equally sensitive and this depend on the disease which have to be diagnosed. It is possible to affirm that biopsy and SAP scintigraphy may be complementary techniques in amyloidosis diagnostics.

In patients suffering from cardiac amyloid infiltration, ECG is currently employed[88]. "Q" wave is usually targeted as a marker of a pathological state since it shows low voltages when a cardiac disorder occurs[89].

Innovative biophysical techniques are currently employed to increase the knowledge about the processes involved in amyloid formation and to furtherly define the high-resolution structure of protein fibrils. The introduction of a series of small and large molecules that are able to modulate these processes opens to the development of diagnostics and therapeutic strategies[27].

Amyloid deposits are currently diagnosed in several ways using cerebral spinal fluid diagnostics, positron emission tomography imaging and immunoassays[90,91]. However, amyloid deposits are mainly formed by fibrils of a specific precursor protein fibrils, but it is not infrequent that different fibrils are found in these deposits and with diverse polymorphism. For these reasons, the in-vivo diagnosis is not easy as well as a personalized treatment against amyloidosis. Great efforts are made to find out new biomarkers able to diagnose specific diseases.

1.10 Drug design strategies against amyloidosis

A pioneering and successful treatment of a human amyloid disease occurred with transthyretin (TTR) amyloid disease involved surgically-mediated gene-therapy. TTR amyloid disease is mostly caused by the aggregation of mutant TTR which is almost completely secreted by liver into the blood[92]. Liver transplantation successfully slow down the process of TTR aggregation[92]. However, this approach presents several limitations due to patients mortality due to surgical complications, limited organ availability and more others. Additionally, TTR is not produced only by liver and a little percentage continues to be secreted by choroid plexus or by the

retinal pigment epithelium cells, leading to relevant ophthalmological and neurological complications.

Another strategy is the "protein reduction" which is currently employed in order to decrease the concentration of precursor and amyloidogenic proteins in blood and tissues by using several approaches such as the employment of chemotherapy agents, the inhibition of specific enzymes, the use of anti-inflammatory drugs and antibiotics and, finally, also the gene silencing[92]. This strategy has been currently employed for the treatment of light chain (LC) amyloidosis and Alzheimer's disease[92]. Another strategy is the "kinetic stabilization approach", aiming to discover molecules able to stabilize and prevent the misfolding of several aggregation prone-proteins[92].

Recently, it has been observed that epigallocatechin-3-gallate (EGCG), a well-known green tea polyphenol, effectively remodel a wide range of different protein aggregates, including amyloid fibrils and oligomers. The amyloid remodelling observed by EGCG treatment appears to be driven by hydrophobic binding of oxidized EGCG molecules to amyloid fibrils, and appears to stabilize the aggregates preventing the formation of toxic oligomers or by directly binding to oligomers, avoiding their interactions with cellular membranes as observed for α-synuclein[92,93].

1.11 Natural compounds: mechanism of interaction and clinical trials

Epidemiological studies suggest that diets rich in flavonoids and polyphenolic compounds as in the Mediterranean diet, correlates with a reduces the aging-risk of AD, dementia, diabetes and so on[94]. Several polyphenols such as resveratrol, curcumin and EGCG have progressed to clinical trials for Alzheimer's treatment[95,96].

Anti-aggregation agents can chemically interact with amyloid species by forming covalent bonds or via non-covalent interactions such as hydrogen bonding, π-π stacking, hydrophobic interactions, charge-charge interactions[97,98]. The type of the

non-covalent protein-ligand interactions depends on the protein or the aggregation stage that wants to be targeted.

Those compounds are usually less-soluble in water and hydrophobic, which permits them to bind preferentially to proteins, instead of remaining dispersed in solution[97,98]. This could explain also why natural hydrophobic compounds are able to bind to protein. The outcome of the interaction between proteins and natural compounds may be different: the same natural ligand can inhibit the early oligomerization, it can accelerate the process, or it can lead to the formation of non-fibrillar insoluble aggregates. As an example, epigallocatechin gallate (ECGC) inhibits the amyloid precursor peptide (PAP) fibril formation and, at the same time, directly prevent Aβ and α-synuclein fibrils formations. Additionally, ECGC promote the formation of high insoluble fraction of light chain immunoglobulin[99,100]. Insoluble aggregates may display a wide range of size distributions and are usually non-toxic, as reported in cell viability assay[99], differently to pre-fibrillar aggregate which are believed to be the most toxic species during the aggregation phenomenon. Due to the complexity of the systems and since the mechanisms underlying both aggregation and its inhibition are not fully characterized and understood, high-throughput screening have been put in place in order to find the best candidates to target a specific protein or a specific stage in the aggregation process among a wide library of molecules[99,101]. In addition, molecular docking and MD simulations help in order to understand the binding mechanism responsible for aggregation inhibition, making also a forecast of the possible binding sites[102,103].

Even if the powerfulness of natural compounds has been tested *in vitro*[95,96], few limitations and factors may have to be taken into account through *in vivo* studies concerning membrane permeability, drug delivery and drug metabolism. Most drugs and natural compounds tend also to crystallize, limiting also their dissolution and solubility: the combination of their hydrophobicity and higher tendency to crystallize, compromise natural compounds bioavailability[99]. In fact, while several compounds showed an outstanding ability in targeting specific aggregated proteins, improving also memory disfunction and reducing inflammatory cytokines levels in animal models, differently they showed poor outcomes when tested with humans due to

their limited bioavailability, to their lower water-solubility and to their lower capability in passing the blood-brain-barrier[99]. In fact, in order to reach the wanted therapeutic effect, large quantities of these compounds must be administered, resulting in an acute toxicity response[99].

Currently, novel approaches are tested in order to fix this problem. Nanoparticle-based drug delivery and lipid-based nanoparticles approaches are currently used in order to improve the bioavailability of natural compounds, including the passage across the blood-brain-barrier[99]. Additionally, liposomal encapsulation has also been used to improve the bioavailability of hydrophobic compounds[104]. Moreover, poorly soluble natural compounds can be mixed with a polymer in order to form an amorphous solid dispersion, which can prevent crystallization, improve also the storage stability and, depending on the amorphous phase used, also the solubility[105].

1.12 Techniques used for IDPs and protein fibrils structure

Being complex phenomena, protein folding, protein aggregation as well as their interaction with ligands have been studied using different techniques and approaches. One of the main concerns working with amyloid fibrils and IDPs is linked to the poor reproducibility of experiments and batch variability within and among different laboratories. Additionally, dealing with *in vivo* fibrils could also worsen the analysis, increasing the complexity and the differences in extracted samples due to a different cellular environment and to a genetic variability. However, in the case of *in vitro* protein fibrils, the use of a multi-techniques approach is recommended for the investigation of these delicate specimens.

Two main aspects of amyloid fibrils and of IDPs, in general, can be characterized *in vitro* such as their structure and the kinetics of fibrillation. In Table 1.2, a clear picture of the techniques used for the investigation of IDPs and protein fibrillation is reported.

The typical cross-β structure of fibrils has been firstly detected using X-ray diffraction, which provides high-resolution measurements reaching the wavelength

typical of a C-C bond (1.5 Å) without limiting its applicability of low molecular weight samples. Unfortunately, the use of X-ray diffraction is limited since IDPs and fibrillar proteins do not crystallize or they form small ordered portions, respectively[106–109]. In fact, X-ray diffraction could be employed only for the characterization of small and ordered amyloid fragments, being conscious that the static structure is only one of the many projections of the dynamic protein structural state[107–109].

Table 1.2: Techniques used for the investigation of protein aggregates.

Techniques	Purposes
ssNRM and NMR X-Ray diffraction Cryo-EM Mass Spectrometry (HD exchange)	Atomic resolution structure
SDS-PAGE Size-exclusion chromatography (SEC) Dynamic Lyght Scattering (DLS)	Size distribution
Circular Dichroism (CD) Fourier Transform Infrared spectroscopy (FTIR) Raman spectroscopy	Secondary and tertiary structure
Transmission Electron Microscopy (TEM) Atomic Force Microscopy (AFM) Scanning Tunnel Microscopy (STM)	Morphological assay
Intrinsic fluorescence spectroscopy Dye-based fluorescence spectroscopy	Kinetics and tertiary structure

Only in the case of amyloid fibrils, being non-crystalline and insoluble, solid-state NMR (ssNMR) has become a method for the investigation of their structures instead of the classical NMR. In fact, the measurement of isotopically labelled fibril samples in milligrams quantity has become feasible. Amyloid fibrils usually show a relatively homogenous structures, resulting in high-quality ssNMR spectra. However, the high amount of sample required and working in solid phase, represent two important limitations for the employment of ssNMR[56,110–112].

The use of cryo-EM has become very common in this field since, differently to X-ray diffraction, the vitrification of the sample maintains it in a native-like state and low quantity of sample (approximately 0.1mg) is required without needing protein crystallization. However, particles are detected with a random orientation and spectra are affected by high levels of noise due to the limited electron doses for disfavouring radiation damage. In fact, only large complexes and large proteins (>50kDa) could be investigated with a good signal-to-noise ratio. Thus, obtaining a homogenous sample vitrification and deposition, improving image resolution as well as increasing also the throughput of the technique still represent major challenges dealing with cryo-EM[56,70,113,114]. In Table 1.3 the advantages and disadvantages of these techniques are summarized.

Table 1.3: Advantages and disadvantages of the atomic resolution techniques.

Techniques	Advantages	Disadvantages
ssNRM	High resolution Easy isotopic labelling	Solid state samples High amount of sample
X-Ray diffraction	Well developed High resolution No sample size limitation	Only crystals Static crystalline structure
Cryo-EM	Native-like structure 3D structure Low amount of material	Sample size limitation Low resolution (3.5 angstrom) Dependent on the sample preparation Costly EM equipment

Despite, they have been considered as golden standard for the investigation of protein structure in general. However, for the case of amyloid fibrils and, more in particular of IDPs, those techniques could not be used or could not be employed alone. Noteworthy, other techniques characterized by a lower structural resolution, are able to investigate a biomolecule without size limitation and in their native-like structure. As an example, the enrichment typical of amyloid fibrils can be easily detect by circular dichroism (CD) and by Fourier Transform Infrared (FTIR) spectroscopy, observing a change in ellipticity and an enhancement of the 1625-1630 cm^{-1} bands of amide I, respectively[115–117]. Raman spectroscopy is widely used for the

investigation of protein secondary and tertiary structure, especially with visible or infrared light. Unfortunately, CD provides only qualitative information on protein secondary structure (far-UV) and tertiary structure (near-UV) and FTIR usually requires high amount of samples since water completely dominate the absorption spectra. A typical Raman spectrum of a biomolecule, instead, contains a lot of multiple and overlapping contributions that could complicate the interpretation of the data, since vibrational contributions arisen from different chromophores could cancel out themselves, leading to a misleading interpretation of the phenomenon[118]. However, both techniques permits to study protein unfolding and secondary structure modification working with native proteins. The morphological characteristics of amyloid fibrils are commonly detected by TEM and AFM microscopy, while the size evolution from monomers towards intermediate specimens and fibrils could be followed by Dynamic Light scattering[119–122]. Dye-based fluorescence as well as intrinsic fluorescence spectroscopy are widely used approaches for studying protein fibrillation and protein interaction with ligands. Intrinsic fluorescence studies are limited by the presence of tryptophan residue, since it is the only aromatic residue, whose fluorescence signal is sensitive to solvent-effect. Differently, the most reported dyes employed in fluorescence are Thioflavin T (ThT), Congo Red, ANS and Nile Red, which are able to change their optical features when exclusively bound to fibrillar proteins[123–125]. Fluorescence spectroscopy is usually employed to follow the kinetics of aggregation and the autofluorescence of aromatic residues, especially tryptophan residue, is useful to indirectly get information on protein 3D arrangement. However, autofluorescence, inhomogeneous binding alongside fibrils, fibrils structure modification upon binding as well as competitive binding when another ligand is present, are considered as source of biases in the interpretation of fluorescence data[124].

In light of this consideration, optical and laser spectroscopies could be used to overcome the abovementioned limitations related to the atomic resolution techniques, but also to complement those results. Differently from cryoEM, ssNMR and diffraction techniques, proteins can be investigated in aqueous solutions without any size limitation[118]. However, to study such complex phenomena it is

important to build a solid multi-technique approach aiming to also to cancel out or reduce the biases of the sample preparation and those induced by the measurement protocol, while giving a solid interpretation of a phenomenon.

Thus, there is the need to find an alternative, less expensive, accessible technical approach to shed light on amyloid fibrils and on the monomeric IDPs structure.

In the next Chapter, the experimental techniques employed for the study of IDPs in their native and fibrillar state as well as general aggregation-prone proteins are presented.

Chapter 2

Experimental techniques

2.1 UV Resonance Raman (UVRR) spectroscopy

2.1.1 Overview of Raman spectroscopy

Raman spectroscopy is a technique employed in the study of vibrational and rotational dynamics of molecules[126,127]. When an incident electromagnetic radiation excites a molecule, it provokes the oscillation of its electron cloud, inducing a displacement of the charge density, which in turn provokes a change in the molecular dipole moment. Charges are accelerated and, therefore, radiate energy producing a scattered electronic radiation. The energy of the irradiated light depends on the interaction with the molecular system. If the scattered light has the same energy as the incident one, the process is "elastic", whereas it is "inelastic" when the scattered light has a shifted energy compared to the initial value. The elastic process is called Rayleigh scattering, whereas the inelastic one is called Raman scattering. The coupling between the electron cloud oscillation at frequency ω_0 and the vibrationally induced one at frequency ω_n give rise to an oscillation at a frequency $(\omega_0 \pm \omega_n)$, energetically shifted respect to the incident light.

In fact, Raman scattered light can be shifted to higher (anti-Stokes) or lower (Stokes) energy compared to the incident one. Stokes scattering are usually more frequent and stronger than anti-Stokes, since molecules are usually in their ground state before excitation.

2.1.2 Classical description

Raman effect can be described using the classical theory. The incident electromagnetic radiation oscillates at frequency ω_0 as described by equation (2.1):

$$E(t) = E_0 cos(\omega_0 t) \tag{2.1}$$

promoting the oscillation of the electric dipoles. The induced dipole moment depends on the molecular polarizability α and on the electric field E as described by equation (2.2):

$$\boldsymbol{\mu} = \alpha \boldsymbol{E} \tag{2.2}$$

Each component can be expressed as:

$$\mu_\rho = \sum_{\sigma=x}^{z} \alpha_{\rho\sigma} E_\sigma \tag{2.3}$$

Where $\alpha_{\rho\sigma}$ refers to the component of the polarizability tensor. Thus, eq. (2.3) becomes:

$$\mu_\rho = \sum_{\sigma=x}^{z} \alpha_{\rho\sigma} E_{0\sigma} cos(\omega_0 t) \tag{2.4}$$

Assuming harmonic vibrations, the time dependence of the nuclear normal mode coordinates Q_k can be expressed as:

$$Q_k = Q_{0,k} cos(\omega t) \tag{2.5}$$

The perturbation of polarizability due to molecular vibrations can be expressed as a Taylor series with respect to the normal mode of vibrations Q_k. Truncating it at the first degree, we obtain:

$$\alpha_{\rho\sigma} \cong \alpha_{0,\rho\sigma} + \sum_k \left(\frac{\partial \alpha_{\rho\sigma}}{\partial Q_\kappa}\right) Q_\kappa = \alpha_0 + \alpha_\kappa' Q_\kappa \tag{2.6}$$

Where α_0 represent the static polarizability and it is representative of Rayleigh scattering and the change in the polarizability along a certain coordinate κ is represented by the term $\alpha_\kappa' Q_\kappa$.

If we combine eq. (2.1) and (2.2), we obtain:

$$\mu_\rho = \sum_{\sigma=x}^{z} \alpha_{\rho\sigma} E_0 \cos(\omega_o t) + \frac{1}{2}\left(\frac{\partial \alpha_{\rho\sigma}}{\partial Q_\kappa}\right)_0 E_0 Q_{\kappa,o} \{cos[(\omega_0 + \omega)t] + cos[(\omega_0 - \omega)t]\} \tag{2.7}$$

We can re-write this equation in a more compact one, like so:

$$\mu = \mu^{Rayleigh} \cos(\omega_o t) + \mu^{Raman} cos[(\omega_0 \pm \omega)t] \quad (2.8)$$

As reported in the eq. 2.8, the interaction between molecules and electromagnetic radiation produces an induce dipole which is able to oscillate at three different frequencies. The first dipole oscillates at the same frequency of the incident electric field and accounts for the Rayleigh scattering. The second term comprises a dipole that oscillates at a frequency shifted respect to that of incident radiation, taking into account the natural frequency ω of the electronic oscillators. If the dipoles radiate photons with higher energy compared to the incident one, we assist to an anti-Stokes process $(\omega_0 + \omega)$; alternatively, if dipoles radiate photons with a reduced energy compared to ω_0 we observe a Stokes process $(\omega_0 - \omega)$ (Fig. 2.1).

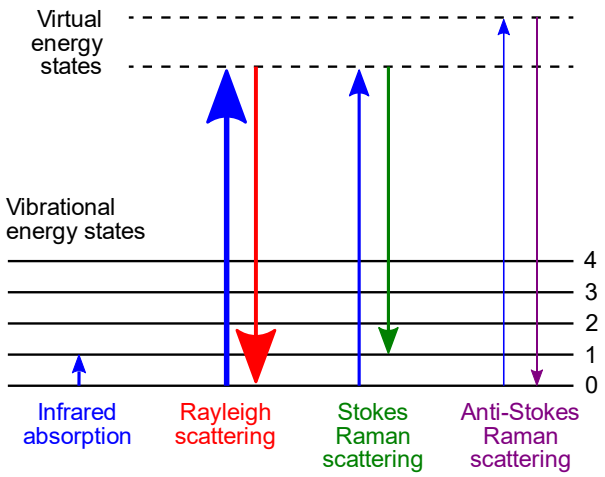


Fig. 2.1: Quantum mechanical diagram describing the vibrational excitation of a medium.

Interestingly, if the first term of Taylor expansion of polarizability, i.e. the derivative of polarizability, is zero, then the vibration is not Raman active but only elastic scattering is possible. Thus, Raman selection rule implies that a vibration is Raman active if there is a change in polarizability $\left(\frac{\partial \alpha_\rho}{\partial Q_k}\right)$.

For a given vibrational mode, the total intensity of the Raman scattering will be:

$$I(\theta) = D(\omega_0 \pm \omega)^4 I_0 \left(\frac{\partial \alpha_{\rho\sigma}}{\partial Q_k}\right)_0^2 sin^2\theta \quad (2.9)$$

Where I_0 and θ are the intensity of the incident light and the angle between the direction of the electromagnetic field and the induced dipole. Noteworthy, the intensity of the Raman spectrum depends on the modification in the polarizability

of the molecule by the vibration and on the fourth-power of the frequency of the incident radiation.

2.1.3 Quantum mechanical approach

While the classical model provide a semi-complete characterization of the Raman process, it does not provide adequate previsions regarding Raman band intensities. To overcome this limit, it is pivotal the introduction of the quantum mechanical description of the process. We adopt the so-called semi-classical approach, where molecular energy is treated as quantized, while electromagnetic contribution is treated classically as explained in the previous section. In quantum mechanics, a molecule can absorb or emit only quantized amount of energy, which correspond to the energy difference of two energy states. If we consider the incoming radiation as a perturbation of the initial energy state $|i>$ and we define as $|f>$ the final state, we can obtain that the induced dipole can be calculated as:

$$\mu_{fi} = < \psi_f | \boldsymbol{\mu} | \psi_i > \tag{2.10}$$

Where $\boldsymbol{\mu}$ is the permanent dipole moment operator, ψ_f and ψ_i are the final perturbed and the initial wavefunctions of the molecule. In order to obtain a correct definition of the total Raman intensity $I(\theta)$, we have to substitute the polarizability term $\left(\frac{\partial \alpha_{\rho\sigma}}{\partial Q_k}\right)_0$ with the Raman scattering tensor $[\alpha_{\rho\sigma}]_{fi}$. As an example, the polarizability tensor describing the transition between the ground state and the first excited state can be expressed as:

$$[\alpha_{\rho\sigma}]_{10} = \frac{1}{hc} \sum_{m,k} \left\{ \frac{<g_1|M_\rho|m_k><m_k|M_\sigma|g_0>}{v_{mk} - v_0 - v_{exc} + i\Gamma_{mk}} + \frac{<g_1|M_\sigma|m_k><m_k|M_\rho|g_0>}{v_{mk} - v_1 - v_{exc} + i\Gamma_{mk}} \right\} \tag{2.11}$$

The terms g_0, g_1 are the vibrational wavefunctions of the ground electronic level corresponding to the starting 0 and the final 1 vibrational levels, respectively. The m, k are the vibrational wavefunctions of the virtual states, Γ_{mk} is the width of the band associated to the $|m, k>$ vibrational state and M_ρ, M_σ are the electron dipole moments, while v_{exc} is the frequency of the incident electromagnetic beam.

2.1.4 Resonance Raman spectroscopy

The total Raman intensity differs between spontaneous and Resonance Raman scattering. Both are characterized by a transition that involves a process between the vibrational states of the electronic ground state; however, in conventional Raman scattering the electronic excited states are virtual intermediate states, whereas in the Resonance Raman scattering they are real ones. In particular, when v_{exc} matches the frequency required for an electronic transition of a molecule, eq. (2.11) becomes:

$$[\alpha_{\rho\sigma}]_{10} = \frac{1}{hc}\sum_k \left\{ \frac{<g_1|M_\rho|m_k><m_k|M_\sigma|g_0>}{v_{mk}-v_0-v_{exc}+i\Gamma_{mk}} \right\} \quad (2.12)$$

Since the second term become negligible, the polarizability becomes dependent on the vibrational quantum number of the electronic excited state k. Intuitively, if the incident radiation is within an absorption band, mainly the electron involved in this electronic transition start to oscillate at their natural frequency, causing an increasing in the Raman scattering intensity.

In particular, when $v_{exc} \rightarrow v_{mk}$, the tensor $[\alpha_{\rho\sigma}]_{10}$ becomes greater compared to that of spontaneous Raman scattering and the Raman intensity associated to the particular normal mode is outstandingly enhanced up to 8 orders of magnitude. In fact, if the incident electromagnetic energy is close to the energy required for the excitation of a specific molecular portion, the associated Raman signal could be orders of magnitude higher compared to the others belonging to non-resonant vibrations. The width Γ_{mk} is inversely proportional to the lifetime of the vibrational excited state $|m,k>$, thus taking into account that the effect of the electromagnetic perturbation is related to the lifetime of the excited state.

Due to this characteristics, the advantages of Resonance Raman are twofold: firstly, the sensitivity of the Resonance Raman scattering is enhanced compared to the conventional Raman spectroscopy and, secondly, it is possible to selectively analyse the contributions arising from different molecular chromophores of a heterogeneous sample or eventually to enhance the Raman signal of one component in a mixture. In few words, Resonance Raman is characterized by an increased selectivity and sensitivity, with the possibility to furtherly enhanced the intensity of the Raman

signal. Taking advantage of the Born-Oppenheimer and Condon approximation, we can furtherly simplified eq. 2.12 as:

$$[\alpha_{\rho\sigma}]_{10} = < g|M_\rho|m >_0 < m|M_\sigma|g >_0 \sum_k \left\{ \frac{<1|k><k|0>}{v_{mk} - v_0 - v_{exc} + i\Gamma_{mk}} \right\} \qquad (2.13)$$

Looking at this equation, we can clearly see that the tensor is non-zero:

- if $< g|M_\rho|m >_0$ and $< m|M_\sigma|g >_0$ are both non-zero ;
- if $< 1|k >< k|0 >$ are non-zero for few values of k.

2.1.5 UV Resonance Raman spectroscopy

UV Resonance Raman (UVRR) spectroscopy is used to probe the vibrational modes of molecular groups or chromophores in a system in the UV region, since most of the organic samples exhibit a strong absorption transition in the UV range[118]. In particular, protein secondary structure, aromatic side-chains as well as nucleic acids vibrations are effectively probed in this energy range. As an example, the UV absorption spectra of hen egg white lysozyme (HEWL) and of tryptophan, tyrosine and phenylalanine are reported in Fig. 2.2. As clearly shown, the highest absorption peak is located at 190-200 nm, where the π-π* transition of the amide peptide bonds is efficiently stimulated giving rise to vibrational peaks characterized by frequency position, intensity and bandwidth which depend on the structural conformation of the system. In particular, this electronic transition results in the excitation of the C=O stretching and of the out-of-phase C-N stretching[118]. Additionally, a second shoulder in the absorption spectra of HEWL is clearly visible at ~230 nm, where the electronic transitions of tryptophan (Trp) and tyrosine (Tyr) residues are promoted. Using with this excitation wavelength, UVRR can efficiently enhance the vibrational modes of aromatic side-chains and, more interestingly, probe and follow the modification of their microenvironments[118,128]. In fact, Trp and Tyr Raman vibrational peaks are good reporters of aromatic residue solvent-exposure, of the formation of non-covalent interactions they made with other residues, with ligands or solvent. What makes the (200-250)nm region particularly appealing is the lack of molecular fluorescence background which simplify the data analysis and the interpretation of the spectra

collected. Differently, excitation wavelength higher than 250 nm are used to stimulate the ^1L_a and ^1L_b transition of Trp (280 nm), Tyr residues (230 and 275 nm) and Phe residues (258 nm), but in this case a strong fluorescence background hides the protein spectra. In particular, the maxima of Trp absorbance bands are located at 220 nm and 280 nm, while the tyrosine one at 230 and 275 nm[118,128].

From Fig. 2.2, we can easily see that the absorption spectrum of HEWL is strongly enhanced compared to those of aromatic residues since it contains 6 Trp, 3 Tyr and 3 Phe residues, but it shows similarity to the Trp residue spectrum.

In light of the diverse UV absorbance spectra, it is possible to tune the UVRR excitation wavelength in order to mostly probe one of this chromophores, by enhancing the corresponding electronic transitions. However, due to the high molecular extinction coefficient, the UVRR spectrum of a protein containing Trp

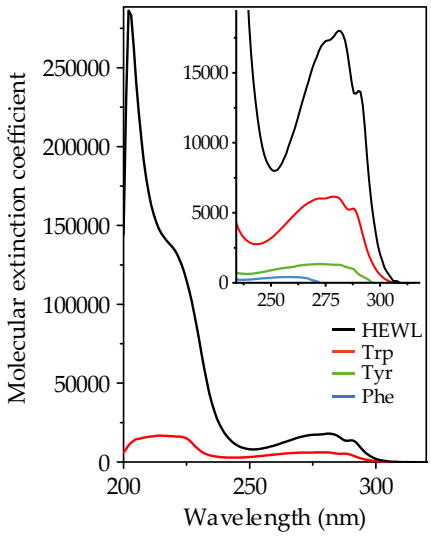


Fig. 2.2: UV absorption of HEWL (depicted in black) and tryptophan residues (Trp, in red) collected in the region 200-310 nm. The inset shows the absorbance spectra of HEWL (in black), of Trp (in red), of tyrosine (in green) and of phenylalanine (Phe, in light blue) in the region 240-310 nm. As clearly shown in the inset, Trp residues are the most strong absorbers in a protein, followed by Tyr and Phe.

residues, collected using any wavelength of the UV range, will present a spectrum completely dominated by Trp vibrational peaks. Interestingly, in the spectral region within 240 and 250 nm, the minima of the absorbance spectra of both aromatic residues and HEWL are located. At this particular energy, the aromatic residues

absorbances are less enhanced, but even far from the resonance conditions, a similar excitation wavelength allows to get information on peptide bond vibrations without requiring a deep UV excitation wavelength close to 200 nm.

The most informative portion of the UVRR spectra of a protein is the region ranging within 800-1800 cm^{-1}, where the peaks addressed to the vibrational modes of aromatic residues and peptide bonds are found. the variation of the chemical environment around the molecular groups of interest and to investigate protein secondary structure modifications[118,128]. In Fig. 2.3 two representative UVRR spectra obtained at 244 nm (a) and 228 nm (b) are presented.

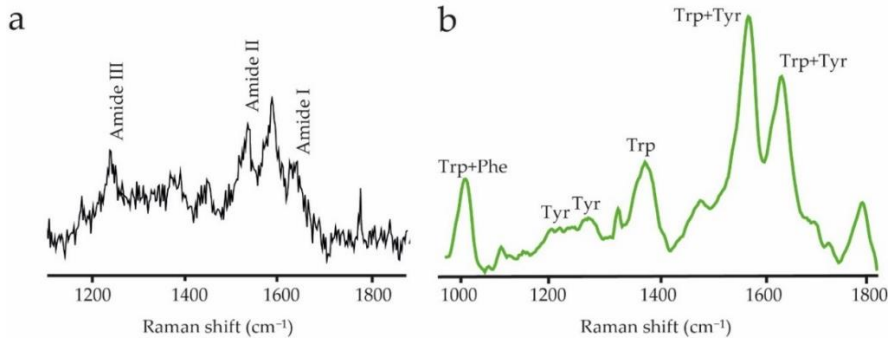


Fig. 2.3: UVRR spectra of 2 biomolecules obtained at (a) 244 nm and (b) 228 nm. The bands addressed Amide I, II and III are well enhanced at 244 nm, while the aromatic amino acids vibrational bands are clearly visible at 228 nm.

As explained before, the 228 nm-UVRR spectra is able to enhance mainly the vibrational modes of aromatic chromophores, while at 244 nm, the intensity of the aromatic bands are reduced and amide bands are more visible. In particular, we observed the emergence of Amide I (Am-I, at ~1650 cm^{-1} consisting of C=O stretching and of out-of-phase C-N stretching), Amide II (Am-II, at ~1550 cm^{-1} consisting of an out-of-phase combination of CN stretching and NH bending) and Amide III (Am-III, ~1200-1340 cm^{-1}, a combination of CN stretching and NH bending) signals. As widely reported in literature, the intensity and energy variation of these bands is used to determine protein secondary structure. In Table 2.2, the position of the amide bands is reported with the assignment of the secondary structure they are related to. In the same region of interest, the aromatic side-chains vibrational bands could be detected

and, depending on the protein and on their microenvironments, they could appear energy shifted of with a modulated intensity.

In Table 2.3, the position of the aromatic side-chains bands is reported with the corresponding assignment (readapted from[128]).

Table 2.2: Assignment of the UVRR bands to peculiar secondary structure components.

	Secondary structure		
	Raman shift (cm-1)		
	α-helix	β-sheets	Unordered
Amide I	1650	1667-1674	1677-1680
Amide II	1550	1564	1554
Amide III	1339 1293 1276	1306 1259 1225-1235	1314 1293 1270 1240-1245
Cα-H		1354 + 1406	1374 + 1397

Table 2.3: Assignment of the UVRR bands to peculiar aromatic side-chains vibrational mode.

Phenylalanine		
Vib. Mode	*Shift (cm-1)*	*Assignment*
F5	1003	symmetric ring stretching
F4	1030	in plane CH bend
F3	1207	phenyl-C stretching
F2	1585-1595	in-plane ring stretching
F1	1604	in-plane ring stretching
Tyrosine		
Vib. Mode	*Shift (cm-1)*	*Assignment*
Y6-Y5	830-850	Fermi resonance doublet
Y9a	1178	in-plane CH bend, CH-C stretch
Y7	1208	totally symmetric stretch
Y8b-Y8a	1601-1615	anti-symmetric/symmetric in-plane ring stretching
Tryptophan		
Vib. Mode	*Shift (cm-1)*	*Assignment*
W17	870-880	indole ring vibration with NH bending
W16	1006	symmetric pyrrole out-of-phase breathing vibration
W7	1345-1360	Trp Fermi doublet - pyrrole ring vibration
W3	1550-1553	symmetric phenyl ring mode
W1	1615	phenyl ring vibration

ù

2.1.6 Advantages and limitations of UVRR spectroscopy

The sensitivity and the high specificity that characterizes Resonance Raman (RR) spectroscopy provide the possibility to investigate the changes induced in the structure of molecular systems, leading to be adapted to a wide range of applications, including biological systems. Lots of systems are rich in spectral features and present a higher degree of spectral complexity, which results in overlapping bands and sometimes in misleading information. The interpretation of the spectral features could be made easier by disentangling the contribution of peculiar vibrational modes of interest from the whole spontaneous Raman spectrum. This could be made only by Resonance Raman (RR) spectroscopy, with which only the vibrational modes that are in resonance with the incident electromagnetic energy are enhanced. In this view, RR allows to simplified the complexity of the conventional Raman spectrum, being highly sensitive and with an increased limit of detection. Moreover, differently from infrared spectroscopy, Resonance Raman spectroscopy as well as the spontaneous Raman spectroscopy permits the investigation of samples in aqueous solution due to the relative low polarizability of water, without the requirement of the hydrogen-deuterium exchange. The infrared absorption of water is very intense and the spectral region where the main biological features fall in (1400-1800 cm^{-1}) are completely hidden by the OH bending signal. Moreover, RR required a reduced sample concentration compared to both conventional Raman and infrared (≥3 mg/ml) spectroscopies[129,130]. In the case of Surface Enhanced Resonance Raman spectroscopy (SERS), samples could be analysed with a nanomolar concentration[131]. Perhaps RR presents several advantages compared to classical vibrational spectroscopies and biophysical techniques, several limitations have to be considered. First of all, the RR cross-section is influenced by a self-absorption contribution which occurs especially when a part of the scattered light is somehow reabsorbed by the sample[132]. This phenomenon could reduce the intensity of the overall spectra, affecting mostly the time required to get a spectrum with a good signal-to-noise ratio. Another limitation, generally encountered dealing with biological systems, is the photodegradation phenomenon due to possible photoreactions occurring particularly in resonance conditions and with a low content of water[133].

Photodegradation phenomenon could be modulate or overcome by several tricks and strategies that will be properly discussed in the following sections and chapters.

2.1.7 Experimental setup

In the previous paragraph, the potentiality of the UVRR technique in the investigation of matter has been presented. Nevertheless, the lack of continuous wave tuneable excitation sources results in a limited exploitation of the full UVRR region, thus limiting the investigation of the outer electronic excitations in matter (i.e. up to 10-15 eV) by selectively probing specific orbitals and bands.

- **Synchrotron radiation source**

The BL10.2-IUVS beamline at Elettra Synchrotron facility in Trieste has developed an optical setup that exploits a tuneable UV synchrotron radiation (SR) source able to collect UVRR spectra from solids, gases and liquids[134]. Fig. 2.4 shows the technical arrangement of the instrument.

A 32 mm period undulator generates a linear polarized SR with an energy between 4.4 and 11 eV, thus corresponding to wavelength ranging from 113 nm to 280 nm. A specific advantage of the presence undulator is the relevant reduction of the total on-

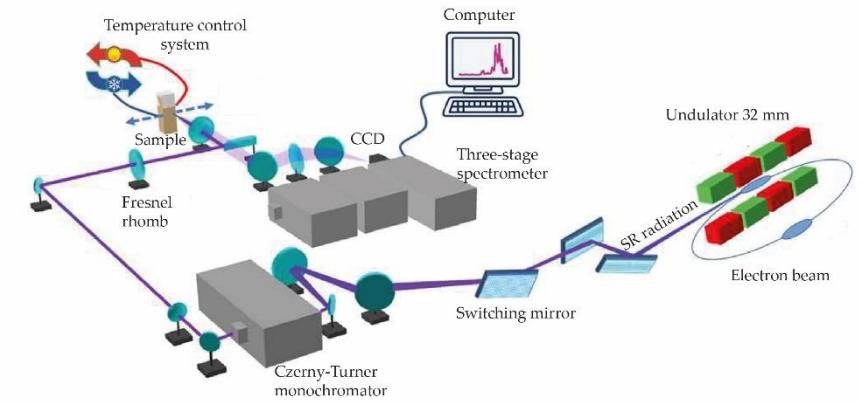


Fig. 2.4: IUVS experimental set-up.

axis power density that is obtained with no penalty on the useful photon flux in the first harmonic of the undulator emission spectrum. The beam coming from the source is cleaned from the higher order undulator harmonics by a gold coated GLIDCOP mirror. This internally water-cooled mirror is able to deviate the photons in the vertical plane with an angle of 60°, while a second externally water-cooled silicon mirror is used to bring back the beam parallel to the floor. Then, a silicon switching mirror placed at 3.5 m from the water-cooled silicon mirrors and a system of UV enhanced coating mirrors (reflectivity of 90-92% and 185-250 nm wavelength) drive and focus the SR into the entrance slits of the monochromator. The transport system delivers to the monochromator a UV beam of 10 mW (measured with an excitation wavelength of 270 nm), with a typical bandwidth of $\Delta\lambda/\lambda \cong 0.01$, corresponding to ~350 cm^{-1}. The SR is monochromatized by a Czerny-Turner (CT) monochromator (Acton SP2750 produced by Princeton instruments), equipped with three exchangeable flat holographic gratings (1800, 2400, and 3600 lines/mm). The monochromator provides a wavelength resolution of 0.012 nm at 270 nm (i.e. 1.6 cm^{-1}). Depending on the selected wavelength, the maximum beam power after the monochromator, while keeping opened the monochromator slits to 100 nm, is presented in Fig. 2.5. Once it has been monochromatized, the UV beam is collimated by a lens and transported to the Raman scattering optical system. At this point, the incident beam of typically few mm^2 is focused on the sample and the diffused light is collected in back-scattering configuration through plano-convex lenses and UV broadband mirrors and transported into the entrance slits of the spectrometer. The main advantage working in back-scattering geometry consists on a simple alignment procedure, because the same lens acts both as focusing and collecting lens. The three-stages spectrometer (TriVista 557, Princeton Instruments) is used to analyse the Raman signal. Each stage is equipped with a two holographic gratings (1800 and 3600 grooves/mm) optimized for both UV and visible radiation. The highest wavelength that can be obtained depends on the CT mechanical rotation capability and results to be 417 nm (2.98 eV) for the 3600 lines/mm grating and 833 nm (1.49 eV) for the 1800 lines/grating. Finally, the scattered photons are detected by a Peltier-cooled UV-enhanced CCD camera. Cyclohexane is used to calibrate the spectrometer. Additionally, UV

Raman spectra could be collected in parallel polarization (VV) or depolarized orthogonal (HV) configuration by inserting in the optical path a Fresnel Rhomb Retarders (Half-Wave Retardance with broader wavelength range) and a polarizer at the

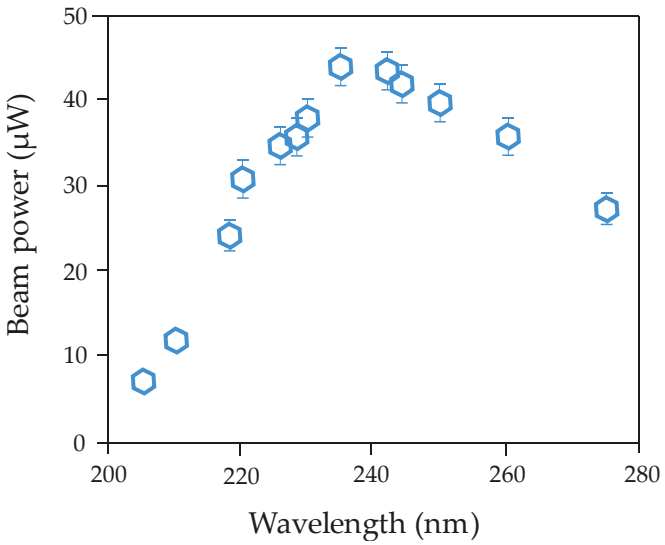


Fig. 2.5: Maximum beam power measured at the exit of the IUVS monochromator in function of the SR wavelength.

entrance of the spectrometer. Due to the low power radiation incoming to the sample, we could investigate also biological samples and dyes which could be affected from photodegradation. In order to minimize this aspect, the sample holder is horizontally oscillated.

- **Table-top laser (244 nm)**

In addition to the SR source, a table-top laser (244 wavelength) is also available for fixed-wavelength Raman experiment. The setup used for the collection of the scattered light is the same explained in the previous paragraph. The energy resolution exploited by the three CT stages working in additive mode reaches an energy resolution of 0.68 cm^{-1} with a maximum incident beam power of several mW's. Since the investigation of biological samples does not require higher beam power, the incoming radiation is usually filtered.

2.1.8 UVRR limitations are overcome by the use of other techniques

UVRR spectroscopy results are complemented by those obtained by other techniques. As an example, even if 244 nm excitation wavelength provided a reduction of the aromatic amino acids cross-section, increasing that of amide bands, there is the need to verify and complemented these data with the results obtained by FTIR analysis. In fact, unfortunately, the experimental setup available at IUVS does not provide a sufficient beam power in order to obtain a good signal-to-noise ratio and a decent spectral resolution in the case of complex biomolecules.

At the same time, the intensity of the UVRR aromatic residues-related bands is a valuable marker of the aromatic ring microenvironments condition, probing the non-covalent interactions where they are involved such as hydrogen bonding, and the modification of the microenvironments' polarity and hydrophobicity[135,136]. The study of aromatic residues is crucial for problems such as protein fibrillation or ligand binding, since they are generally involved both in protein self-assembly and in the interaction with an aromatic ligand. However, the intensity variation of an aromatic band peaks can be related to multiple factors. As an example, the increasing of a peak intensity could be caused by the increasing of the microenvironments' hydrophobicity around the ring or to the establishment of a strong H-bond. In the case of Trp-based proteins, UVRR spectroscopy can be combined to intrinsic fluorescence spectroscopy in order to get major insights in the characterization of the Trp microenvironments' hydrophobicity[137]. In fact, the latter techniques shows an outstanding sensitiveness to Trp residues fluorescence: Trp maximum peak shifts in energy and quenches its intensity depending on the solvent exposure and polarity[137]. Hence, Trp peak redshifts upon increasing solvent polarity, probably due to hydrogen bond interactions.

For this reason, UVRR spectroscopy and intrinsic fluorescence spectroscopy could be combined and are particularly useful for the characterization of the tertiary structure of the Trp-based proteins. However, despite its sensitiveness to aromatic residues fluorescence, intrinsic fluorescence spectroscopy does not provide crucial information in the case of Trp-lacking proteins. In fact, tyrosine and phenylalanine

residues are completely insensitive to solvent-effect, resulting in a possible misleading interpretation of the observed effect. In the last cases, UVRR spectroscopy demonstrates to be one of the most useful techniques able to probe the state of these aromatic residues, carrying more information regarding the phenomenon studied.

2.2 Fourier Transform Infrared spectroscopy

The interval of the electromagnetic spectrum between visible and microwave regions is the InfraRed portion. The IR region is conventionally divided into three spectral regimes, the name of which reflects their distance from visible light:

- Near-Infrared (NIR), from 14000 to 4000 cm^{-1} (0.8-2.5 um)
- Mid-InfraRed (MIR), from 4000 to 400 cm^{-1} (2.5 – 25 um)
- Far-Infrared (FIR), from 400 to 10 cm^{-1} (25-1000 um).

The energy of the IR photons ranges between 1.25 meV and 1.7 eV, that corresponds to the energies needed to induce vibrational excitation of covalently bonded atoms. At temperature higher than 0 K, these bonds could vibrate in a variety of ways. However, IR light could be absorbed by a molecular vibration only when the energy of the incoming light matches perfectly with that required for the stimulation of a molecular vibration within the same electronic state. The frequency and the extent of absorption is influenced by inter- and intramolecular effects. In this view, band position, bandwidth and absorption coefficient may be used to get insight on the structure of the molecules investigated. From here on, we focus only on the MIR region since the transition energies of the systems investigated lies in this spectral range.

Due to its sensitiveness to proteins' secondary structure and biomolecules structure in general, InfraRed spectroscopy (IR) in the MIR region has been considered for decades a valuable tool for studying these systems[117]. The most interesting spectral region dealing with proteins ranging within 1200 cm^{-1} to 1800 cm^{-1}, where amide bands are clearly visible. C=O stretching vibrations and a minor contribution of the our-of-phase CN stretching vibration, CCN deformation and in-plane NH bending,

give rise to Amide I vibration at approximately near 1650 cm^{-1}. Amide I is used as a reporter of the proteins secondary structure arrangement, depending mainly on protein backbone structure. Transition dipole coupling theory explains the oscillating charges in the dipole approximation: in particular, when the dipoles of the neighbouring amide groups start to resonantly oscillate, they may couple depending on their relative orientation and distance. The way protein oscillators absorb depend of the backbone structure: α-helix give rise to band at ~1655 cm^{-1}, anti-parallel β-sheets present a strong band at 1630 cm^{-1} and a weaker one approximately at 1685 cm^{-1} and, finally parallel β-sheets present two bands at 1625 cm^{-1} and 1695 cm^{-1}. The position of the β-sheets band depends strongly on the number of b-strands forming the sheets. As a rule of thumb, when sheets are composed by a large number of strands, the position of the characteristic band is shifter to lower wavenumbers. Obviously, when parallel β-sheets present several strands or vice versa, the difference between parallel and antiparallel β-sheets could be less pronounced. The NH in plane bending and the CN stretching vibration give rise to the amide II band, which absorbs approximately at ~1550 cm^{-1}. Both Amide I and II are less affected by the side-chains absorption bands, even if small contributions are clearly detectable in the low energy flank of both bands. Additionally, also side-chain absorption band could be detected in this region of interest, even if these bands may deviate from their absorption bands depending on their microenvironment. Proteins provide a special environment due to the modulation of strength and polarity of bonds.

As a recap, a schematic table reporting the assignment of amide I positions to secondary structure is reported (Table 2.4):

Table 2.4: Assignment of amide I band positions to secondary structure

Amide I	
Wavenumber (cm^{-1})	*Assignment*
1600-1620	Side-chains (tyrosine)
1620-1635	β-sheets
1635-1648	Unordered structure
1648-1658	α-helix
1659-1670	Unordered structure
1671-1680	β-turn
1681-1697	β-sheets

2.2.1 Classical description

A mechanical description of IR phenomenon can be derived studying the simple case of a diatomic molecule: the atoms are modelled as two masses m_1 and m_2 connected by a spring (i.e. interatomic bond) and free to move only along the molecular axis[117,138,139]. If the system is perturbed from equilibrium, the masses may be displaced by a factor X_1 and X_2 from their original positions. In the case where $(X_2 - X_1)$ is the variation of the bond strength respect to the equilibrium value, the spring tends to recover their initial position through an elastic force F that counteract the external force. Applying the Newton's law, we can derive the following equations (2.14):

$$F(X_2 - X_1) = m_1 \frac{d^2 X_1}{d^2 t} \quad \text{and} \quad F(X_1 - X_2) = m_2 \frac{d^2 X_2}{d^2 t} \tag{2.14}$$

These differential equations can be solved by using an expression of the displacement as:

$$X_{1/2} = A_{1/2} \quad \cos(2\pi\nu t + \delta) \tag{2.15}$$

Thus, the mass oscillate following a simple harmonic motion. Since ν and δ are supposed equal for X_1 and X_2, differentiating eq. (1.14) we obtained

$$F(A_2 - A_1) = -m_1 A_1 4\pi^2\nu^2 \quad \text{and} \quad F(A_2 - A_1) = m_2 A_2 4\pi^2\nu^2 \tag{2.16}$$

Both those equations can be expressed in function of the ratio A_1/A_2:

$$\frac{A_1}{A_2} = \frac{F}{F - m_1 4\pi^2\nu^2} \quad \text{and} \quad \frac{A_1}{A_2} = \frac{F - m_2 4\pi^2\nu^2}{F} \tag{2.17}$$

By eliminating the ratio A_1/A_2 we obtain a equation in function of ν ,which is solved if the frequency is:

$$\nu = \frac{1}{2\pi}\sqrt{F\left(\frac{1}{m_1} + \frac{1}{m_2}\right)} = \frac{1}{2\pi}\sqrt{\frac{F}{u}} \tag{2.18}$$

where $u = \left(\frac{1}{m_1} + \frac{1}{m_2}\right)$ is the reduced mass of the system. Looked at the equations, we can easily deduce the following points:

- the frequency ν increases when the force constant F increases;
- the larger the masses, the lower is the frequency of oscillation.

Hence, the strength of the bond between two atoms and the mass of the interacting atoms are linked to the frequency of vibration. An increase in bond strength leads to corresponding frequency increase and vice versa, while increase in the atoms mass corresponds to frequency decrease. In order to provide a classical description of the IR absorption, we could take advantage of the mechanical description above in order to extend it for a simplification of the electromagnetic approach. Noteworthy, we could deduce the same conclusion if we replace the system of masses m_1 and m_2 with two charge $+q$ and $-q$ displaced a distance R. The two charges schematize the displacement of the positive and negative charges of the sample induced by the external perturbation F because the molecular geometry variation affects the electron redistribution. In fact, the mechanical force F is the analogous of the infrared radiation. The oscillating electric field tends to continuously change the spacing between the positive and negative charges, i.e. it induces an oscillating electric dipole with a dipole moment $\mu(t) = qR$. Depending on the electric field energy, a dipole moment oscillation could activate a nuclear vibration which provokes a change in dipole moment. The more the electric dipole changes, the easier is the activation of a molecular vibration otherwise no vibrations are activated.

In general, to observe an IR absorption, two conditions have to be satisfied:

- The frequencies of light and the molecular vibrations have to coincide in order to observe an absorption;
- The oscillating electromagnetic radiation provokes a change in the dipole moment $\left(\frac{\partial \mu}{\partial R}\right) = q$ of a molecular vibration.

2.2.2 Quantum mechanical description

In quantum mechanics, the energy levels are quantized and separated by an energy $h\nu$. Equally to the description above, an IR photon carrying an energy $h\nu$ could be absorbed by the molecule, i.e. the oscillator, if the energy required for the vibration perfectly matches that of the photon. In this way, the system exhibits a transition between the ground state and the first excited state of the same electronic state.

In quantum mechanics, a diatomic molecule only have discrete energy levels with an energy corresponding to $E_N = (n + \frac{1}{2})h\nu$, where n is the quantum number. According to selection rule and in the limit of harmonic energetic potential, only transitions between adjacent energy levels are allowed (Δn=1). The approximation of harmonic potential is in good agreement with experimental data when the energetic transition occurs between the ground state and the first vibrational quantum level. Differently, the model fails for transition towards higher energetic levels. In those specific cases, the harmonic model has to be revised in order to take into account the repulsion and attraction of electron clouds at the extremes of the vibration.

According to Fermi's golden rule, the transition probability induced by the interaction with an electric field from the vibrational state |i> to the final state |f> of the same electronic ground state ψ_0 can be expressed by the Fermi's golden rule. Using the Born-Oppenheimer approximation, the nuclear wavefunctions φ_f and φ_i can be separated from the electronic wavefunction ψ_0 and the transition probability could be written as follows:

$$P_{i\to f} = \frac{2\pi}{\hbar}\left|< \psi_0\varphi_f|\boldsymbol{\mu}|\psi_0\varphi_i >\right|^2 \delta(E_f - E_i - h\nu) \tag{2.19}$$

where $\boldsymbol{\mu}$ is the dipole moment operator and the last term accounts for the conservation of the energy.

The matrix element $\left|< \psi_0\varphi_f|\boldsymbol{\mu}|\psi_0\varphi_i >\right|$ could be split in a nuclear and in an electronic part obtaining:

$$\left\langle\psi_0\left|\frac{\boldsymbol{\partial\mu}}{\boldsymbol{\partial R}(R_0)}\right|\psi_0\right\rangle\left(\frac{h}{8\pi^2 u\nu}\right)^{0.5} \tag{2.20}$$

Where the first is the electronic term, R_0 is the equilibrium position, the second term is the nuclear contribution and u is the reduced mass. Noteworthy, the electronic contribution accounts for the change in the dipole moment of the molecule. Interestingly, using quantum mechanics, we have obtained the same classical results. In order to observe an IR absorption, two selection rules have to be satisfied:

- The vibrational transitions occurs only when $f = i \pm 1$, thus they occurs only to the next vibrational level;

- IR absorption takes place only when the dipole moment of the molecule changes upon interaction with the electric field and the larger is the change, the stronger is the absorption.

Dealing with a polyatomic molecules with N atoms, we have that 3N-6 degrees of freedom are the fundamental vibrations for non-linear molecules, while linear ones have 3N-5 degrees of freedom. As derived before, only those vibrations producing a net change in the dipole moment of the molecule are IR active. However, the motion of two connected atoms in a large molecule cannot be treated as separated from the motion of all the other atoms, thus the vibrations of the two bonded atoms are coupled. If one bond contracts, the other could either contracts (in-phase) or expand (out-of-phase). There are barely two different vibrational modes: stretching and bending. The stretching could produce a contraction and an expansion of the bonds, and could be symmetric or asymmetric whether the vibrations are in-phase or not, respectively. Differently, the bending determines a variation of the angle between two bonds. To name a few, rocking, wagging and twisting are bending vibrations.

2.2.3 Lambert-Beer law

An IR spectrum usually represents the modification of the transmittance/absorbance versus incoming radiation wavenumber ($\bar{\upsilon} = 1/\lambda$). The amount of light transmitted by the sample carry information on the extent of energy absorbed at a certain frequency. The absorption of light by matter could be explained by the Lambert-Beer law. When the incoming light comes across a sample, part is absorbed, part is reflected and part is transmitted. The transmitted light is depleted by three physical phenomena:

- concentration of the material c, i.e. the amount of absorbing molecules;
- the optical path length of the material l, i.e. the distance that light has to travel in order to pass across and escape from the material;
- the molar extinction coefficient ε, which reflects the probability of absorption by the sample.

Considering the intensity of the incoming light as I_0 propagating in 1D and I the intensity of the transmitted light, we can observed experimentally that $I < I_0$ by a factor that depends on the peculiar sample investigated as follows:

$$-\frac{\mathrm{d}I}{I} = \varepsilon c(\mathrm{d}l) \tag{2.21}$$

Integrating the equation (2.21) we obtain:

$$I\ (l) = I_0 e^{-\varepsilon c l} \tag{2.22}$$

Thus, keeping fixed the geometrical condition of the sample, i.e. the path length, the transmitted intensity decreases exponentially with increasing the concentration of absorbers or with increasing the molar extinction coefficient. From the last equation, it is possible to calculate the amount of light absorbed by the sample as

$$A = log\frac{I_0}{I(l)} = -logT = \ \varepsilon c l \tag{2.23}$$

Where $T = \frac{I(l)}{I_0}$ is the transmission coefficient. The Lambert-Beer law shows that the absorbance is directly proportional to the concentration, to path length and to the extinction coefficient.

The Lambert-Beer law is valid only if few conditions are satisfied such as:

- the medium have to absorb homogeneously;
- the total absorbance of a multi-composed material is the sum of the absorption of these independent absorbers;
- light must not modify the molecular structure.

2.2.4 Experimental setup

Broadband emission, high collimation, polarization and pulsed structure are four main reasons that have induced the construction of the BL9.1-SISSI beamline at Elettra Synchrotron facility in Trieste[140]. Even if the synchrotron source does not provide more total flux compared to the common IR sources, the high collimated nature of the infrared synchrotron-radiation (IRSR) allows the illumination of very small areas, improving the signal-to-noise ratio and reducing also the acquisition time. The last aspects are particularly important in IR imaging and microscopy.

The radiation, extracted from the 9.1 bending magnet over acceptance angles of 70 mrad (H) and 25 mrad (V), enters the first vacuum chamber, equipped with a plane mirror M1 and an ellipsoidal one M2. The former is placed at 3.5 m from the source and collects the radiation from 65 mrad in the horizontal and 25 mrad in the vertical plane. Differently, M2 is at 0.1 m from M1 and it is used to focus the radiation beyond the shielding wall of the synchrotron hall to the intermediate focal point F1, which is a at a distance of 11.5 m from the source. A second vacuum chamber is placed at a distance of 14.50 m from the source, containing another plane mirror an and ellipsoidal one (M3 and M4, respectively). In order to minimize the aberrations in the transfer optics due to the wide emission angle, M3 and M4 are disposed in a symmetric optical configuration with respect to M1 and M2. Since SISSI beamline consists of two branches, one dedicated more to life science and the other to solid-state physics, an additional retractable plane mirror (M5) after mirror M4 at 15 m from the source was added in order to provide radiation also to the life science branch. Hence, if the solid-state branch works, M4 focuses the IRSR on F2 which is a chemical-vapor-deposited diamond window at 15.5 m and, then, the radiation pass through a Bruker IFS-66v Michelson interferometer coupled to a Hyperion-2000 infrared microscope, which can work in the spectral region from far-IR to visible in both reflection and transmission modes. The station is equipped with a He cryostat and bolometer. Differently, if the life science branch is operating, M5 deviates the IRSR beam by 90° and focuses it on a wedged 20 mm diamond window and IR radiation enters a dry- nitrogen-purged Bruker VERTEX 70 interferometer coupled to a Hyperion-3000 infrared microscope.

The reproducibility and the noise level of the spectro-microscopy set-up in the mid-IR range have been determined using a gold coated Si-wafer test sample placed on the microscope stage and measuring the ratio of two subsequent spectra acquired with a MCT detector in reflection mode by summing 128 scans at a resolution of 4 cm^{-1}. For apertures smaller than about 40x40 μm, IRSR becomes advantageous. Indeed, IRSR provides a good root mean square of the noise on the 100% line (RMS) down to 4x4 μm. In particular, in the whole spectral range using this aperture, a noise level with an excellent RMS ~ 0.3% and ~ 0.75%, respectively in the 2450-2550

cm^{-1} and 1100-1200 cm^{-1} spectral region, is obtained. At 2x3 μm the RMS ~ 1.5% obtained above 1000 cm^{-1} is still acceptable for several applications.

2.2.5 Advantages and limitations of FTIR spectroscopy

IR is a non-destructive, label-free and highly sensitive technique, which has been considered one of the classical tools used for the investigation biomolecules, without any limit of biomolecule dimension. Being label-free, IR technique could provide information of a wide variety of biological systems without requiring the presence of a probe that might alter the structure of the system. Apart from those advantages, the FTIR investigation of biological samples in aqueous media is more difficult since the characteristic bands addressed to proteins (especially amide bands), lipids, nucleic acids and carbohydrates are completely hidden by water absorption. In fact, water is a strong IR absorber and its absorbance spectrum is characterized by three main bands located at 1643, 3450 and 3600 cm^{-1}, respectively. Even if the extinction coefficient of water in the region 1700-1600 cm^{-1} is about 20 L $mol^{-1}cm^{-1}$ and that of proteins is higher than 30, water still cover the whole region with its contribution. In order to increase the signal-to-noise ratio, to perform a proper buffer subtraction and to reduce the water content in the sample, a high concentration of the biological sample is required together with longer acquisition time.

2.3 nanoFTIR spectroscopy

nanoFTIR spectroscopy resolves the spatial resolution limit of FTIR spectroscopy and microscopy, introducing a scattering-type scanning nearfield optical microscopy (s-SNOM, which returns an infrared image combined with a morphological profile. This technology is particularly suitable for the identification of the chemical composition of molecular thin films and nanocomposites and nanostructures, usually employed in nanoscale devices. nanoFTIR spectroscopy has become a

valuable tool for the characterization of protein fibrils, single viruses and biobased nanostructures.

2.3.1 Basic principles

As reported in Fig. 2.6, an infrared beam is focused onto a probing metalized tip, which actually acts as an antenna, concentrating the incident light at the apex of the tip.

The tip vehicles the incoming infrared radiation and collect also the scattered light due to the near field interaction between the apex and the sample. While collecting the scattering light from the sample, the tip also scans a set surface and the instrument merge the information regarding the amplitude and the phase of the scattered light compared to the reference radiation, providing a chemical and morphological map of the sample. The resolution of the near-field image depends only on the size of the tip apex, which is usually 20 nm and the maximum supported tip scan speed at higher resolution is 20 μm/s. Differently, the spectral resolution is of 6 cm^{-1}.

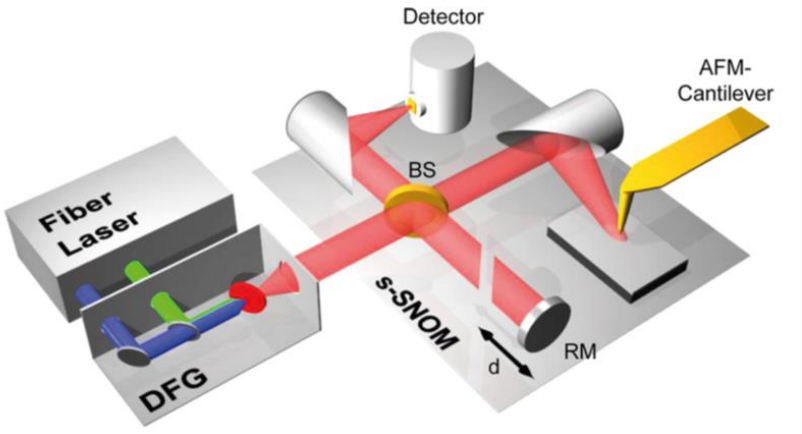


Fig. 2.6: nanoFTIR experimental setup. A fiber laser system emits two pulse trains at two different frequency, which are superimposed in the difference frequency generator (DFG) unit and directed onto a non-linear crystal, which in turn emits a mid-infrared beam. The infrared beam is directed onto the AFM tip and the backscattered light is analysed by the asymmetric Michelson interferometer, composed by a beamsplitter (BS) and a moving mirror (RM) used for the reference light, and a detector. The figure has been readapted from[141].

For our experiments, a NeaSNOM instrument (neaSpec GmbH) has been used. It combines an AFM and an asymmetric Michelson interferometer. The tip used to

collect the phase and the amplitude of the scattered near-field signal is coated with aurum and illuminated by a coherent broadband mid-infrared beam. The incoming radiation is generated by a difference frequency generator (DFG, lasnix.com), where two near-infrared, 100-fs pulse trains from a fiber-laser system (FemtoFiber pro IR and SCIR, toptica.com) are superimposed in a GaSe crystal. The mid-infrared source radiation could be tune in frequency within the range 700-2500 cm^{-1}, depending on the crystal orientation.

The sample is located in one of the interferometer arms and the detector allows to record both the amplitude s(ω) and the phase ϕ(ω) of the scattered light. The scattering coefficient is reported in eq. (2.31)

$$\sigma(\omega) = s(\omega)e^{i\phi(\omega)} \tag{2.31}$$

and linked the scattered electric field to the incoming one:

$$E_s(\omega) = \sigma(\omega)E_{inc}(\omega) \tag{2.32}$$

To extract the near-field signals, the detector signal has to be demodulated at a higher harmonic nΩ of the tip vibration at frequency Ω by the translation of the reference mirror using a piezo stage. By subsequent Fourier transformation of the interferogram, we obtain the near field spectra:

$$E_n(\omega) = \sigma_n(\omega)R(\omega)E_{inc}(\omega) \tag{2.33}$$

Where n is the demodulation order and $R(\omega)$ is the optical response of the instrument.

The $E_n(\omega)$ signal is normalized by a reference spectrum $E_{n,ref}(\omega)$, that could be obtained by recording the near-field spectrum of a spectrally transparent and flat sample of Si with a constant $\sigma_{n,Si}(\omega)$, yielding

$$E_{n,ref}(\omega) = const.R(\omega)E_{inc}(\omega) \tag{2.34}$$

thus, the near-field measured on the sample will depend on $\sigma_n(\omega)$ and, in particular, on the demodulation order n. This factor describes the near-field interaction between the tip and the sample, carrying information on the local dielectric function and on the refractive index. Additionally, the nano-FTIR absorption a_n is equal to the $\Im m(\sigma_n(\omega)) = s_n(\omega)\sin(\phi_n(\omega))$.

2.4 Intrinsic fluorescence spectroscopy

Fluorescence spectroscopy is a frequently used technique for indirectly studying the conformation and dynamics of the macromolecules[137,142,143]. Fluorescence is a form of luminescence, defined as the emission of light from any substance. It can be performed using a site-specific external fluorophore (such as a dye) or taking advantage of fluorescent molecule in the system under investigation: the first approach is called "extrinsic fluorescence", while the second one "intrinsic fluorescence". Since fluorescence spectroscopy has been widely applied to the study of protein folding, protein: ligand complexes and so on, an extensively used extrinsic fluorophore is thioflavin T and an intrinsic fluorophore is tryptophan residue. Fluorescence spectroscopy is ideal for probing micro- and nanodomains, where the fluorescent probe is embedded.

2.4.1 Basic principles

When light is absorbed by a molecule, an electron is promoted from the ground state to a higher excited state. To observe pure fluorescence, the excited electron has to be in a singlet state, thus it has to be paired (i.e. has an opposite sign) to the electron left in the ground state. The population of the excited state through the absorption is a very fast process (>10^{15} s^{-1}), provoking no significant displacement of nuclei (Franck-Condon principle) and the population of the excited state can be depleted in 10^{8} s^{-1} by fluorescence, via the emission of light as well described in Fig. 2.7.

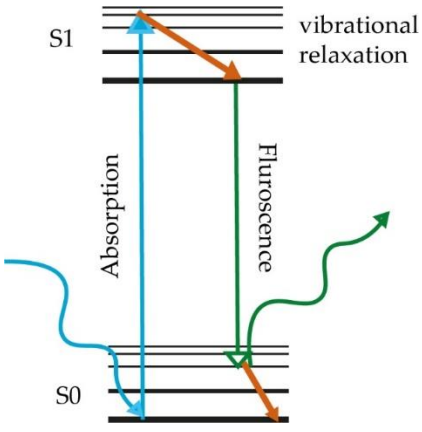


Fig. 2.7: Jablonski diagram of fluorescence effect.

The number of photons released per unit time $-\mathrm{d}N/\mathrm{d}t$ accounting for pure fluorescence is proportional to the number of excited fluorophores per unit time $N(t)$. Hence, we could derive that:

$$-\mathrm{d}N/\mathrm{d}t = k_F N(t) \tag{2.24}$$

Where k_F is the rate constant of fluorescence. Integrating the equation, we obtain:

$$N(t) = N_0 e^{-k_F t} \tag{2.25}$$

where N_0 are the number of excited fluorophores after excitation.

The Jablonski diagram is reported in Fig. 2.7 could illustrate the processed occurring between the absorption and the emission of light. The ground and the first electronic states are depicted as S_0 and S_1, respectively.

The fluorophore could be extrinsic or intrinsic and could exist in a different number of energetic level within the ground state. By absorbing light (generally in the ultraviolet region), the fluorophore could be excited to a higher vibrational level of the excited electronic state. From this high energy vibrational level, the molecule relaxes by a non-radiative *internal conversion* (10^{-12} s) to the lower energy vibrational state of the excited electronic state S_1. The return to the ground electronic state occurs by the emission of light and the molecule is left in a high vibrational level of the electronic ground state, letting it to decay to the original vibrational level.

The probability of the transition $S_1 \rightarrow S_0$ depends on a change in the dipole moment during the transition and on the energy difference between the two states.

This simple description has to be integrated by listing all the possible processes and phenomena that could take place simultaneously to fluorescence emission, contributing to modify the k_F. The Jablonski diagram could be integrated taking into account these phenomena as in Fig. 2.7. In fact, the excited state can get depopulate by radiative and non-radiative ways. As an example a molecule can lose its energy by energy transfer to another absorbing species of by collision with other species. Phosphorescence is another mechanism of emission due to the conversion of the molecule in S1 to the first triplet excited state T1 when the vibrational energies of the two excited states overlap. Vibrational relaxation could be another effect to take into

account if the molecule is not rigid, namely a relaxation where the molecule lose partially its energy by modifying its conformation.

Experimentally, the rate constant k_{exp} is actually replaced by the sum of the rate constant k_i off all the processes depopulating the excited state, including k_f as:

$$k_{exp} = \sum_i k_i \tag{2.26}$$

The fluorescence quantum yield ϕ, i.e. the ratio between the number of emitted photons N_F and the absorbed ones N_A, is lower than one and is attenuated by the quantity $\frac{k_{exp}}{\sum_i k_i}$. Accordingly, the fluorescence lifetime, typically close to 10 ns, and reflects the occurrence of other emission processes since $\tau_F = 1/k_{exp}$.

The energy of the emission occurring via pure fluorescence is typically less compared to that of absorption (Stokes shift). However, the occurrence of these alternative radiative and non-radiative processes could furtherly modified the energy of the emitted photons.

2.4.2 Fluorescence quenching

As explained in the previous paragraph, all the non-radiative processes that depopulate the excited state weaken the emission intensity. The quenching of fluorescence signal requires the close contact of the compound and the fluorophore. In this view, the compound is called "quencher". Quenching experiments are used to explore the accessibility of quencher to a fluorophore, in order to monitor the formation of complexes or the conformational changes. Thus, fluorescence spectroscopy is able to get insights on a small volume element around the fluorophore, probing its micro- and nano-environment.

There are two kind of quenching:

- Dynamic (collisional)
- Static

The dynamic quenching takes place when the fluorophore and the quencher come in close contact during the lifetime of the excited state, i.e. "collide", and the energy of the fluorophore is transferred to the quencher without the emission of a photon.

Differently, in the case of static quenching a complex between the fluorophore and the quencher is formed. A fraction of the fluorophore is bound in a complex and makes no contribution to the emission, whereas the other part exhibits fluorescence. The fluorescence quenching of a protein could be analysed by the classic Stern-Volmer equation, which is employed to elucidate which is the mechanism and the rate of quenching.

$$\frac{F_0}{F} = 1 + K_{sv}Q = 1 + k_q\tau_0 Q \tag{2.27}$$

Where F_0 and F are the fluorescence intensity without and with the quencher, respectively; k_{SV} and Q are the Stern-Volmer constant and the concentration of the quencher, respectively. Quenching data are usually present al plots of $\frac{F_0}{F}$ or $\frac{F_0}{F} - 1$ versus Q and the slope is a direct measure of k_{SV} and of the collision constant k_q. In several system, the Stern-Volmer plot is not linear because it is possible that a combination of both dynamic and static quenching occurs, usually when k_q is higher than $2.0*10^{10}$ $M^{-1}s^{-1}$. In these cases, the data have to be fitted with modified expression of the Stern-Volmer equation, taking into account the effective accessibility of the fluorescent side chains (f_a) and the binding constant K[144]:

$$\frac{F_0}{F_0 - F} = \frac{1}{f_a} + \frac{1}{f_a K}\frac{1}{Q} \tag{2.28}$$

In fact, when f_a is equal to 1, all the fluorescent residues are accessible to the solvent. Additionally, from a double logarithmic equation, it is possible to estimate the number of binding sites n and the association constant K_a:

$$\log\frac{F_0 - F}{F} = \log K_a + n\log Q \tag{2.29}$$

2.4.3 Intrinsic fluorescence

Proteins are unique in displaying intrinsic fluorescence in the UV region via three aromatic side chains such as tryptophan (Trp), tyrosine (Tyr) and phenylalanine (Phe)[137]. Due to the largest extinction coefficient as reported in Fig. 2.2, Trp residues are widely used as intrinsic fluorophores thanks to their strong signal and sensitiveness to their local environment and their solvent-exposure. Differently, Phe

displays a minor contribution in the fluorescence spectra and tyrosine emission is relatively insensitive to solvent polarity.

Trp intrinsic fluorescence is particularly useful when the interaction with other molecules and/or the proteins' self-assembly are studied. However, when a protein accounts for the presence of more than one Trp residues, the emission spectrum becomes complex and it is difficult to separate each contribution. The sensitivity to solvent polarity is conferred by the presence of two nearly isoenergetic excited states of Trp residues, namely ^1L_a and ^1L_b, which present different absorption, emission and sensitivity to solvent polarity due to the diverse indole photo physics. Different is the case of tyrosine residues, the emission of which appears to occur from a single electronic state. These overlapping excited state could give rise to the excitation anisotropy spectra due to the different direction of the electronic transition in the fluorophore as described in Fig. 2.8.

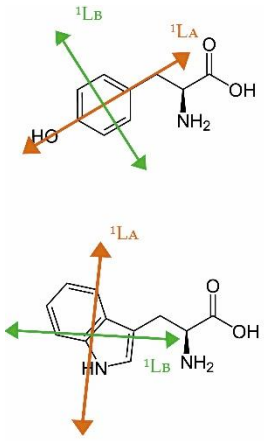


Fig. 2.8: Electronic absorption transition in tyrosine (top) and of tryptophan (bottom). The figure has been readapter from[137] .

In fact, lowest electronic transition of tyrosine (260-290 nm) is due to the ^1L_b transition, while the ^1L_a transition is the origin of the stronger absorption below 250 nm. However, tyrosine displays a constant anisotropy across the absorption band. Differently, the indole of Trp does not and the electronic transition are always perpendicular to each other. This anisotropy presents a minimum at 290 nm, where the absorption spectrum has a maximum due to the ^1L_b state, and increases within

the range 250-280 nm due to the emission of the ^1L_a state. Since ^1L_a state involves the polar nitrogen atom of the indole group, its emission is more sensitive to the solvent effect. For this reason, proteins are generally excited at this wavelength and the emission spectrum depends on the Trp residues microenvironments status. In particular, the emission spectrum shifts to lower energies whether the Trp micro-environment becomes more polar, when Trp is solvent-exposed or when Trp residues are involved in a hydrogen bond network. Differently, a high quantum yield and a higher energy-position of the emission peak is observed when Trp residues are buried in a hydrophobic core. Moreover, in a less polar environment, the ^1L_b state dominates the spectrum, which becomes more structured. Apart from the case where only a Trp is present in the proteins, the emission spectrum of a protein is made up by an unequal Trp residues contribution. Thus, the quantum yields and lifetimes of proteins are due to multiple interactions, which depends on the protein structure.

2.4.4 Advantages and limitations of intrinsic fluorescence spectroscopy

Proteins represent a unique class of biomolecules, displaying a natural intrinsic fluorescence thanks to the presence of three aromatic residues (tyrosine, tryptophan and phenylalanine)[137]. However, these amino acids are rare in proteins, especially tryptophan which actually shows the most intense intrinsic fluorescence. In fact, when at least one Trp residue is present, the fluorescence spectrum of the protein is completely dominated by its contribution. In fact, the indole ring of Trp residues have a high sensitivity lo its local environment, and the modification of Trp fluorescence spectrum accounts for conformational variation, ligand binding, solvent-exposure, denaturation and interactions with other residues. Additionally, fluorescence measurements are performed in aqueous environment and at μM concentrations. Differently from extrinsic fluorescence spectroscopy, the intrinsic one permits to study the interaction between biomolecules and ligands avoiding competitive binding effects due to the presence of another ligand. All these aspects

permit to infer about protein folding and ligand binding. However, this technique presents several limitations. In fact, only Trp residues are sensitive to solvent-exposure and polarity, to ligand binding and denaturation, whereas Tyr is insensitive to all these effects and Phe has the lowest fluorescence signal, being rarely observable when proteins are studied. Additionally, Tyr emission is often quenched by the interaction with the peptide chain or via energy transfer to Trp[137,145]. Additionally, the interpretation of the fluorescence spectra of Trp-rich proteins could be misleading, since each Trp contributes unequally to the total intensity of the peak. Thus, the interpretation of the fluorescence spectra is facilitated if the proteins possess one or a few Trp residues. In order to discriminate the fluorescence signal arisen from each tryptophan, usually proteins undergo a the site-directed mutagenesis process, where only the Trp residue to be study is maintained and the others are replaced with other aromatic residues, usually phenylalanine residues[146]. Similarly, in order to understand the role of a specific residue in a Trp-lacking protein such as a tyrosine, usually the replacement of that residue with Trp is usually made[147]. This *modus operandi* could induce modification of the protein structure under investigation, possible adding biases in the interpretation of the results.

2.5 Atomic Force Microscopy (AFM)

The atomic force microscope belongs to the class of scanning probe microscopes and it consists in the use of a proximal probe for the investigation of the surface properties with a sub-nanometric resolution[148]. The possibilities opened by AFM, especially in the biological field, overcome the limitations of the others scanning probe microscopies since it does not long sample preparation, it is non-destructive, it has a high resolution and it does not require electrical conduction of the sample material. AFM is currently used for the investigation of the structural, chemical and mechanical properties of a sample.

2.5.1 Basic principles

A typical AFM instrument scheme is reported in Fig. 2.9 AFM works by scanning a sample using a very sharp tip mounted on a flexible cantilever[149,150]. The actual shape of the object analysed is reproduced with a sub-nanometric resolution, modulating the force or the sample-tip distance using a piezoelectric actuator. The interaction with the sample induces a cantilever deflection or also a torsion, changing the tip-sample distance and interaction. The sample-tip interaction and the modification of the cantilever could be monitored with a very high resolution by a laser beam, focused on the back of the cantilever. The reflected light is directed towards a photodetector and any modification could be followed by a modification in the position of the reflected beam compared to the reference one. The force required to move the beam through tens of angstrom can be as small as 10^{-18} N. This level of sensitivity is approximately equals to that of interatomic forces. Depending on the modification of the cantilever induced by the interaction with the sample, the distance or the force between the sample and the tip is restored using a feedback circuit. The amount of the feedback signal is an indirect measure of the tip deflection and torsion and concurs to form a 3D image of the sample.

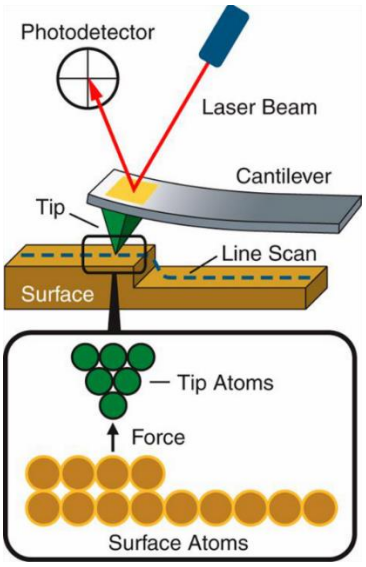


Fig. 2.9: AFM Instrument scheme. The reference laser beam illuminated the back of a vibrating cantilever, which scans the sample. The reflected light is directed to a photodetector and the angles and intensity of the reflected component depends on the morphology of the sample. The tip measures the interaction with the surface and depending of the force measured, a feedback mechanism modulates the response of a piezoelectric material which recover the original force or distance. Figure readapted from[151].

The cantilever and the tips are the most critical components. The cantilever are made with several elastic properties in order to response to different forces and to work in several condition. The operation modes of the AFM is strictly related to the force field between the tip and the sample. Depending on the order of magnitude of the force explored, three working methods can be defined: contact, non-contact and intermittent. If the force between the tip and the sample is always repulsive and the tip is in closely contact with the sample, AFM works in the contact mode. On the contrary, if the tip experiences only an attractive force with the sample, the microscope works in a non-contact mode. It the tip feels both attractive and repulsive forces, the AFM works in the so called intermittent mode. In contact mode, AFM could work by fixing the sample-tip interacting force or by fixing the sample-tip height. In the former case, the tip approach the sample until the positive force reference setpoint is reached (from tens of pN to tens of nN). At this point, the tip start to scan the sample and the feedback circuit restores the force to its initial value as deflections or torsions are detected, in order to maintained a constant value of the force. In this case, the stiffness of the cantilever should be as low as possible. When the sample-tip height is maintained fixed, the deflection signal is acquired as the sample is scanned in the (X,Y) plane and the image resulting from this scanning is a direct force/map of the tip-sample interaction. This mode could be used under aqueous condition due to a reduction of the interaction force between the tip and the sample (i.e. it reduces the capillary effect). However, a relevant problem dealing with the contact mode id the development of dragging forces associated with the lateral movement of the tip over the sample. In the case of biological sample, we could overcome this problem working in the so called intermittent condition. In this case, the cantilever oscillates close to the resonance mechanical frequency of the cantilever until it touches the surface. At this point, the interaction with the sample modifies the frequency of oscillation and, by recording the feedback signal required to keep the amplitude constant, the topography of the sample could be obtained. In this view, the tip only experiences both attractive and repulsive forces. In the intermittent working regimes, the cantilever oscillate at its mechanical frequency ω_0. Considering

the cantilever as a system of spring and mass, we consider $\omega_0 = \sqrt{\frac{k}{m}}$. The amplitude of oscillation could be written as:

$$A = A_0 \sin(\omega_0 t + \varphi) \tag{2.28}$$

The cantilever oscillates obeying to the Hooke's law harmonic oscillation; in fact, the force exerted on the cantilever is:

$$F = -k\Delta z \tag{2.29}$$

Where Δz is the difference between the initial and the final positions along the z axis. Thus, if the tip approaches the sample, the feedback circuit counteract this force modifying the oscillation frequency from ω_0 to $\omega_0{}'$. In the approximation of small oscillation respect the equilibrium position z_0, we could write:

$$F \sim F_0 + \left(\frac{\partial F}{\partial z}\right)\Delta z = -\left(k - \frac{\partial F}{\partial z}\right)\Delta z = -k'\Delta z \tag{2.30}$$

Thus, $\omega_0{}' = \sqrt{\frac{k'}{m}} = \omega_0\left(1 - \frac{\frac{\partial F}{\partial z}}{2k}\right)$ and the measurement of the force modification is an indirect measure of the force gradient.

The force gradient could be measured in two ways: the frequency and the amplitude modulation. In the first case, the ratio $\Delta\omega/\omega = -\left(\frac{\partial F}{\partial z}\Big/2k\right)$ is kept constant, while in the last case, the ratio $\Delta A/A = -Q\left(\frac{\partial F}{\partial z}\Big/2k\right)$, where $Q = m\omega_0/2\alpha$ is the quality factor and α is the attenuation rate. In the regime of constant frequency, the tip experiences only attractive forces and works in non-contact regime (with an average tip-sample distance of 50-150 Å), whereas in the regime of constant amplitude the tip works in tapping regime. In those dynamical modes, the stiffness of the cantilever is generally higher in order to reduce noise and possible instabilities.

Depending on the sample investigated, it is important to decide which working regime has to be adopted. In general, the contact mode could be more destructive, the non-contact has a lower resolution and the tapping has both a high resolution and a less destructive potential on the sample. The force sensitivity (10^{-11} N) is

guaranteed by the spring constant of the cantilevers which range from 0.01 N/m to 100 N/m.

2.5.2 Advantages and limitations of AFM

AFM was created to overcome the problems linked to the use of Scanning Tunnel Microscopy (STM), since STM is able only to explore conductive samples in vacuum. In fact, AFM is a highly versatile technique able to take measurements in air, in liquids or in vacuum, opening new perspectives, especially in life sciences. AFM is able to measure different sample properties such as mechanical, chemical and topographical ones, without be destructive and gaining information on the sample surface. However, due to the very tight and physically close interaction between the tip and the sample, the tip could undergo contamination or breakage, resulting in a sample damaging, in a sample contamination or also in the image damaging.
All these limitations could be avoided if we set the measurement parameters with care; anyway, AFM sensitivity is restricted to surface properties.

Chapter 3

HEWL and insulin as model proteins for studying tryptophan and tyrosine interactions with ligands and solvent by a UVRR-based spectroscopy

Proteins are responsible for nearly every task of cells life. The maintenance of cells shape, organization and waste cleaning as well as the control of external and internal signals and stimuli are received by proteins. They are as structurally diverse as the functions they perform even if the linkage between protein structure and their function is still widely debated. However, as explained in Chapter 1, proteins may lose their biological activity due to different exogenous and endogenous causes and the fluctuations of their structure, represented as diverse conformational states in the energy landscape, correlate with the progressively loss or modulation of their biological activities. When proteins' structural modifications become irreversible, inducing their dysfunctional behaviour, generally proteins follow a favourable energetic path leading to the formation of β-sheets-rich fibrils. In several cases, the process of aggregation could be perturbed by the interaction between proteins and other molecules. To this purpose, the interaction between small molecules and aggregation-prone proteins is widely studied, willing to select or to design small ligand able to inhibit the process or reduce the cellular cytotoxicity of these aggregates. Great interest has been directed to natural compounds such as

polyphenols or flavonoids, which could be taken through a diet. Lots on natural compounds actually show to inhibit the aggregation process, to disassemble pre-formed fibrils and reduced also their cytotoxicity *in vitro*. Even if the mechanism of inhibition is still not fully understood, the aromatic and hydrophobic character of these small ligands seems to play a key role in promoting the physical binding with biomolecules, generally to their hydrophobic portion and involving aromatic amino acids. Apart from the benefits observed *in vitro*, the translation of natural compounds from the "bench to bedside" for human use is limited due to their inefficient delivery, poor bioavailability and high dosage typically required for optimal response. Thus, great efforts are still made to improve the delivery and the therapeutic performances of these compounds, developing nanocarriers (polymers, proteins, nano-sponges and so on) or engineered liposomes in order to deliver them in peculiar organs or tissues without reaching a toxic dosage[152]. Despite from those limitations, the structure of natural compounds, especially of polyphenols, is considered as a model to unravel the mechanism of interaction with dysfunctional proteins.

In this Chapter, we present the study of the fibrillation of two aggregation-prone model globular proteins, namely hen-egg white lysozyme (HEWL) and insulin, and the impact on fibrillation of a model natural antioxidant, which is resveratrol. These systems have been chosen due to their large characterization in literature as explained in the Section 3.1. The study has been conducted by a UVRR-based multi-technique approach. Thanks to the possibility to work in aqueous solution and taking advantage of working in resonance conditions, UVRR opens up the possibility to study complex biological systems as well as their interaction with UV non-resonant ligands, by selectively enhancing only the vibrational signals arising from specific molecular vibrations of biomolecules. The selective enhancement provided by UVRR permits to deeply investigate the behaviour of aromatic amino acids (14), especially tyrosine (Tyr) and tryptophan (Trp) residues, which play an active role in the self-assembly process involved in fibrils' formation. In order to verify the potential of UVRR spectroscopy compared to other biophysical techniques for the investigation of both phenomena, a Trp-rich protein (HEWL) and a Trp-lacking one (insulin) were considered as ideal candidates for the study. UVRR results were

successfully complemented by FTIR and intrinsic fluorescence spectroscopy as well as AFM ones, willing to provide a clear picture of the structural and morphological modifications occurring in proteins upon fibrillation and upon their interaction with a polyphenolic compound.

In this context, UVRR spectroscopy demonstrated to overcome the limitations of intrinsic fluorescence spectroscopy, especially dealing with Trp-lacking proteins; in fact, while UVRR Tyr markers are sensitive to the modification of non-covalent interactions and to the change in microenvironments' hydrophobicity, intrinsic Tyr fluorescence manifest its insensitiveness to both phenomena, hence resulting less applicable in these cases.

Additionally, thanks to the support of FTIR spectroscopy, the secondary structure modifications of HEWL and insulin induced by both phenomena and detected by UVRR spectroscopy are confirmed. Noteworthy, by observing the modifications of the Trp and Tyr vibrational bands in UVRR spectra, we could target the bands mainly affected by fibrillation as well as those affected by the interaction with resveratrol. In particular, we demonstrated that resveratrol mildly modifies protein fibrils secondary structure, while relevantly remodelling their morphology. Additionally, UVRR results, together with computational evidences reported in literature, lead us to speculate about the involvement of Trp (in the case of HEWL) and Tyr (in the case of insulin) residues as resveratrol binding sites.

Overall, our results show that UVRR spectroscopy, complemented by other spectroscopical and morphological techniques, is mandatory to infer about protein fibrillation and their interaction with ligands, and that UVRR may play a pivotal role in this type of studies. Part of the results presented in this chapter has been organized in a scientific paper, which is currently under peer-review[153].

3.1 Model systems for studying protein fibrillation and interaction with ligands

As already discussed in Chapter 1, amyloid fibrils are extended insoluble aggregates formed from misfolded proteins, whose the secondary structure is rich in β-sheets.

Lots of proteins with a well-defined native structure as well as with a natively disordered structure undergo fibrillation *in vivo*, forming the so-called cross-β structure. This evidence indicates that fibrillation is a natural tendency of proteins and the formation of a β-sheets-like structure upon fibrillation is a generally observed phenomenon. The need to study protein fibrils structure and the mechanism leading to their formation promoted the investigation of the conditions inducing proteins fibrillation *in vitro*. The advent of *in vitro* studies has solved many aspects of the previous *in vivo* researches such as the extraction of aggregates from tissues, the possibility to work with higher amount of material and in controlled conditions, reducing also the variability and the polymorphic nature of fibrils structure. Simultaneously, thanks to recombinant technology, large amount of purified proteins can be produced and dedicated for *in vitro* experiments. For how they are designed, *in vitro* studies allow to describe the mechanism of aggregation both thermodynamically and kinetically, simplifying the investigation of *in vivo* processes; on the other hand, they are less affected by the organismal context of the structural change since the conditions inducing protein aggregation are standardized. In confirmation of this, many proteins or protein fragments not forming aggregates *in vivo* produce aggregates *in vitro*[154], demonstrating that proteins work differently in the cellular environment compared to isolated proteins. Since protein aggregation assays are typically affected by poor reproducibility, i*n vitro* studies reduce the number of factors inducing variability.

In order to expand the knowledge regarding the molecular mechanism and the pool of factors inducing protein aggregation as well as to develop therapeutic agents, *in vitro* fibrillation of different proteins have been studied. Taking advantage that, generally, fibrillation leads to the formation of intermediate species first, and of mature fibrils characterized by a cross-β structure at the end, we can simplify the study of fibrillation employing proteins widely studied in literature, small in size and with a well-established 3D structure, i.e. model proteins. To this purpose, the study of model proteins helps in the development of experimental assays which could be extended for the investigation of IDPs.

3.1.1 Hen-egg white lysozyme (HEWL)

First discovered in 1922, lysozyme is a 14.3kDa enzyme comprising 129 residues, which is able to catalyse the hydrolysis of β-1,4 glycosidic bond between N-acetylmuramic acid and N-acetylglucosamine in peptidoglycan, a carbohydrate found in Gram+ bacteria cells wall. For this reason, lysozyme has been widely studied due to its antibacterial properties. Human lysozyme was found in tears, saliva and blood serum, whereas egg white has been identified as a good alternative animal source of the protein. Lysozyme native structure was firstly published in 1965, indicating that it is formed by two domains (α and β) linked by a long α -helix around its hydrophobic core (Fig. 3.1)[155].

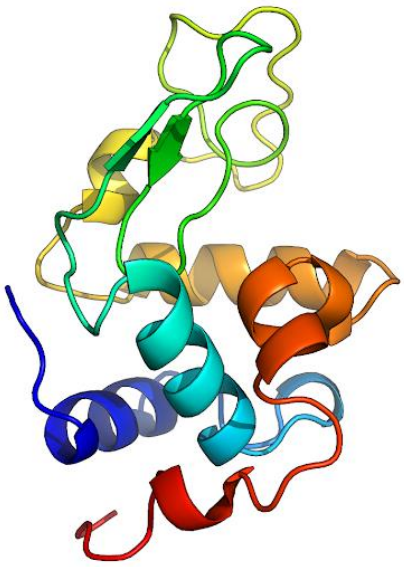


Fig. 3.1: 3D structure of hen-egg white lysozyme (HEWL). Figure readapted from[156].

The structure of lysozyme is stable, kept tightly folded by four disulphide bridges (S-S), exposing polar groups on the surface and the hydrophobic residues buried in the core. Two S-S bonds are responsible for its thermal and chemical stability, since lysozyme retains its activity even after 1-2 minutes at 100°C and also in acidic condition (pH 3-4). A solvent-protected catalytic portion of lysozyme is mainly characterized by two key amino acids such as Glu35 and Asp52[157]. Noteworthy, lysozyme contains 6 tryptophan residues (Trp), 3 tyrosine (Tyr) and 3 phenylalanine residues (Phe).

Over the years, extensive studies on lysozyme folding have shown how different regions of this protein dynamically interact with each another and spatially arrange in order to allowing free access to the substrate, needed for the catalysis. Despite lysozyme has been characterized mostly for its enzymatic activity, numerous experimental studies have been pointed out that lysozyme has been employed as a model protein to study amyloid fibrillation[46,158,159]. In fact, not only lysozyme undergoes *in vitro* fibrillation under several different environmental conditions, but its self-association has been correlated with the development of a hereditary systemic amyloidosis in human[86,160,161], characterized by a huge deposition of lysozyme fibrils in kidneys, gastrointestinal tract, blood vessels, spleen and liver. Several natural mutations in the human lysozyme have been reported (I56T, F57I, W64R, D67H, T70N, and F57I/T70N or W112R/T70N), leading to multiple variants and early onset of the pathology.

Due to the structural similarity (approximately 60%), *in vivo* fibrillation of human lysozyme has been modelled by taking hen egg white lysozyme (HEWL) fibrils[162–166]. Under acidic conditions, lysozyme follows a hierarchical nucleation-like aggregation pathway, as presented in Chapter 1[167].

3.1.2 Recombinant human insulin (rHu)

Recombinant human insulin (rHu) is a hormone (5.8 kDa) involved in important biological tasks such as storage, usage and cellular uptake of glucose, amino acids and fatty acids. At the same time, it also inhibits the degradation of proteins, glycogen and fat. rHu is characterized by a predominant α-helical structure and it is commonly used as a model protein for *in vitro* studies of fibrillation. In fact, its biological dysfunction induces human diabetes and it is prone to form local aggregates after repetitive injections in patients with diabetes.

rHu consists of two polypeptides (A and B chains). A and B chains comprise 21 and 30 amino acids, respectively, which are covalently bonded by disulphide bridges (Fig. 3.2). Three regions of the A-chain, namely residues 1-2, 12-17 and 19, and the 8-25 residues of the B-chain are fundamental for insulin biological function. In fact,

mutations like SerB24, LeuB25, LeuA3, which actually induce the early onset of human diabetes, provoke a weakened *in vitro* insulin activity. Three disulphide bridges keep insulin in their native conformation, bringing the N-terminal tripeptide (Gly-Ile-Val) to the proximity of the invariant C-terminal TyrA19. Additionally, GlyB20 and GlyB23 pushes C-terminal aromatic residues (B24-B26) close to the termini of the A-chain. This architecture constitutes the foundation of the insulin receptor recognition site: the N-terminal of the A-chain (residues A1-A8) is important for receptor activation, while ValB12 and TryB16 associate to the α subunit of the insulin receptor, facilitating the contact between ValB12, PheB24 and PheB25 with the receptor[168]. The aromatic character of PheB24 and PheB25 has been retained of particular importance for insulin biological activity, since non-aromatic substitutions of both residues provokes a loss of the biological activity, whereas the substitution with tyrosine residues maintain full activity. Differently from HEWL, insulin does not present tryptophan residues, but it is characterized by the presence of 4 tyrosine (TyrA14, TyrA19, TyrB16, TyrB26) and 3 phenylalanine (PheB1, PheB24, PheB25), making the protein an ideal Trp-lacking system to be studied.

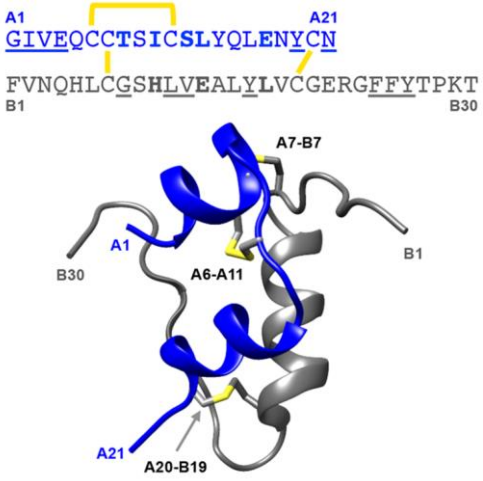


Fig. 3.2: Sequence and 3D structure of human insulin. Yellow lines represents the disulphide bridges, blue sequence the A-chain and grey sequence the B-chain. The figure has been readapted form[169].

As observed for HEWL, insulin is quite stable but under harsh conditions of temperature and pH rapidly converts into amyloid fibrils[170,171]. The kinetics of the fibrillation process could be described as a nucleation-elongation pathway, where

the formation of stable nuclei is observed during the lag time, followed by an exponential growth of the fibrils. However, insulin presents a wide polymorphism capability and it is present in multiple forms such as spherulites, fibrils and other intermediate forms.

3.1.3 An example of a model ligand: resveratrol

Due to their dimension, small molecules have reached great interest in the pharmaceutical field since they are able to cross the blood-brain-barrier (BBB)[172–174]. Nonetheless, if not properly characterized and tested, these molecules could become dangerous, becoming endocrine-disrupting chemicals, which may bind to proteins promoting their dysfunction[175]. However, the study of the mechanisms that rule out protein-ligand binding can be useful in therapeutics and in diagnostics fields. In fact, small ligands may have the potential to influence protein aggregation and/or reducing deposits cytotoxicity[93]. In particular, the interaction between natural compounds and protein during fibrillation or at the end of the process has been widely studied *in vitro*, showing that these small ligands could interact with aromatic or hydrophobic pockets limiting o slowing down the process of fibrillation[93,176–178]. Interestingly, fibrillation inhibitors are usually small molecules rich in aromatic moieties, indicating that hydrophobic forces play a pivotal role in aggregation and, consequently, in fibrillation inhibition.

Among the stilbene non-flavonoid family, resveratrol is a natural compound of great importance which naturally can be found in several foods and in red wine. It is characterized by two aromatic rings connected by a methylene bridge. The chemical structure of resveratrol is characterized by two isoforms, which are *cis-* and *trans-* (Fig. 3.3). It is one of the smallest polyphenols, usually used as model molecules to study protein-ligand interaction.

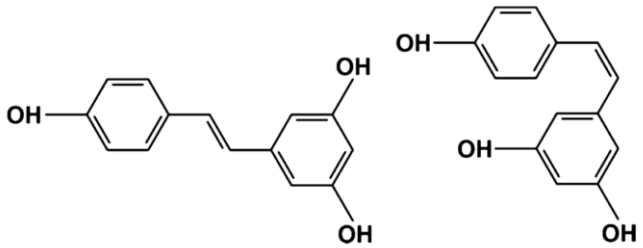


Fig. 3.3: (left) *trans*- and (right) *cis*- isoforms of resveratrol. Figure readapted from[179].

Resveratrol acts as an antioxidant, but also modulate the toxicity of Aβ proteins in cell culture, promoting cellular mechanism of degradation of amyloids[180,181]. Additionally, resveratrol may also directly interact with amyloid fibrils, reducing their toxicity as in the case of hIAPP or depolymerizing them[182].

Unfortunately, similarly to other compounds, resveratrol solubility in aqueous media is poor and scientists have faced out with the problems linked to its delivery *in vivo* and to its bioavailability[64,99]. On the other hand, the poor solubility in aqueous environment lets resveratrol more prone to bind to protein hydrophobic side chains, which usually tend to stay exposed in amyloid toxic species[64]. In fact, clinical trials have been shown that resveratrol is safe to use in patients with mild to moderate Alzheimer's disease (AD)[174]. Patients treated with resveratrol for one year showed a lower decline of Aβ fibrils levels in cerebrospinal fluid (CSF)[173,183]. Additionally, despite its lower bioavailability, significant amount of resveratrol has been found in CSF, indicating its ability in crossing the blood-brain barrier[173,183]. Treatment of patients with AD and stroke was beneficial suggesting that resveratrol can be effective dealing with neurological disorders[172,174,183]. Moreover, in diabetes patients, resveratrol seems to increase insulin sensitivity and decrease blood glucose levels[184]. Nowadays, the major efforts are directed to improve the bioavailability of resveratrol *in vivo* in order to reduce the dosage[185]. In fact, depending on its concentration it can has opposite biological effects[186]. Martins and colleagues found out that resveratrol could have a pro-oxidative effect in a dose-dependent manner, leading to cells oxidative stress: while low resveratrol doses (0.1-1 μg/ml) could improve cells proliferation, higher doses (more than 10 μg/ml) induces cellular apoptosis[187].

3.2 Experimental methods

3.2.1 Preparation of protein samples

Hen-egg white lysozyme (HEWL, L6876), human recombinant insulin (91077C) and resveratrol (R5010) were purchased from Sigma-Aldrich and used without further purification. Resveratrol was dissolved in 100% ethanol and further diluted in phosphate buffer saline (PBS, pH 1.8). HEWL and insulin solutions were prepared in acidic condition (PBS at pH 1.8, 1.02 mM adjusted by a solution of HCl 37%). The samples were then diluted to the final concentration of 1mM using the solution of resveratrol, reaching a (protein:resveratrol) molar ratios of (1:0), (1:1) and (1:2). HEWL/insulin (1% EtOH) were defined as "HEWL" and "insulin", HEWL/insulin + resveratrol (1:1) as "HR1"and "IR1", and HEWL/insulin + resveratrol (1:2) as "HR2" and "IR2", respectively. Part of these solutions were measured as freshly prepared solutions, while the remaining parts were kept at 65°C, shacked at 750 rpm in a dark place to induce fibrillation. The fibrillation was stopped when the gelatinous phase was achieved, i.e. after 7 days for HEWL and after 1 day for insulin. Then, the solutions were centrifuged at 16000 g for 40 min at room temperature. The supernatant solutions were used to estimate the concentration of the gelatinous fraction. Finally, the pellet was re-suspended in PBS prior to measurement, achieving a final ethanol concentration of 1% EtOH.

3.2.2 UV Resonance Raman spectroscopy

UV Resonance Raman measurements were carried out at the BL 10.2 – IUVS beamline at Elettra Sincrotrone Trieste using the experimental setup reported elsewhere[134]. Samples spectra were measured using two different wavelengths, namely 228 and 244 nm. Synchrotron radiation was used to get the 228 nm source, whereas the measurements taken at 244 nm were carried out using an Ar+ laser (Lexel Laser Inc., USA). Proteins spectra were acquired at room temperature. The diffused radiation was collected in backscattering geometry and energetically

dispersed using a spectrometer with 1800 g/mm (Trivista, Princeton Instrument). The final spectral bandwidth is 20 cm^{-1} at 228 nm and is 8 cm^{-1} at 244 nm. Samples excited at 228 nm were irradiated for 3 hours, accumulating one spectrum every 15 minutes with an incoming radiation of 16 μW. Differently, with 244 nm of excitation wavelength, samples were irradiated for 10 seconds at 140 μW. The final spectrum was the average of 40 spectra. Spectra of buffer and of resveratrol solutions were taken as well and their contributions were carefully removed from the measured spectra of proteins. Samples were horizontally oscillated during the measurement to avoid photodamage.

The spectral region of interest was fitted by using Gaussian curves, which in our case represent a good choice to avoid high correlation between the fit parameters. Considering the same number of curves in order to reproduce the experimental data, the use of Gaussian curves instead of Voigt ones does not induce significant variations of χ^2 but reduces the number of parameters for each peak (3 for Gaussian and 4 for Voigt curves). The high correlation of those peaks is highlighted by the analysis of the errors deriving from fitting the experimental data by using only Voigt curves. The use of Gaussian curves re-duces the error of the parameters from 50% (in the case of Voigt) to 10%, which is close to the S/N ratio of the spectra collected, without altering the final result in term of peaks relative intensity. The Gaussian curves were initialized to the value corresponding to the maximum of the peaks and kept fixed for a first run. Then, the parameters were made free and the fit runs until its convergence.

3.2.3 Fourier Transform Infrared spectroscopy

Infrared spectra of the samples were collected at the infrared beamline SISSI (Synchrotron Infrared Source for Spectroscopic and Imaging), Elettra Sincrotrone Trieste (Trieste, Italy)[140]. Attenuated-total-reflectance (ATR) spectra of the samples were collected using MIRacle™ Single Reflection ATR box (PIKE Technologies) with Germanium (Ge) internal reflective element (IRE). Spectra were acquired using the Vertex 70 interferometer (Bruker Optik GmbH) equipped with a MIR DLaTGS

detector and a KBr beam-splitter. 2μL drop of each sample was deposited onto the IRE and the measurements were performed during drop dehydration at ambient temperature, under a gentle nitrogen flux to prevent water vapor spectral interference. To the same purpose, interferometric compartment and ATR accessory were also purged with nitrogen. For each spectrum, 15x128 scans-interferogram were collected with a scanner velocity of 7 kHz and 4 cm^{-1} of resolution. The spectral region explored spanned a range within 5000 to 500 cm^{-1}. A 256-scan interferogram of the background was collected under identical external condition.

3.2.4 Atomic Force Microscopy

A 15 μL-drop of the diluted solution containing fibrils was spotted on freshly cleaved mica and let absorb for 15 minutes (dilutions were optimized to obtain a sufficient number of fibrils on the surface). Then, the deposited samples were washed with 1 ml of deionized water and we dried with a gentle stream of nitrogen. To avoid sticking effects due to the residual water on the surface, we kept the samples overnight in a N2 chamber. We performed topographic measurements in dynamic mode in air with commercial cantilevers (NSC36b, Mikromasch, nominal spring constant=2 N/m, radius of curvature < 10nm; NSG03 NT-MDT, nominal spring constant=1.7 N/m, radius of curvature < 10nm) on MFP-3D (Oxford Instruments/Asylum Research) and Solver Pro (NT-MDT) instrument at the NanoInnovation Laboratory at Elettra. Images were acquired with a pixel resolution of roughly 5 nm, acquiring 5x5 μm^2 images at 1024x1024 $pixels^2$ and then analysed with Gwyddion software to remove background with standard flattening processes[188]. To obtain length distribution of fibrils we performed a statistical analysis using the Persistence software, developed by De La Cruz lab[189].

3.2.5 Intrinsic fluorescence spectroscopy

Fluorescence spectra were collected at 24°C using a Perkin Elmer LS-55 spectrofluorometer. Solutions of HEWL (5μM) were excited at 280 nm, while insulin ones (20μM) at 276 nm, recording the emission in the range of 200–500 nm. 500 μL of each solution were loaded inside a quartz cell and measured at different times to follow the aggregation kinetics. For quenching experiments, small aliquots of resveratrol solution from a stock solution of 1mM were added to 500 μL solution of HEWL and insulin, reaching a maximum concentration of 27,45 μM and 23μM of resveratrol, respectively.

3.3 Results

3.3.1 Modifications of HEWL and insulin secondary structure changes upon fibrillation

As reported in Chapter 2, proteins secondary structure is conventionally investigated by recording proteins FTIR spectra, focusing on amide I (1600-1700 cm^{-1}) and amide II bands (1490-1580 cm^{-1}) (see Fig. 3.4). This spectral range is dominated by specific normal modes of vibrations related to peptide linkage[117,130,190]. In particular, due to the high sensitivity of C=O stretching to its own H-bond state, amide I is commonly used to discern about protein secondary structure composition[117,138]. Amide bands of freshly prepared and fibrils solution of HEWL and insulin are shown in Fig. 3.4A(a) and Fig. 3.4B(a), respectively. In order to resolve the overlapping amide I band components arising from different secondary structural elements, the second-derivative analysis of the FTIR absorbance spectra was performed (see Fig. 3.4A(b) and 3.4B(b), respectively)[130,190]. The positions of the minima found in the second derivative spectra are reported in Table 3.1, accompanied by a complete assignment. Additionally, in order to follow the spectral modification of protein secondary structure content, FTIR spectra were fitted employing Gaussian curves, the centres of which were initialized using the minima found in the second derivative spectra (indicated by the dashed lines in Fig. 3.4(b)

panel A and B). The fitting results, reported in Table 3.1, highlight that the secondary structure of both freshly prepared HEWL and insulin is mostly characterized by a predominant content of α-helix (58% and 40% of the amide I band area, respectively), in agreement with literature[192,193]. The remaining band areas (42% for HEWL and 60% for insulin) account for other vibrational contributions arising from lateral chains, unordered structures and β-sheets. As reported in Table 3.1, the 1616 cm^{-1} has been assigned to a mixed contribution arising from lateral chains, in particular to Tyr-OH vibrational modes, and intermolecular β-sheets. In particular, depending on the protein investigated, a plethora of papers assigned the sub-band below 1620 cm^{-1} to intermolecular β-sheets, especially when amyloid fibrils are detected[6,102,193–203], while many others addressed this band to lateral chains vibrations[117,138,193]. Additionally, the presence of β-sheets signals below 1620 cm^{-1} is not restricted to the formation of amyloid fibrils, but it has been observed during the formation of small oligomeric species or the presence of small aggregates, especially when proteins are kept under harsh conditions[50,195,204]. In our case, multiple factors such as high protein concentration and acidic pH may induce fast modification on protein native structure[50,204]. Consequently, we could not exclude that our experimental conditions may have induced the formation of a small percentage of additional intermolecular β-sheets, which may contribute to the 1616 cm^{-1} band intensity, conventionally solely addressed to Tyr-OH vibrations (i.e a mixture of the in-plane C-C stretching ν_{CC} and to the deformation of the ring (δ_{HCC}).

As shown in Fig. 3.4 and Table 3.1, fibrillation of HEWL and insulin is accompanied by a significant change in the FTIR spectra line-shapes, compared to the freshly prepared protein solutions. By following the modification of amide I shape, we observed that both fibrils present two intense β-sheets bands located approximately

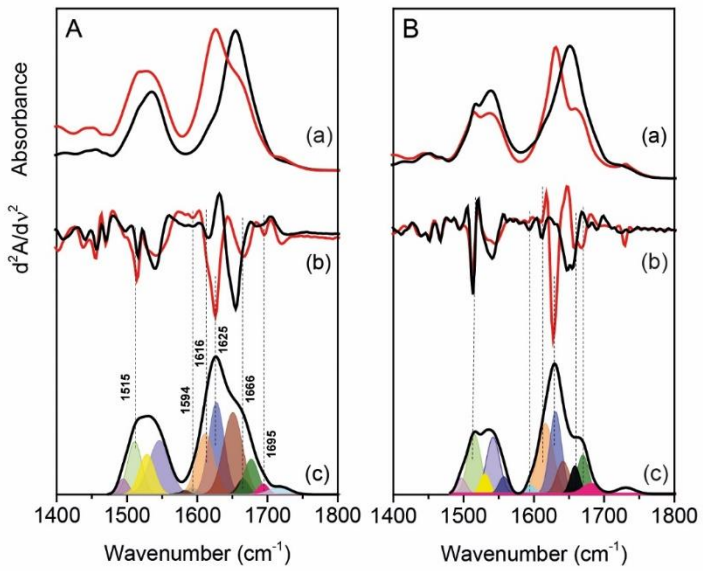


Fig. 3.4: (A) FT-IR spectra of (a) HEWL as freshly prepared and fibrils solution (in black and in red, respectively), (b) of the second-derivative spectra of HEWL freshly prepared and fibrils solutions (in black and in red, respectively) and (c) the fit of the HEWL fibrils spectrum with amide I and II sub-bands highlighted with colours. Dashed lines are used to distinguish different secondary structure components. (B) FT-IR spectra of (a) insulin as freshly prepared and fibrils solution (in black and in red, respectively), (b) of the second-derivative spectra of insulin freshly prepared and fibrils solutions (in black and in red, respectively) and (c) the fit of the insulin fibrils spectrum with amide I and II sub-bands highlighted with colours. Dashed lines are used to discern about the secondary structure components.

in the regions (1626-1631) cm^{-1} and a little one at 1695 cm^{-1}, which are typical of intermolecular β-sheets. The position of the Am-I maximum when mature fibrils are formed is informative on the β-sheets organization, namely if they have a parallel or an anti-parallel orientation. Generally, when Am-I maximum peak is located below 1630 cm^{-1} and a minor contribution is detected in the region 1686-1695 cm^{-1}, the intermolecular β-sheets are anti-parallel, while when Am-I maximum is located at 1630 cm^{-1} or higher wavelengths, and the high energy band is no longer detected, β-sheets have a parallel orientation[204–206]. Anyway, small deviations to this rule are commonly observed: in fact, as reported for transthyretin and other systems, the lower energy position of β-sheets band can shift depending on their length, namely on the number of strands composing the sheets[201]. Due to this evidence and to the fact that proteins could be composed also by a mixed composition of parallel and

anti-parallel β-sheets[207], it is preferable to refer to those bands as generic intermolecular β-sheets bands. Moreover, in both cases, fibrillation induces also the

Table 3.1: Assignment, spectral positions and content of amide I secondary structures probed by FTIR spectroscopy. The assignment of amide I secondary structures is made on the basis of the minima found in the second-derivative spectra of the freshly prepared and of the fibrils solutions of HEWL and insulin (1%EtOH). The spectral position and the content of each protein secondary structure is reported as well. The error estimated for each area is approximately of 2%.

HEWL (1% EtOH)				Insulin (1% EtOH)				Assignment[117,130,138,171,190,193,204]
Freshly-prepared		Fibrils		Freshly-prepared		Fibrils		
(±2cm^{-1})	(%)	(±2cm^{-1})	(%)	(±2cm^{-1})	(%)	(±2cm^{-1})	(%)	
1616	10.5	1616	20	1615	10	1616	31	Lateral chains/β-sheets
		1626	31	1631	8	1630	28	Intermolecular β-sheets
1641	12.5			1648	32	1645	16	Unordered structure
1654	58	1651	3	1656	40	1654	8	α-helix
		1666	30			1668	12	Irregular helix/loops
		1677	10	1678	3			turns
1683	19	1695	2	1691	6	1695	5	β-sheets

rise of loops/irregular helical structures located at 1666-1668 cm^{-1} and the increasing of the 1616 cm^{-1} (see Table 3.1) and of 1515 cm^{-1} band intensities, together with a frequency redshift of the latter band. Both the 1616 cm^{-1} and the 1515 cm^{-1} bands arise from a mixed contribution of the in-plane C-C stretching and from the deformation of Tyr ring (δ_{HCC}). While the 1616 cm^{-1} band has not been extensively studied since its contribution rarely appears as a well-resolved peak, the intensity and the band frequency modifications of the 1515 cm^{-1} are easier to monitor as they are generally more distinguishable from the contribution of amide II[208]. This band is considered as a local reporter of the tertiary structure modification and on the H-bond network where Tyr residues are on average involved[209,210]. In particular, as reported also for other systems[208,211,212], the 1515 cm^{-1} band increases its intensity (see Fig. 3.4(a) panel A and B) and redshifts of ~2 cm^{-1}, upon fibrillation. A similar redshift could be thought as meaningless, but it is actually larger than other already

reported in literature[212,213]. It accounts for a change in the hydrogen bond network where Tyr residues are involved, and it has been reported many times working with aggregated proteins[208,214], indicating the formation of a stronger H-bond network where tyrosine OH group are involved as H-donor [208,214,215].

The secondary structure arrangement predicted by FTIR spectroscopy is confirmed by UVRR spectra obtained at 228 nm and 244 nm. However, due to the higher spectral resolution achieved at 244 nm, we were able to identify precisely the bands related to protein secondary structure (i.e. amide I and III). Interestingly, Raman spectroscopy is widely used to investigate proteins and peptides secondary structure and UVRR has been extensively employed for elucidating protein fibril core structural organization and to monitor all the steps of fibrillation[118,136,216–218]. Protein secondary structure is usually derived from the spectral position of Am-I and Am-III, since the intensity of Am-II is hampered only with H/D exchange[217]. The correlation between the position of amide bands and the composition of protein secondary structure was made using chemometrics. Hence, UVRR spectra of purely α-helix, β-sheets and unordered secondary structures-like proteins, with a well-known X-ray structures, were collected using an incoming radiation energy of 193-206 nm, and their amide I, II, III and S positions were used to directly unravel the unknown secondary structure of other proteins[219]. Since the radiation energies used in this work do not satisfy amide bands resonance conditions, those bands were decomposed by a linear fitting using a routine based on the Levenberg-Marquardt non-linear least squares method as proposed in[129]. In particular, in Fig. 3.5A (b) and 3.5B (b), the 244nm-UVRR spectra of freshly prepared and fibrils solutions of HEWL and insulin are presented and the vibrational peaks are assigned using the same nomenclature reported in[220]. In both cases, the modification of the secondary structure could be easily observed by the shape variation and by the energy shift of amide I and III bands (Am-I and Am-III in Fig. 3.5A(b) and 3.5B(b), indicated by green boxes). Similar to FTIR data analysis, a fitting procedure has been applied to the 244 nm- HEWL and insulin freshly prepared and fibrils spectra (see the inset of Fig. 3.5B) in order to follow the modification of those bands upon fibrillation (see results in Table 3.2). From Table 3.2 we can deduce that both freshly prepared HEWL

and insulin solutions have Am-I and Am-III typical of mixed α-helix/unordered populated protein[129,221,222], whereas fibrils solutions are characterized by a redshifting of Am-III and an Am-I located approximately at ~1670 cm^{-1}, which is typically found in a β-sheets-rich protein and in amyloid fibrils [136,216,220,221,223]. In particular, Am-III band position located at 1242 cm^{-1} corresponds to a Ramachandran angle of 150°, associated to an anti-parallel β-sheets structure. This observation furtherly justifying our choice to not define the orientation of β-sheets, since this evidence is in contrast with our FTIR results[224].

Table 3.2: Assignment of protein secondary structure made by UVRR spectra analysis, through amide I and III spectral positions. Am-I and Am-III positions evaluated by the fit of the 244 nm-UVRR spectra of freshly prepared and fibrils solutions of HEWL and insulin. A complete attribution of the secondary structure has been made according to[221].

	HEWL		Insulin	
	Am-I (cm^{-1})	Am-III (cm^{-1})	Am-I (cm^{-1})	Am-III (cm^{-1})
Freshly prepared	(1656, 1674) ± 2 (α-helix, β-sheets)	1246 ± 1	(1656, 1683) ± 2 (α-helix, unord.)	1257 ± 2
Fibrils	1672 ± 3 β-sheets	1229 ± 3	1671 ± 1 β-sheets	1242 ± 2

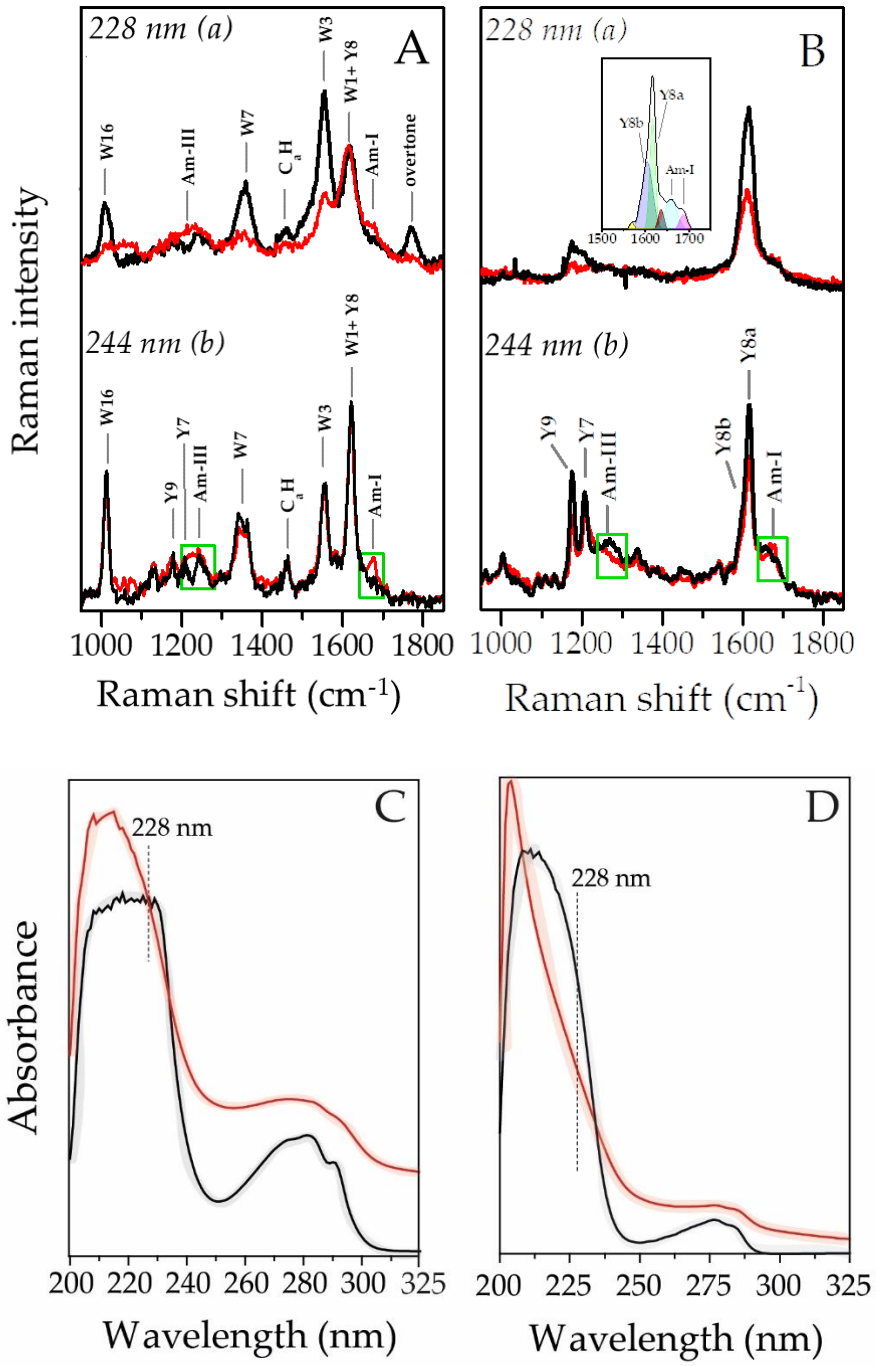


Fig. 3.5: UVRR spectra of freshly prepared (in black) and fibrils (in red) (A) HEWL and (B) insulin solutions taken at (a) 228 nm and (b) 244 nm-excitation wavelength. Peaks assignment has been made according to the nomenclature reported in[220]: Trp, Tyr and amide bands are indicated with symbols W*, Y* and Am-*, respectively. The inset in panel B shows an example of the fit of freshly prepared solution of insulin taken at 244 nm. UV absorption spectra of freshly prepared and fibrils solutions (depicted in black and in red, respectively) of (C) HEWL and (D) insulin. in the spectral region (200-328) nm. Shaded contours represent the variability of the absorption spectra in 3 different experiments.

3.3.2 Tertiary structure changes upon HEWL and insulin fibrillation

UVRR spectra of freshly prepared and fibrils solutions of HEWL and insulin could be used to evaluate proteins' tertiary structure modification upon fibrillation (see Fig. 3.5A and 3.5B). Differently to the 244 nm-spectra, the 228 nm-ones present a furtherly enhanced signal of Trp (indicated with W* in Fig. 3.5A) and Tyr bands (i.e. Y* in Fig. 3.5A, 3.5B), letting this wavelength particularly suitable for the study of protein tertiary structure. In fact, the spectrum in Fig. 3.5A(a) is characterized by an overtone band at approximately 1760 cm^{-1} arisen from the sum of two Trp bands located at 1009 cm^{-1} and 757 cm^{-1}. The presence of the overtone is rarely observed working in off-resonance conditions, explaining why it is absent in the 244 nm-spectra (Fig. 3.5A(b))[128]. Thus, the 228 nm-spectra of HEWL are dominated by the vibrational features addressed to Trp bands[135,220,223,225], whereas the insulin ones by Tyr bands since no tryptophan residues are present in the aminoacidic sequence. Both the 228 nm-UVRR spectra of insulin and HEWL solutions are characterized by the damping of the Trp and Tyr residues peaks' intensities when both mature fibrils are formed (see Fig. 3.5A (a) and 3.5B (a), respectively). These evidences confirms previously obtained UVRR results concerning HEWL and insulin fibrillation[226–228], where Phe UVRR band cross-section (at 1003 cm^{-1}) decreases its intensity upon HEWL fibrillation[136,229], and similarly, both Phe and Tyr bands decrease their Raman cross-sections during insulin aggregation[226–228]. Despite the intensity damping of aromatic residues bands has been correlated to the establishment of a strong H-bonding network where aromatic residues are involved[230–234], a quantitative evaluation of the H-bond sensitivity of these bands is complicated by environmental influences and only the energy shift of specific UVRR bands has been straightforwardly associated to the formation of H-bond[235]. Indeed, modified solvent exposure, hydrophobicity and solvent polarity are considered the main causes of aromatic amino acids UVRR band cross section modification[135,136,219,229,235,236]. In fact, Xu and colleagues demonstrated that Phe peak at 1003 cm^{-1} is particularly sensitive to water exposure and solvent polarity[136], and Asher et al. explained the correlation

between the intensity increasing of both Trp and Tyr UVRR bands and the increased solvation of their microenvironments, as the result of a blue shift of their absorbance spectra upon hydration[236]. In our case, the UV-Vis spectra of both HEWL and insulin fibrils are characterized by a blue-shift of the main absorption peak, while maintaining a comparable or even a reduced absorption degree (see Fig. 3.5C and 3.5D). This imply that self-absorption plays a minor role in the modification of aromatic amino acids band, which actually due to water exposure and hydrophobicity variation of Trp and Tyr residues. Thus, the intensity damping of these UVRR bands upon fibrillation suggest that part of the Trp and Tyr residues in HEWL and insulin surely reside in a less hydrophobic and water-exposed environment when mature fibrils are formed and, eventually, could hydrogen bond to solvent molecules.

In the case of HEWL fibrils, UVRR results could be complemented by the data collected with Trp intrinsic fluorescence spectroscopy. Previously reported fluorescence experiments on HEWL indicates that among the 6 Trp residues of the protein, only Trp62 and Trp108 are the dominant emitters and contributes to the 80% of the total fluorescence signal of HEWL. Trp63 and Trp123 are non-fluorescent, while Trp28 and Trp111 comprise for the 20% of the total fluorescence signal of HEWL. In particular, the fluorescence peak addressed to Trp28 and Trp111 is located approximately at 320-330 nm, indicating that those residues are embedded in a non-polar hydrophobic matrix. Differently, Trp62 and Trp108 emit at 352 nm and 342 nm, indicating that the former is solvent exposed and is in a polar environment, while Trp108 is not solvent-exposed but in a polar environment. In fact, Trp108 establishes an interaction with Glu35, which seems to be impaired when the NAG substrate binds to the active site provoking a quenching of the fluorescence signal[237,238].

During fibrillation, we observed a quenching and a redshift of the Trp maximum peak from 342 nm to 345 nm (see Fig. 3.6 (left)), indicating a major solvent exposure of Trp residues or the establishment of intra- or inter-molecular interactions between Trp62, Trp108 and other polar side chains [137,239]. Nevertheless, when HEWL mature fibrils are formed in similar fibrillation conditions, the sequence 57-107 was found inaccessible to pepsin digestion[240]. In light of this evidence, we can conclude that

Trp62 remains buried in the hydrophobic core of the fibrils, being somehow involved in a non-covalent interaction with other residues, while Trp108 becomes more solvent-exposed. In light of these consideration, the damping of Trp UVRR intensity could be interpreted as a gradual solvent exposure of those residues when mature fibrils are formed.

Unfortunately, in the case of Trp-lacking proteins, intrinsic fluorescence spectroscopy cannot infer about protein residues solvent-exposure[147,241]. In fact, as reported in Fig. 3.6(right), the insulin Tyr fluorescence peak located at 305 nm does not shift in energy upon fibrillation, but it only undergoes quenching. However, the intensity damping of the peak does not correlate straightforwardly to a major Tyr solvent-exposure, as already evidenced by Bekard[241] on bovine insulin and by Dusa[147] on α-synuclein.

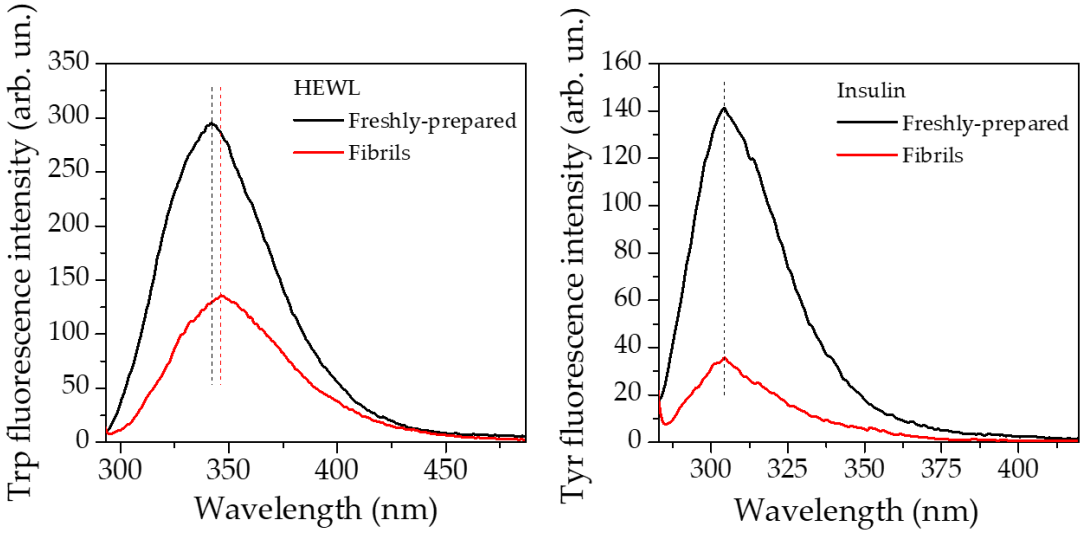


Fig. 3.6: (left) Trp fluorescence intensity of HEWL freshly prepared and fibrils solutions (in black and in red, respectively). Trp peak maximum red-shifts from 342 nm towards 345 nm during fibrillation. (right) Tyr fluorescence intensity of insulin freshly prepared and fibrils solutions (in black and in red, respectively). Tyr maximum peak does not shift in energy upon fibrillation.

In light of this limitation and despite its applicability in the case of Trp-rich proteins, intrinsic fluorescence spectroscopy has to be complemented by UVRR spectroscopy in order to get important insights regarding Trp-lacking protein tertiary structure and aromatic residues microenvironments' hydrophobicity.

Noteworthy, Fig. 3.5B(a) shows the quenching of UVRR Tyr peaks intensities (especially the Y8 band, which is actually a doublet Y8a and Y8b), suggesting an average gradual solvent exposure of Tyr residues upon fibrillation. Additionally, by considering the 244 nm-UVRR spectra (Fig. 3.5B (b)), other markers sensitive to fibrillation have been analysed[242]. To this purpose, the Y8b band position has been used as an indicator of the strength of H-bond where Tyr is involved as H-donor, and the modification in the relative intensity of Tyr Fermi doublet $R = I_{853}/I_{828}$ has been used as a marker of Tyr average micro-environment hydrophobicity variation[242–245]. Table 3.8 summarized the result of the 244nm-UVRR spectra analysis of insulin.

Table 3.3: Y8b band position and R ratio extracted by band fitting procedure applied on insulin freshly-prepared and fibrils spectra.

	Y8b band position (cm^{-1})	R ratio (I_{853}/I_{828})
Freshly prepared insulin	1606 ± 3	2.2 ± 0.3
Insulin fibrils	1597 ± 2	1.6 ± 0.2

In particular, the red-shifted position of Y8b band after fibrillation suggests that some Tyr residues act as H-donor when mature fibrils are formed[243]. Additionally, from the band fitting of the Tyr Fermi doublet located at 828-853 cm^{-1}, we estimated that the R ratio decreases in the case of fibrils. Phenomenologically, the decreasing of the 853 cm^{-1} component is associated to a reduced hydrophobicity of Tyr environment or it is observed when the hydroxyl group acts as a strong H-bond donor [242].

In light of the previous considerations, we deduce that the overall intensity decreasing of Tyr bands, the red-shifted position of Y8b band and the value of R ratio are all clear hints of Tyr increased solvent-exposure and of the phenolic rings involvement in a H-bond network when mature fibrils are formed[232,233]. These findings are in agreement with literature, since among the 3 solvent-protected Tyr residues (TyrA14, TyrB16 and TyrB26) in native insulin[241], only TyrA19 was found

completely solvent-protected when insulin fibrils were formed, while TyrB16 was found partially protected and the other two (TyrA14 and TyrB26) solvent exposed[246]. The major solvent-exposure of Tyr residues when mature fibrils are formed predicted by the UVRR markers is confirmed by results already reported in literature [242,243,245–249], confirming the ability of UVRR to infer about solvent-exposure/change in hydrophilicity of aromatic residues microenvironment.

3.3.3 Quenching experiments: interaction between HEWL, insulin and resveratrol

Quenching experiments were performed on both HEWL and insulin with the addition of resveratrol as a quencher. It could be easily seen in Fig. 3.7 (a) and 3.7 (e) that the addition of varying concentration of resveratrol induces a noticeable quenching of the fluorescence intensity of both HEWL and insulin, respectively, which is related to the saturation of both proteins binding sites. Generally, intensity quenching takes place when the quencher comes closer to Trp or Tyr residues[250].

In the case of HEWL (see Fig. 3.7 (a)), the intensity quenching of the Trp emission peak is accompanied also by a redshift from 342 nm to 351 nm when the concentration ratio (HEWL:resveratrol) passes from (1:0) to (1:6), indicating that the interaction with resveratrol induces conformational changes leading to a further solvent-exposure of Trp side chains or to an increased polarity of fluorophores' microenvironments.

Both Trp and Tyr emission fluorescence peak' shapes are influenced by the presence of a continuously higher amount of *trans-* resveratrol, which increases the quantum yield of its fluorescence maximum peak is located at 400 nm once continuously added. Noteworthy, an isosbestic point is visible at 389 nm in the case of HEWL and at 350.5 nm in the case of insulin, which is indicative of equilibrium between bound and free forms of the quencher[251]. The formation of the isosbestic point occurs when (HEWL:resveratrol) concentration ratio is approximately over (1:1) and (insulin:resveratrol) is over (8:1).

The fluorescence quenching of HEWL and insulin were analysed by the following modified Stern-Volmer equation, which was employed to elucidate which is the mechanism and the rate of quenching taking into account that the quenching is not only dynamic[252]:

$$\frac{F_0}{F} = (1 + K_{sv}Q)e^{VQ}$$

where Q is the molar concentration of the quencher and V is the static quenching constant.

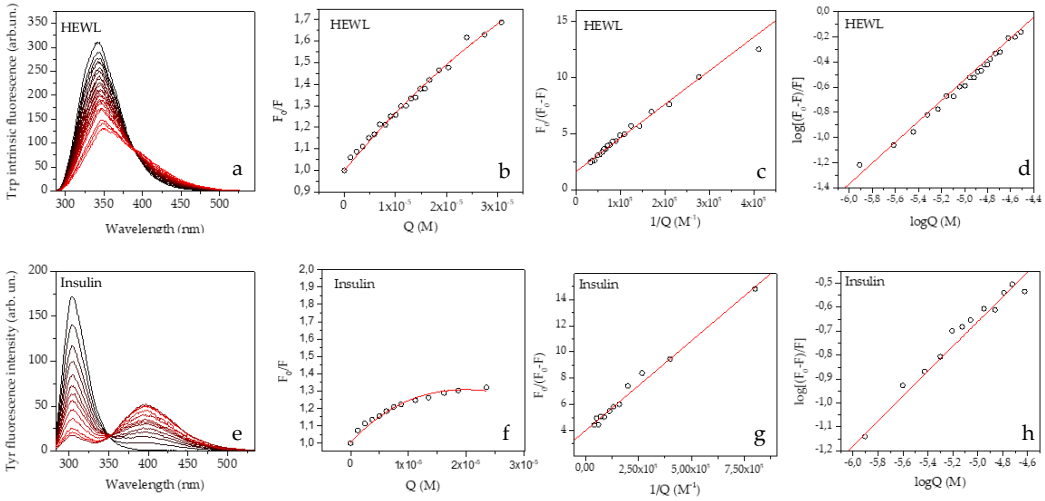


Fig. 3.7: (a) quenching experiment on HEWL+resveratrol at λ_{ex} = 280nm; (b) Stern–Volmer plot and (c) modified Stern Volmer plot for fluorescence quenching of HEWL; (d) Double logarithmic curve of HEWL-resveratrol solution. (e) quenching experiment on insulin +resveratrol at λ_{ex} = 276nm; (f) Stern–Volmer plot and (g) modified Stern-Volmer plot for fluorescence quenching of insulin; (h) double logarithmic curve of insulin-resveratrol solution.

The parameters extracted from the fit of both protein quenching data reported in Fig. 3.7 (b) and 3.7 (f) are summarized in Table 3.4, assuming that τ_0 is nearly 10^{-8} s for a typical biopolymer. The high k_q values (k_q>10^{10} $M^{-1}s^{-1}$), estimated in both cases suggest, suggest that the collisional quenching is not the only mechanism of interaction, but HEWL-resveratrol and insulin-resveratrol complexes actually form. In the case of insulin, the effect of the non-linear behaviour of the Stern-Volmer plot is particularly marked (Fig. 3.7 (f)). Thus, for static quenching, the Stern-Volmer constant could be interpreted as the binding constant K. In order to extract both the effective number of quencher-accessible sites and the binding constant, the eq. (2.28)

reported in Chapter 2 has been applied. The data for HEWL and insulin are reported in Fig. 3.7(c) and 3.7(g), respectively, and the parameter extracted from the fits are reported in Table 3.5. Additionally, to estimate the number of binding sites n and the association constant K_a for each protein, eq. 2.29 reported in Chapter 2 has been applied and the data are summarized in Table 3.5 and reported in Fig. 3.7 (d) and (h).

Table 3.4: Stern-Volmer constant and collision constant extracted from the fit of the data reported in Fig. 3.9 (b) and 3.9(f) of HEWL and insulin, respectively.

	K (M^{-1})	k ($M^{-1}s^{-1}$)
HEWL	$(3.6 \pm 0.2)*10^4$	$(3.6 \pm 0.2)*10^{12}$
Insulin	$(6.5 \pm 0.3)*10^4$	$(6.5 \pm 0.3)*10^{12}$

Table 3.5: Effective solvent-accessibility f_a and binding constant K from the fit of the data reported in Fig. 3.9 (c) and 3.9(g) and association constant K_a and number of the binding sites n extracted from the fit of the data reported in Fig. 3.9 (d) and 3.9 (h) for HEWL and insulin, respectively.

	f	K	$\log_a$	n
HEWL	0.63 ± 0.02	$(5.4 \pm 0.4)*10^4$	3.41 ± 0.08	0.8 ± 0.02
Insulin	0.25 ± 0.01	$(3.0 \pm 0.2)*10^5$	1.9 ± 0.2	0.52 ± 0.02

In the case of HEWL, approximately 63% of Trp residues (i.e. 3.78) seem to be accessible to the quencher. Docking study confirms this evidence, since resveratrol seems to preferentially bind to the regions (Trp62-Leu75) or (Ile98-Val107), actually located in the most hydrophobic portion of the protein, containing Trp62 and Trp63 and close to Trp108[253]. However, the number of the estimated resveratrol binding sites n per protein is closed to 1. Since Trp63 is non fluorescent, Trp62 or Trp108 could be considered part of the resveratrol binding site. Similar effects and similar binding sites were also found by studying the interaction between HEWL and curcumin, chloramphenicol, myricetin and many other compounds[254–256]. Non-

covalent interactions such as hydrogen bonding, aromatic interaction and van der Waals are the most recurrent mechanism of interaction between the two compounds[256 and literature therein].

Differently, in the case of insulin, only the 25% of the total tyrosine residues are accessible to the quencher. In fact, as estimated by the association constant K_a, resveratrol shows a lower affinity to bind to insulin compared to HEWL and, consequently, the number of binding sites per protein slightly reduces. Docking studies exploiting the interaction between insulin and other simple aromatic compounds highlighted that they actually bind to a specific portion of insulin B domain, composed by (Tyr26, Phe24, Phe25, Tyr16)[249,258–260], which is the same portion able to self- assemble independently[247]. In light of the quenching experiments and docking results, the formation of a complex between insulin and resveratrol is reported, where approximately one Tyr residue seems to be close to resveratrol binding site.

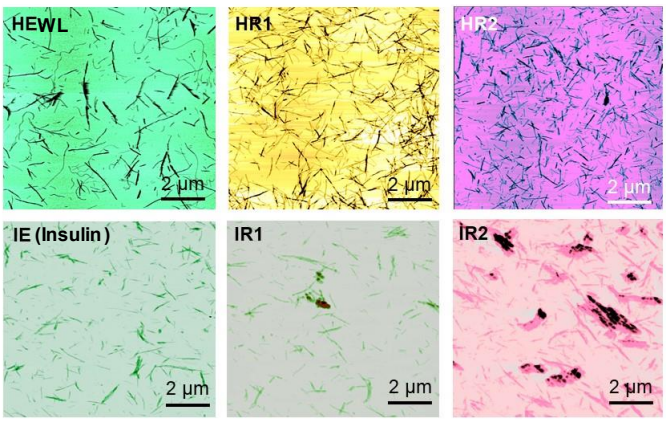
3.3.4 HEWL and insulin interaction with resveratrol upon fibrillation: morphological aspects

Polyphenols and natural compounds, in general, have been demonstrated to act as anti-amyloidogenic compounds[261]. In light of this results, the effect induced by the incubation of HEWL and insulin in presence of resveratrol has been evaluated. Since the interaction between proteins and ligands occurs via non-covalent interaction (H-bonding, van der Waals, aromatic and hydrophobic interactions), resveratrol has been used as one of the smallest aromatic model drugs to infer on the molecular mechanism of interaction.

Resveratrol has been incubated with HEWL and insulin in a concentration-dependent manner and in environmental conditions prone to induce aggregation. AFM images reported in Fig. 3.8 shows that resveratrol actually remodelled the structure of both HEWL and insulin fibrils in a dose-dependent manner. In order to quantify this qualitative behaviour, AFM images were statistically analysed to estimate the distribution of fibrils contour lengths with and without resveratrol. The

results of AFM analysis are summarized in Fig. 3.8, where the most probable value for fibrils contour length corresponds to the 50th percentile of the distribution. In particular, the long and well-separated HEWL fibrils obtained in absence of resveratrol appear gradually shorter and straighter by increasing the antioxidant concentration, whereas the contour length of the insulin fibrils is mildly modified by the presence of resveratrol, which actually induces the formation of amorphous aggregates by increasing resveratrol concentration (see Fig 3.8 left).

Noteworthy, the reduction of HEWL fibrils contour length in acidic condition does not depend on HEWL hydrolysis, a phenomenon widely discussed in literature when HEWL is incubated at pH<3[240]. HEWL, HR1 and HR2 have been obtained in the same acidic condition, thus the reduction of contour length depends on the interaction with resveratrol.



	Contour length (nm)
HEWL	640 [421:964]
HR1	445 [234:755]
HR2	300 [155:577]
Insulin (IE)	247 [151:391]
IR1	242 [153:412]
IR2	218 [143:352]

Fig. 3.8: (left) 10x10 µm^2 AFM topographic images of HEWL (HEWL in black, HR1 in red and HR2 in blue) and insulin (IE in black, IR1 in red, IR2 in blue) fibrils. False colour scales represent the 3D topographic height. (right) the distributions of fibrils contour length are presented. The most probable value for contour lengths as well as the 25th and 75th percentiles (in squared brackets) are summarized in the table.

Similarly, in the case of insulin, IR1 and IR2, we observed the formation of non-fibrillar spherical aggregates coexisting with mature fibrils of comparable length

with increasing the concentration of resveratrol, as observed also elsewhere[262,263]. These findings are compatible with literature, where the remodelling action induced by antioxidant on protein fibrils as well as the modification of the fibrillation pathway is widely reported [264,265].

3.3.5 FTIR spectroscopy probes secondary and tertiary structure modification upon the interaction between HEWL, insulin and resveratrol

The relevant shrinking of HEWL fibrils induced by dose-dependent interaction with resveratrol is not accompanied by significant secondary structure modifications as evidenced by FTIR analysis performed on HEWL, HR1 and HR2 samples reported in Fig. 3.9 (a). The results of FTIR band fitting of HEWL, HR1 and HR2 spectra are reported in Table 3.6, indicating that resveratrol induces an increased concentration of β-sheets structure (from 29% to 40%) and a reduced contribution of α-helix (from 3% to 10%) upon increasing its concentration.

Differently, resveratrol induces a modification in insulin secondary structure. In fact, as clearly visible from the second derivative spectra reported in Fig. 3.9(b), the presence of resveratrol induces the increasing of the α-helix contribution (1658 cm^{-1}) instead of the unordered and turns populations (1666 and 1670 cm^{-1}, respectively). Additionally, band fitting of insulin spectra highlights that the low energy β-sheets band (at approximately 1630 cm^{-1} in the IE case), blueshifts to 1627 cm^{-1} upon increasing the concentration of resveratrol as reported in Table 3.6. This evidence indicates a reorganization of the β-sheets architecture passing from shorter to longer ones when resveratrol concentration is doubled that of insulin.

More interestingly, from the FTIR spectra of insulin we observe a further 2 cm^{-1}-redshift of Tyr-OH band at 1515 cm^{-1} upon interaction with resveratrol. This redshift indicates a modification of H-bond network where Tyr residues are involved with respect to insulin mature fibrils[213,266–269]. Additionally, the intensity of this band increases with increasing the concentration of resveratrol. Interestingly, the low-

energy flank of amide I present a resveratrol-concentration dependent variation: the tyrosine doublet band at 1594-1606 cm^{-1} (ν_{8b}, ν_{8a}) modifies its relative intensity by increasing the 1606 cm^{-1} and by decreasing the 1594 cm^{-1} one (see Fig. 3.8(b))[213]. Noteworthy, those bands are sensitive to tyrosine ring deformations and symmetry losses, mainly due to intermolecular interactions, confirming that tyrosine residues could be involved in a direct interaction with the antioxidant or could reflect structural modifications resulting from the interaction between insulin and resveratrol.

Table 3.6: The impact of resveratrol on fibrils contour lengths and secondary structure probed by FTIR. Position of β-sheets extracted from FTIR spectra of HEWL, HR1, HR2, insulin, IR1, IR2 both fibrils together with the associated contour length distribution. The 25th and the 75th percentiles of the distribution are indicated in squared brackets.

	FTIR β-sheets position (2 cm^{-1}) and content (%)	FTIR α-helix position (2 cm^{-1}) and content (%)
HEWL	1626 (29%)	1656 (3%)
HR1	1626 (39%)	1656 (9%)
HR2	1624 (40%)	1656 (10%)
IE	1630 (32%)	1658 (7.5%)
IR1	1627,1631 (28%)	1658 (18%)
IR2	1627 (34%)	1658 (18%)

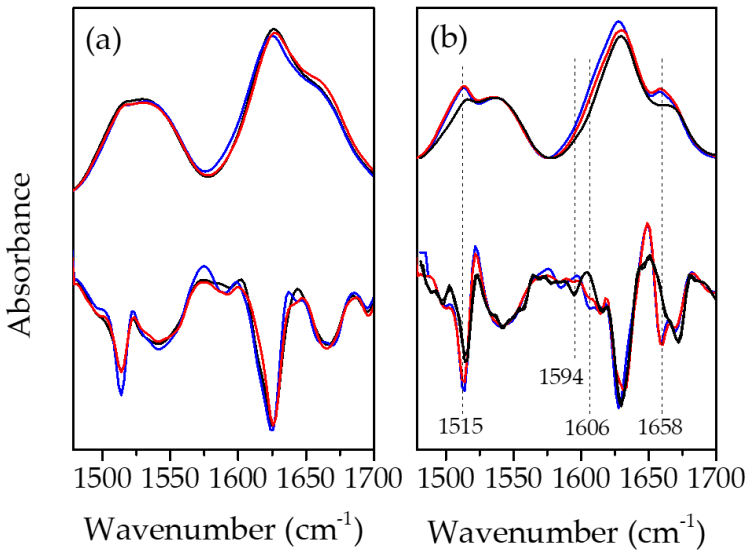


Fig. 3.9: FT-IR spectra of amide I and II bands (1450-1750 cm^{-1}) of (a) insulin fibrils in absence (in black) and presence of resveratrol (IR1 in red and IR2 in blue) and (b) of HEWL fibrils in absence (in black) and presence of resveratrol (HR1 in red and HR2 in blue). In (b) and (c) the respective second-derivative spectra presented.

3.3.6 Tertiary structure of HEWL and insulin fibrils in presence of resveratrol

The modification of protein tertiary structure upon fibrillation in presence of ligands is usually exploited in literature by fluorescence spectroscopy. In this case, intrinsic fluorescence spectroscopy has been used to indirectly investigate the effect of resveratrol during the fibrillation of both HEWL and insulin. Fig. 3.10 shows the fluorescence data obtained for HEWL (a) and insulin (b) at different concentration of resveratrol, namely when the protein:antioxidant ratio is (1:0) (in black), (1:1) (in red) and (1:2) (in blue).

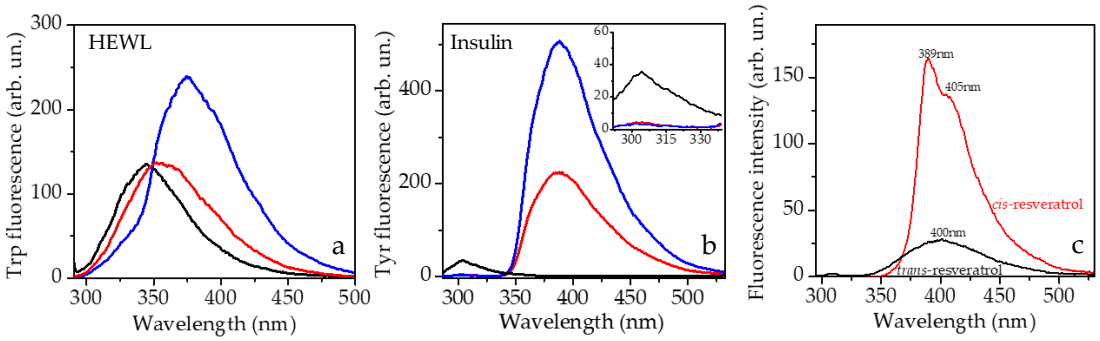


Fig. 3.10: (a) Trp fluorescence of HEWL fibrils solution in absence (in black) and in presence of resveratrol, i.e. HR1 (in red) and HR2 (in blue). (b) Tyr fluorescence of freshly prepared solution of insulin in absence (in black) and in presence of resveratrol, i.e. IR1 (in red) and IR2 (in blue). (c) *trans*- (in black) and *cis*- (in red) resveratrol spectra. Trans- isoform is characterized by a low quantum yield symmetric peak with a maximum at 400 nm, while cis- isoform by a high quantum yield contribution characterized by two peaks located at 389 nm and 405 nm.

Once HEWL fibrils are formed in absence of resveratrol, Trp maximum peak shifts from 342 nm to 345 ± 0.5 nm. Since Trp62 and Trp108 are the dominant emitters and Trp62 has been found buried in the hydrophobic core of the HEWL fibrils[240,270], it is likely that Trp108 results more solvent-exposed when HEWL fibrils are formed. The presence of resveratrol in the case of HR1 is characterized by a redshifted Trp emission peak located 352± 0.5 nm, indicating that the interaction with resveratrol during fibrillation actually stabilizes HEWL fibrils in a different conformational structure with a more solvent-exposed Trp residues. Docking studies reported that resveratrol could bind to specific portions, involving Trp62[177]. In light of this

consideration, we could speculate two possible effects induced by the presence of resveratrol at this specific concentration: a possible scenario could suggest that resveratrol provokes the complete solvent-exposure of Trp108, and the other possibility is that resveratrol binds to Trp62 during fibrillation and stabilizes protein in a different fibrillar conformation, with a more solvent-exposed hydrophobic core. In the case of HR2, the emission peak results further redshifted to 376 ± 0.5 nm and increases its quantum yield, which is actually not compatible with a typical Trp signal. In fact, in this case, the protein contribution is completely hidden by that of resveratrol which is no more in a totally *trans-* conformation. In fact, the *trans-* to *cis-* isomerization of resveratrol provokes the observed hyperchromic effect with a modification of the emission peak shape as reported in Fig. 3.10(c). However, in the case of HR2, the emission peak does not resemble that of a *trans-* nor that of a *cis-* resveratrol one (see Fig. 3.10 (c)). Hence, we can conclude that the peak detected in the HR2 spectrum may arise from a mixed population of free and bound resveratrol in both the conformations, or may arise from resveratrol blocked in an alternative conformation by the binding with HEWL. Since proteins underwent fibrillation stored in a dark place, the *trans-* to *cis-* isomerization is not induced by UV light, but it is actually promoted by the direct interaction with HEWL [271–273].

At the same time, Fig. 3.10(b) shows that insulin fibrils obtained in presence of resveratrol (IR1 and IR2) presents a strong intensity damping of the Tyr emission peak compared to insulin fibrils. More importantly, the fluorescence spectra of IR1 and IR2 are completely dominated by the signal arisen from resveratrol, showing a single broad peak at 388 nm. The hyperchromic effect observed even in these cases suggests that the interaction between insulin and the antioxidant promotes the emergence of a resveratrol emission peak, whose shape does not resemble that of its *trans-* nor that of its *cis-* conformations. Differently from HR2, the maximum peak of the resveratrol contribution when interacting with insulin appears redshifted [272].
Thus, due to the hyperchromic effect of resveratrol and to the insensitiveness of Tyr bands to protein fibrillation and interaction with ligands, an alternative tool has to be used to shed light into protein tertiary structure when fibrils are formed, especially in presence of a fluorescent ligand. Due to its exclusively sensitiveness to

protein chromophores, UVRR spectroscopy has been used to discern about the structural modifications provoked by resveratrol on both proteins.

In particular, Fig, 3.11(a) and (b) shows the 228 nm-UVRR spectra HEWL, HR1, HR2 and insulin, IR1, IR2, respectively. In the case of HEWL (Fig. 3.11 (a)), the 228 nm-UVRR spectra show a strong damping of the symmetric phenyl ring mode of Trp (W3 band) in a resveratrol concentration-dependent manner. To better investigate this peculiar behaviour, the spectra were fitted in order to quantify the relative area variation of W3 band with respect to the W1 one and to evaluate a possible W3 band shift. The results reported in Table 3.7 indicate that, in the HR2 case, the W3 band of

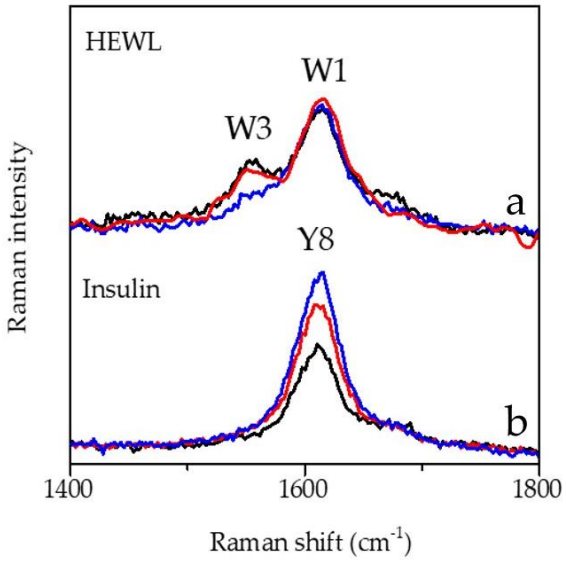


Fig. 3.11: UVRR spectra of (a) HEWL (in black), HR1 (in red) and HR2 (in blue) and (b) of insulin (black), IR1 (red) and IR2 (blue) obtained using 228 nm of excitation wavelength and presented in the region 1400-1800 cm^{-1}.

Trp residues results quenched and redshifted. As reported elsewhere[232], W3 band position is mostly sensitive to the variations occurring in the torsional angle $\chi^{2,1}$ of proteins. The observed modification of W3 peak intensity and position indicates that the interaction with resveratrol affects and modifies HEWL torsional angles[232,274].

Interestingly, the modification of peptide torsion angle occurs only in the HR2 case, when possibly a mixed population of *trans-* and *cis-* resveratrol occurs. Since UVRR spectra of HR1 and HR2 fibrils are mainly dominated by proteins' signal, we could deduce that the effect observed on W3 band are direct consequences of the Trp-resveratrol interaction in a concentration dependent manner. As observed by

molecular dynamic simulation, the modification of dihedral angle of protein residues actually is crucial for binding affinity and can be strongly modified by the interaction with small aromatic compounds [275].

Table 3.7: Resveratrol changes the relative intensity of Trp peaks in a concentration-dependent manner. W3/W1 area ratio and W3 position of 228 nm-UVRR spectra of HEWL, HR1 and HR2.

	W3 area/W1 area	W3 position(cm^{-1})
HEWL	0.63 ± 0.04	1557 ± 2
HR1	0.6 ± 0.03	1557 ± 1
HR2	0.33 ± 0.03	1554 ± 1

Conversely, the 228nm-UVRR spectra of insulin show that the Y8 band (in-plane stretching of the ring) of the fibrils increases with increasing the concentration of resveratrol. Similar results have been obtained by Wang and colleagues studying the interaction between Aβ proteins and myricetin[276]. In fact, they reported that the intensity of both Phe and Tyr bands increased upon addition of myricetin, which actually inhibits the formation of Aβ fibrils, concluding that the inhibition occurs consequently to the ligand-aromatic amino acid interaction[277]. Therefore, the formation of insulin amorphous aggregates and the increasing of the UVRR Tyr band intensity in a resveratrol concentration-dependent manner, could be the results of the direct interaction between tyrosine residues and resveratrol. The interaction with resveratrol could increase the hydrophobicity of Tyr microenvironments or could underlying the formation of a strong H-bond where Tyr are involved[278]. Hence, the formation of non-fibrillar aggregates in a resveratrol concentration-dependent manner together with the modification of the 1515 cm^{-1} and 1594-1606 cm^{-1} bands detected by FTIR spectroscopy in Fig. 3.9(b), which are sensitive to ring deformation induced by the establishment of intermolecular interactions where Tyr residues are involved, lead us to infer on the involvement of tyrosine residues as direct resveratrol binding sites via intermolecular forces.

3.4 Discussion

Protein fibrillation as well as the mechanism of interaction between proteins and potential drugs are currently hot topics in science. Although fibrillar structure share a cross-β representative structure, investigating how a specific protein is structurally arranged once fibrillated is crucial for drug design.

HEWL and insulin have been used as model proteins and their β-sheets-rich structure once fibrillated have been elucidated by both FTIR and UVRR spectroscopy. In both cases, fibrillation induces an intensity damping of the UVRR Trp and Tyr bands, indicating their major solvent exposure once mature fibrils are formed. Thanks to the Trp fluorescence results and literature data, we can speculate that fibrillation of HEWL induces a progressive solvent exposure of Trp108, while maintaining Trp62 buried in the hydrophobic core.

Differently, in the case of insulin, intrinsic fluorescence spectroscopy does not provide clear information regarding the state of Tyr residues. In fact, the insensitivity of Tyr fluorescence peak to solvent-effect has been overcome by UVRR and FTIR spectroscopy, looking at specific markers able to report the H-bonding state of Tyr residues. UVRR Y8b spectral position, Tyr Fermi doublet relative intensity modification, the Tyr intensity damping as well as the intensity and energy modification of the Tyr bands detected by FTIR spectroscopy, indicate that Tyr residues seem to be on average more water exposed when fibrils are formed, reporting a modification of the H-bonding network where they are involved.

In order to unravel the mechanism of interaction between resveratrol and both proteins, we performed quenching experiments. The formation of HEWL:resveratrol and insulin:resveratrol complexes has been demonstrated, estimating an average of 0.8 and 0.52 binding sites, respectively. The interaction with resveratrol during HEWL and insulin fibrillation induces the formation of different morphological species in a dose-dependent manner: in the case of HEWL, fibrils become straighter and shorter with increasing the concentration of the antioxidant, while in the case of insulin, the formation of spherical aggregates mixed with fibrils is reported with increasing the concentration of resveratrol.

Intrinsic fluorescence data evidenced that the presence of resveratrol during HEWL fibrillation induces the formation of fibrils with a different structural arrangement of

Trp residues, which result more exposed to solvent. Since Trp62 has been hypothesized as a possible binding site of resveratrol and other aromatic ligands[177,279,280], and in the case of HEWL fibrils is found buried in their hydrophobic core[270,281], it is likely that resveratrol binds to Trp62 letting the residue unable to be part of the hydrophobic core. However, in the HR2 case, the fluorescence spectrum is completely dominated by the signal arisen from resveratrol and no information could be derived from this spectrum. Interestingly, the ligand is found not in a *cis-* nor in a *trans-* conformation, and this partial conversion cannot be ascribed to a UV-exposure of the solution.

Similarly, when resveratrol is incubated with insulin during fibrillation, it completely quenches Tyr intensity and completely dominates the fluorescence spectra of both IR1 and IR2, assuming not a *trans-* nor a *cis-* conformations. The effect of resveratrol concentration, incubation temperature and low pH could surely affect the stabilization of the *trans-* population[282]. However, without UV exposure for days at the resveratrol concentration explored, *trans-* conformation remains stable[282]. Additionally, when protein-resveratrol concentration ratio is (1:1) we observed that resveratrol has a different impact on HEWL and insulin, since in the former case the signal of HEWL is still well-distinguishable from the antioxidant one. This evidence indicates that the interaction with HEWL and insulin could affect the stabilization of the *trans-* conformation, depending on the protein studied[132].

Even in this case, UVRR spectroscopy has been used to overcome the limitations of intrinsic fluorescence spectroscopy, thanks to its sensitivity only to peculiar protein chromophores vibrational modes. In the case of HR2, when an increased fluorescence quantum yield is detected, the W3 band of Trp residues resulted particularly quenched. Since W3 is a marker of the modification of dihedral angle of the Trp residues, particularly sensitive to the interaction with ligands, it could suggest that Trp residues indole rings are on average relevantly perturbed by the interaction with resveratrol[283]. Since docking study suggests that Trp62 is involved in the interaction with resveratrol and the number of estimated binding site is approximately 1[177], we speculate that the UVRR and fluorescence Trp modifications are induced by the involvement of Trp62 in the interaction with resveratrol. In the

case of insulin, the increasing of the Y8a band intensity suggests that tyrosine residues are sensitive to the modification of their microenvironments hydrophobicity. In particular, the intensity increasing of this band is likely due to the direct interaction between tyrosine residues and resveratrol, similarly to the case reported by Wang and colleagues, probing the interaction between Aβ and myricetin by UVRR spectroscopy[276]. Since FTIR results suggest that Tyr-OH bands changed their intensity and the 1515 cm^{-1} peak furtherly red-shifts upon adding resveratrol in a concentration-dependent manner, we could deduce that resveratrol interacts with insulin, perturbing the H-bond network where Tyr residues are involved.

3.5 Conclusions

The potential of UVRR spectroscopy in monitoring modifications occurring in protein secondary and tertiary structure upon fibrillation and interaction with a hydrophobic ligand has been exploited in combination with FTIR spectroscopy, fluorescence spectroscopy and AFM. Specifically, due to its high sensitivity to hydrophobicity modifications of aromatic residues micro-environment, UVRR is more informative than intrinsic fluorescence spectroscopy when proteins do not contain Trp residues. Additionally, UVRR exploiting electronic resonances allows enhancing only selected proteins vibrational modes peculiar of the protein-ligand interaction, thus decreasing the background caused by unwanted spectral contributions. FTIR spectroscopy, effective to probe protein secondary structure, complemented with morphological information that can be extracted from AFM measurements, allows for a more comprehensive interpretation of the UVRR data. To benchmark the potential of UVRR, this multi-technique approach has been employed on two model proteins prone to fibrillation (HEWL and insulin), studying as well their interaction with a model ligand: resveratrol. We observed that HEWL and insulin fibrils morphology is completely remodelled by the incubation with resveratrol in a concentration dependent manner: in the case of HEWL fibrils becomes shorter and straighter, while in the case of insulin resveratrol induces the formation of non-fibrillar aggregates. Differently to fluorescence spectra where the

resveratrol completely hides the proteins' contribution, UVRR offers the possibility to follow the proteins' tertiary structure modification due to a different concentration of resveratrol, by completely removing its contribution by selectively probe peculiar proteins vibrations. However, UVRR manifests its powerfulness in the case of Trp-lacking proteins, since Tyr fluorescence peak is completely insensitive to solvent-effect and to non-covalent interactions. Together with FTIR results, UVRR demonstrate that part of the Trp and Tyr residues of HEWL and insulin could be considered close to the resveratrol binding sites.

Indeed, a reliable interpretation of UVRR spectra of proteins undergoing fibrillation has been proposed in this Chapter, providing a firm multi-technique experimental approach that can be followed for the investigation of biologically more relevant systems such as IDPs and of their interaction with drugs.

From a general point-of-view, the characterization of fibrillar molecular architecture of model proteins by UVRR spectroscopy is important not only to answer fundamental questions related to protein folding, but it is also crucial for the development of therapeutics interventions and for diagnostics goals.

Chapter 4

The peculiar effect of spermine on Dihydrofolate Reductase

Enzymes are an important class of proteins able to promote the catalysis of a peculiar reaction and are fundamental for living cells, as anticipated in Chapter 1. A widely known and studied enzyme is Dihydrofolate Reductase (DHFR), which is a key protein in folate metabolism of prokaryotes and eukaryotes. Due to its small size, to the easy production and to the extensive biochemical and structural investigation, DHFR is still considered a model enzyme.

DHFR catalyses the NADPH-dependent reduction of dihydrofolate (DHF) to tetrahydrofolate (THF), a precursor of cofactors required for the biosynthesis of purines, pyrimidines and several amino acids. However, since DHFR is involved in the mechanism of cell growth and proliferation, it has been considered as a molecular target for the development of antineoplastic and antimicrobial drugs. Two well-known molecules, namely methotrexate (MTX) and trimethoprim (TMP), have been developed and administered for at least 60 years to the population in order to inhibit the enzymatic activity of DHFR in pathological situations. The enzymatic inhibition by MTX and TMP has been extensively studied, finding that their interaction actually confined DHFR in a stable inactive and low-energetical conformation. Due to the high affinity to DHFR compared to NADPH and DHF, both MTX and TMP do not induce DHFR misfolding or the recovering of the native active state.

Despite many studies have elucidated the structural features of the intermediate conformations where *E. coli* DHFR is blocked by inhibitors, very little is known about different high-energy conformations that could have a potential impact both on catalysis and misfolding/aggregation propensity[284].

In collaboration with Prof. Stefania Pucciarelli and dr. Giusi Serena De Fronzo of University of Camerino, the interaction between DHFR and spermine has been investigated. Spermine is a polyamine involved in many cellular processes, whose metabolism is fundamental for cell growth and differentiation. Interestingly, DHFR and polyamines are both involved in mechanism related to cell proliferation; in fact, THF, namely the active form of folic acid, participates in different enzymatic reactions related to the synthesis of DNA, RNA and proteins, covering a key role in the biosynthesis of polyamines.

In this study, we combine the 228 nm-UVRR results with the preliminary data reported by dr. De Fronzo in her PhD thesis, for studying the interaction between DHFR (20μM, 40μM) and spermine 4mM (pH 7.3, Hepes buffer). The UVRR data, complemented by Trp fluorescence spectroscopy, circular dichroism and TEM data, highlight that spermine induces a partial inhibition of the catalytic activity of DHFR, blocking the enzyme in a high energy conformer that shows a tendency to slowly aggregate and form fibrils at neutral pH. Since DHFR has 5 tryptophan residues, the effect of spermine on the enzyme tertiary structure have been evaluated using UVRR spectroscopy in order to probe the modification of their microenvironments and the eventual establishment of non-covalent interaction, following also the structural modification induced by the spermine-dependent fibrillation.

4.1 DHFR: structure, function and inhibition

4.1.1 Dihydrofolate Reductase structure and function

Dihydrofolate reductase (DHFR) from *E. coli* is an 18kDa enzyme consisting of an α/β structure. DHFR is characterized by eight stranded β-sheets, four α-helices and loops[285]. Its active site cleft divides the protein into two sub-domains: the adenosine

binding subdomain and the major subdomain. The former comprises the residues 38-88 and provides the binding site for the adenosine moiety of the cofactor NADPH, while the latter is formed by 100 residues and it is characterized by three loops surrounding the active site. These loops are the Met20 (residues 9-24), the F-G (from 116 to 132) and the G-H (from 142-150), with the Met20 loop directly covering the active site, protecting it from the solvent. Both cofactor, inhibitors, and substrate bind to the active hydrophobic cleft, while F-G and G-H loops stabilize the structure via hydrogen bond interaction with Met20 loop.

DHFR structural flexibility promotes a catalytically-activated transfer of a hydride from NADPH to 7,8-dihydrofolate (H_2 folate, DHF) with the protonation of 5,6,7,8-tetrahydrofolate (H_4 folate, THF), which is the reactive form of folic acid and critical in DNA nucleotides' biosynthesis. As explained above, the flexibility and the conformational mobility of Met20 loop, together with that of the other loops near the active site, are key factors in promoting the release of THF. The conformational freedom of Met20 loop helps in the nicotinamide ring stabilization of NADPH in order to promote the transfer of the hydride from NADPH to DHF. The catalytic cycle follows a specific pathway composed by multiple steps, incorporating five important intermediates (see Fig. 4.1): firstly, NADPH and the substrate attach to the binding site of DHFR, then the protonation and the hydride transfer from the cofactor NADPH to the substrate take place. The last two steps comprises the oxidized cofactor NADP+ release and, finally, the THF release. The enzymatic activity of DHFR depends on pH since the binding site of DHFR is in contact with water. Asp27 is the only charged hydrophilic residue in the binding site and plays a critical role in the catalytic mechanism, providing the protonation of the substrate. In order to follow its catalytic pathway, the conformation of the Met20 loop has to adapt to each step of the chemical reaction. In fact, DHFR Met20 loop adopts an occluded conformation when substrate is bound, whereas it becomes closed when also the nicotinamide ring of NADPH binds to its DHFR binding site, protecting the active site from the solvent.

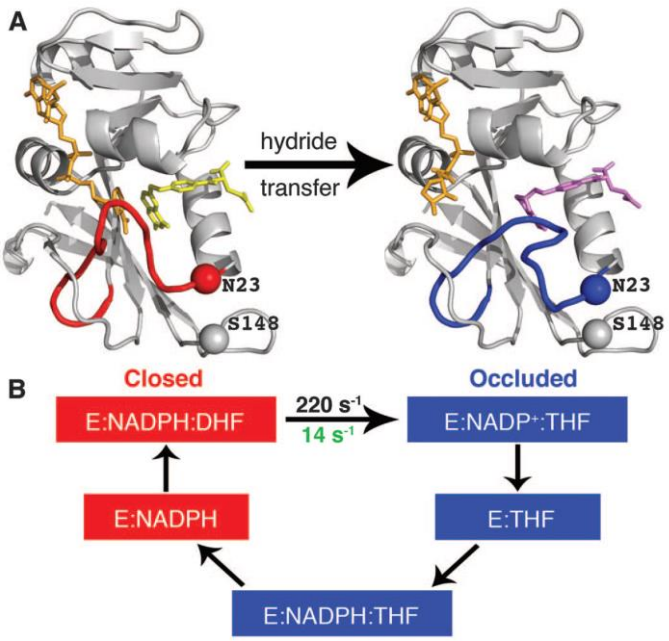


Fig. 4.1: Conformational changes during *E. coli* DHFR catalytic cycle. (A) On the left: the complex formed by DHFR:NADP+:DHF (model for the Michaelis complex) with the Met20 loop (depicted in red) in a closed conformation. On the right, the complex formed by DHFR:NADP+:THF when Met20 (in blue) is in a occluded conformation. NADP+ is depicted in orange, while DHF and THF in yellow and purple, respectively.(B) intermediate steps of the catalytic cycle. The figure has been readapted from[286].

4.1.2 The inhibitory effect of TMP on DHFR

The inhibition of DHFR enzymatic activity results in a reduction of DNA nucleotides biosynthesis, which in turn blocks cell proliferation. Thus, an increasing number of pharmaceuticals have been developed since 1940, considering their positive effects against tumour cells and bacteria proliferation. Methotrexate and trimethoprim are two well-known antifolates used to inhibit the enzymatic activity of DHFR[287]. Computational models and DHFR point mutations experiments helped in characterizing the molecular mechanism leading to inhibition by these two analogues. Despite their effect on the enzyme is similar, TMP and MTX bind to different binding sites located in the DHFR's active site cleft and their binding affinity to the protein is three to more orders of magnitude higher than folate substrates. As such, all of these inhibitors are important model for host-guest molecular recognition.

In particular, trimethoprim is a synthetic, broad-spectrum antibiotic and successful inhibitor of bacterial DHFR (Fig. 4.2). It is mainly used for the treatment of bladder and urinary tract infections. Due to its high affinity compared to DHF, TMP preferentially binds to DHFR, preventing the conversion of DHF to THF.

The widespread use of TMP by the population has driven the development of antibiotic-resistant DHFRs, the structures of which allow a lower binding affinity of the drug while retaining its enzymatic activity. For this reason, the understanding of the molecular mechanism of interaction between DHFR and TMP is vital for the development of new TMP-like analogues, being characterized by a high affinity to TMP-resistant DHFR strains.

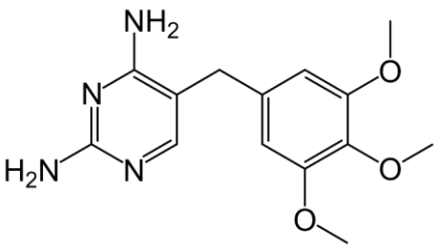


Fig. 4.2: Structure of trimethoprim.

The complex DHFR/TMP has been widely studied in literature by multiple techniques, highlighting that TMP actually binds to DHFR active site, where its pyrimidine ring occupies the interior of the hydrophobic cleft, via a plethora of different non-covalent interactions, namely van der Waals and π-π interactions, hydrogen bonding and hydrophobic interactions[288–290]. Recently, NMR experiments elucidated the interaction between TMP and two DHFR mutants (P21L and W30R, near the active sites) demonstrating the formation of a less stable complex with the drug, since they promote modulation of the protein structure and modify also the rigidity of the Met20 loop[288,290].

4.1.3 The inhibitory effect of spermine on DHFR

Firstly reported in 1678, spermine is an aliphatic polyamine that, together with spermidine and putrescine, are normal cellular constituents. Polyamines play important roles in many cellular processes including the regulation of transcription and translation, modulation of kinase activity, control of cell proliferation and differentiation. Additionally, they control the activity of ion channels, protect from oxidative damage and maintain the membrane structure and function. In mammalian cells, polyamines are found in millimolar concentration[291,292], but their presence decreases with age in many organisms. Although they ameliorate oxidative stress resistance, abnormalities in the control of polyamine metabolism and uptake might be responsible for high levels of polyamines in cancer cells.

Polyamine functions can be predicted from their chemical structure; since these molecules are positively charged at physiological pH, exerting both electrostatic and hydrophobic interactions. The ability of polyamines to alter DNA-protein and protein-protein interactions might be disruptive to cellular functions. As previously demonstrated by dr. De Fronzo in her PhD thesis, polyamines have an effect on the enzymatic activity of *E. coli* DHFR. While putrescine did not affect the catalytic activity, spermidine and spermine induced a partial inhibitory effect. Additionally, spermine behaves as non-competitive inhibitor toward DHF and as uncompetitive inhibitor toward NADPH. Spermine seems to act as a DHFR enzymatic activity regulator, blocking DHFR structure in a less efficient allosteric form, while stabilizing the interactions involved in the Met20 loop closure. Thus, a great interest in the mechanism of inhibition as well as in the investigation of the structure of DHFR complexed with spermine has emerged in recent time.

4.2 Experimental methods

4.2.1 Sample preparation

Dihydrofolate reductase was overexpressed and isolated in the laboratory of dr. S. Pucciarelli (University of Camerino) from *E. coli* strain AG-1 (Stratagene) containing the plasmid pWT1-3, kindly supplied by Dr. C. R. Matthews (University of

Massachusetts Medical School). All the experiments have been carried out with a highly purified (> 99%) dihydrofolate reductase from *E. coli* cells. The concentration of *E. coli* DHFR was determined spectrophotometrically using a molar extinction coefficient of 31100 M^{-1} cm^{-1}). The specific activity of this preparation was 3.39 Units/mg. Before use, the purified protein, stocked in 90% ammonium sulphate, was buffer-exchanged through two changes of dialysis against 2 L of potassium phosphate 10mM and HEPES buffer 10mM pH 7.3.

Solutions containing 50 mM HEPES and 50 mM of polyamine spermine dihydrate (Fluka BioChemika) were prepared by dissolving the powders together, adjusting the pH with HCl to the value of 7.3. Lower concentration solutions were obtained through stepwise dilutions with 50 mM HEPES buffer, checking the pH.

4.2.2 UV Resonance Raman spectroscopy

UVRR spectroscopy experiments were carried out by using the setup descripted in Chapter 1. Spectra were collected at 228 nm, using a holographic grating of 1800 g/mm, reaching a spectral resolution of 20 cm^{-1} in order to achieve a good signal to noise ratio. The power of the beam on the sample was maintained at 17 μW and the sample were horizontally oscillated in order to avoid photodamage and local heating of the sample. UVRR measurements have been collected at 278 K. Original spectra were treated removing the blank spectrum (only addressed to electronic noise), the solvent and the empty cell contributions. Finally, a polynomial baseline has been subtracted in order to remove the small fluorescence contribution present on the spectra.

4.3 Results

The inhibitory effect of spermine on native *E. Coli* DHFR has been monitored by measuring the enzyme residual activity upon the addition of a different concentration of spermine (Fig. 4.3). Noteworthy, the enzymatic activity of native *E. Coli* DHFR lowers to 58% in a spermine concentration-dependent manner. Since the inhibition of the enzymatic activity is due to the stabilization of the enzyme in an

inactive conformation, the characterization of the structural arrangement of DHFR interacting with spermine is of crucial importance. Noteworthy, it is of great importance also understand if the mechanism of inhibition is similar to one of those reported for MTX or TMP. In order to evaluate these aspects, we compared the effect of spermine and of TMP on DHFR.

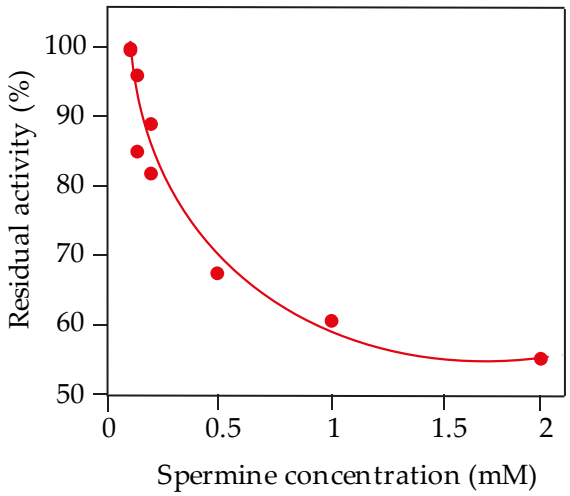


Fig. 4.3: residual activity of DHFR upon addition of different concentration of spermine. The picture has been readapted from the PhD thesis of dr. De Fronzo after the permission of prof. Pucciarelli.

The 228 nm-UVRR spectra of non-interacting DHFR at two different concentrations (40 and 20 μM) are presented in Fig. 4.4 (a) and they are compared to the spectrum of DHFR+TMP (see Fig. 4.4 (b) and Fig. 4.4 (c)). As a first remark, DHFR spectra taken at two different concentrations do not show spectral modification or peak intensity quenching, with the exception of a Phe band located at 790 cm^{-1} that is characterized by a high intensity peak when DHFR is measured at 20μM. Anyway, both spectra are completely dominated by tryptophan residues contributions (W*), with the exception of the Y8 peak addressed to tyrosine and located approximately at 1614 cm^{-1}.

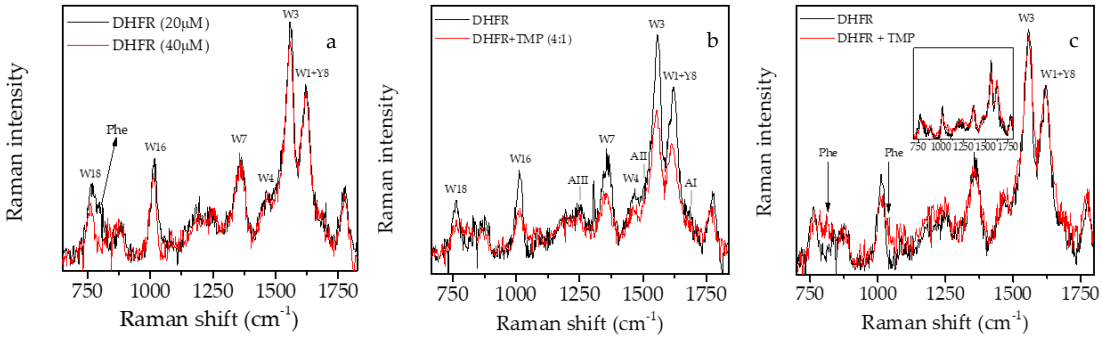


Fig. 4.4: (a) 228 nm-UVRR spectra of DHFR collected at 40μM (in red) and at 20 μM (in black). The assignment of vibrational peak is reported in agreement with the Table 2.3 in Chapter 2. (b) 228 nm-UVRR spectra of DHFR collected at 40μM (in black) and DHFR+TMP (40μM:10μM) (in red). (c) 228 nm-UVRR spectra of DHFR collected at 40μM (in black) and DHFR+TMP (40μM:10μM) (in red) normalized to the W1+Y8 band at 1621 cm^{-1}. The inset shows the same UVRR spectra with the DHFR concentration lowered to 20μM.

As reported in Fig. 4(b), the addition of TMP actually quenches the intensity of all the Trp residues bands of DHFR, without actually inducing relevant spectral modification of amide bands (see Fig. 4.4 (b)). Band fitting confirms that secondary structure remains almost α-helix-like by the inhibition of TMP and the results are reported in Table 4.1. In Fig. 4.4 (c), UVRR spectra of DHFR (40μM) and DHFR+TMP (40μM:10μM) are normalized to the intensity of the W1+Y8 band approximately at 1621 cm^{-1}, in order to detect minor spectral differences. The normalization of the spectra actually highlights that Trp bands quenches of a constant factor (approximately 1.5) when TMP is added, but no remarkable spectral shape modification of Trp bands can be observed. It is likely that the interaction with TMP does not directly involve Trp residues. Differently, Phe bands located at 790 cm^{-1} and 1030 cm^{-1}, addressed to symmetric in-plane ring breathing and CH bend respectively, actually increase their contributions with the addition of TMP, possibly suggesting the establishment of a non-covalent interaction between the aromatic ring of Phe and TMP. The inset of Fig. 4.4 (c) report the UVRR spectra of DHFR (20μM) and DHFR+TMP (20μM:10μM), normalized at the W1+Y8 band, indicating that the effect of TMP on DHFR does not depend on protein concentration.

In Fig. 4.5, the 228-nm UVRR spectra of DHFR, DHFR + TMP and DHFR + spermine just prepared are collected, both when DHFR is measured at 20 μM and 40 μM are presented and normalized to the W1+Y8 band (see Fig. 4.5(a)). By this normalization,

the effect of spermine on UVRR vibrational modes can be easily pointed out. As a first remark, only peculiar Trp-related UVRR bands result more damped respect to the DHFR+TMP spectrum, namely W3 (~1558 cm^{-1}), W7 (~1350 cm^{-1}) and W16 (~1009 cm^{-1}) bands. Those bands are particularly sensitive to Trp residue torsion angle, to the hydrophobicity of Trp microenvironments and to the establishment of non-covalent interactions where Trp residues are involved, respectively[135,220]. Due to the low spectral resolution at this energy, we could not extract information on Trp microenvironments' polarity from the band fitting of W7 Fermi doublet peak, since the two peaks (1340-1360 cm^{-1}) appear convoluted into a broader one. However, from the position of the W17 band at 879 cm^{-1}, we could ensure that most of the Trp residues are H-bonded[220]. The strong damping of the W3 band respect to the other Trp bands and to the TMP case once spermine is added, actually indicates that the mechanism of interaction between spermine and DHFR is different from that of TMP. Trp residues and, more importantly, their torsional angles appear strongly perturbed by the interaction with spermine. In Table 4.1, the relative intensities $W3/_{W1+Y8}$ extracted from the band fitting are presented. The effect of spermine on 20μM DHFR is similar to that obtained at higher concentration of the enzyme, and for this reason is not presented. Additionally, band fitting of the UVRR spectra reported in Fig. 4.5 (a) highlights that the addition of spermine induces the formation of a β-sheets-rich secondary structure population (see Table 4.1), while the α-helix/unordered secondary structure typical of DHFR is not perturbed by the presence of TMP.

Table 4.1: Band fitting results of UVRR spectra of DHFR, DHFR+TMP and DHFR +spermine using 40μM of DHFR. The ratio of W3 over W1+Y8 intensity is reported for each case with the position of Am-I band.

	DHFR	DHFR+TMP	DHFR+spermine (4 days)
W3/W1+Y8	1.9 ± 0.2	1.5± 0.2	0.9± 0.1
Am-I position (cm^{-1})	1658 ± 1	1645 ± 2	1673 ± 3

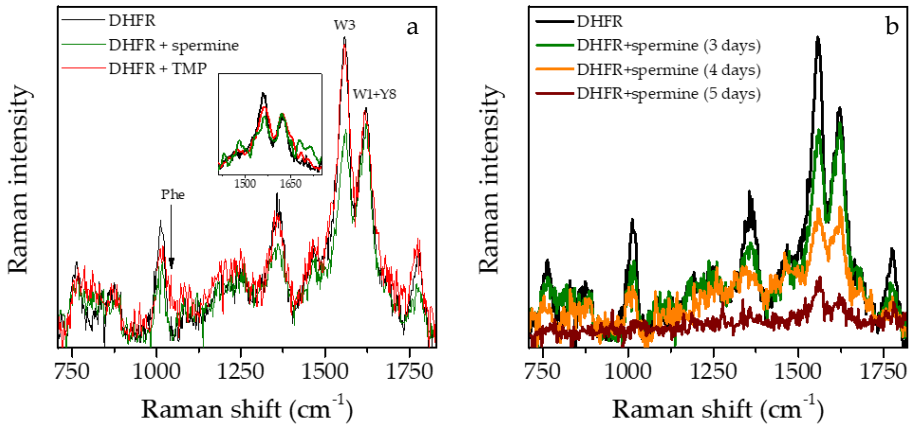


Fig. 4.5: (a) UVRR spectra of DHFR (40μM, in black), DHFR+spermine (40μM:4mM; in green) and DHFR+TMP (40μM:10μM; in red) normalized to the W1+Y8 band. The inset shows the UVRR spectra obtained with DHFR at 20 μM. (b) UVRR spectra of DHFR (depicted in black) and the complex DHFR+spermine taken at different days on incubations, namely 3 days (in green), 4 days (in orange) and 5 days (in dark red).

However, while the TMP stabilizes DHFR in an inactive low energy conformation during time, the effect of spermine has been monitored over time taking measurements at 3, 4 and 5 days of incubation with DHFR. In Fig. 4.5 (b) the UVRR spectra of DHFR + spermine are presented: the 3days-UVRR spectrum is identical to that collected at 0 days and for this reason only one of them is presented, indicating that the Trp side chains are stabilized in the same conformation for 3 days. However, as clearly visible in Fig. 4.5 (b), the intensity of the Trp-related bands results quenched during the incubation: even if after 4 days of incubation, the relative intensity of the Trp peaks of DHFR+spermine are maintained, after 5 days a strong intensity damping is clearly reported. A similar effect usually underlies protein photodegradation or the precipitation of pre-existing insoluble aggregates in solution. Unexpectedly, spermine at the reported concentration is able to induce the formation of fibrillar species of DHFR after 4 days of incubation even at neutral pH as demonstrated by TEM images in Fig. 4.6 (d). More importantly, the formation of DHFR fibrils at neutral pH has never been reported in literature. Preliminary results obtained by dr. De Fronzo showed that fibrillation is observed neither after the incubation of a lower concentration of DHFR (20μM) with the same amount of spermine used in this section (Fig. 4.6(e)), nor after 4-5 days of DHFR incubation at 37°C (data not shown). Thus, this evidence implies that both the presence of

spermine at the concentration used in this work, and the concentration of DHFR play a leading role in the fibrillar aggregates formation and in the fibrillation kinetics.

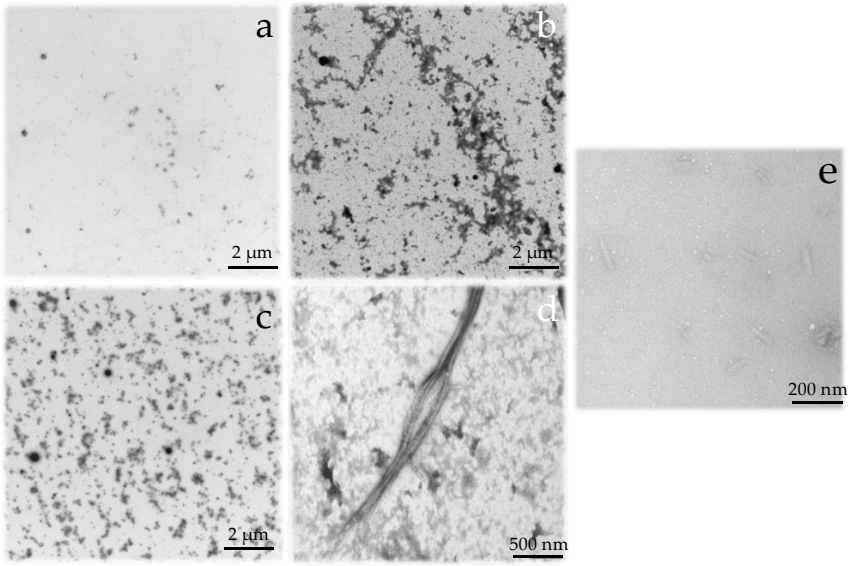


Fig. 4.6: TEM images of (a) HEPES buffer at pH 7.3; (b) DHFR (40μM) in HEPES 10mM, pH 7.3; (c) spermine (4mM) in HEPES buffer; (d) DHFR+spermine in HEPES buffer, pH 7.3 and (e) DHFR (20μM)+spermine in HEPES buffer, pH 7.3.

4.4 Discussion

The recent findings presented in the previous section represent a continuation of a project started in collaboration with dr. Stefania Pucciarelli at the University of Camerino. The spectroscopic information presented in this thesis have been complemented by the results of Trp fluorescence spectroscopy, circular dichroism and differential scanning calorimetry, already reported in the PhD thesis of dr. De Fronzo who demonstrated that spermine induces a partial inhibition of DHFR catalytic activity, by destabilizing its structure differently from well-known antifolate analogues.

UVRR spectroscopy has been used taking advantage of the possibility to work in resonance condition, by selectively exciting Trp residues vibrational modes. Noteworthy, since both spermine and TMP are not resonantly enhanced using 229 nm-excitation wavelength, the modification of the DHFR UVRR spectrum in presence of ligands could be ascribed entirely to the enzyme conformational changes

upon binding with drugs. As widely demonstrated in the Results section, 228 nm UVRR spectra of DHFR is completely dominated by the Trp residues vibrational bands which quench their intensity upon the addition of TMP and spermine.

To unravel the mechanism of interaction between spermine and DHFR and to elucidate if Trp residues are involved in the interaction, we collected also the spectrum of the complex DHFR and TMP, whose binding sites and molecular mechanism of interaction have been already reported. To exclude protein concentration-dependent effect, two different concentrations of DHFR have been used and no relevant spectral modifications have been detected.

E. coli DHFR is characterized by the presence of 5 Trp residues among which Trp22 is located on the Met20 loop, Trp30 is near the substrate-binding site and should be a good UVRR spectroscopic indicator of ligand binding, Trp47 and Trp74 are close together but far from the active site and interacts with each other via exciton coupling when the enzyme folds and, finally, Trp133 is located at the final part of the β-sheets near the FG loop (see Fig. 4.7)[146].

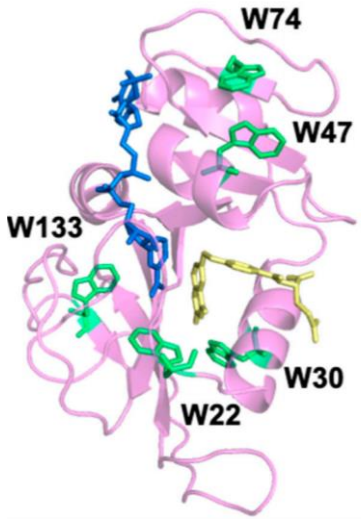


Fig. 4.7: structure of DHFR with Trp residues highlighted in green; the substrate NADPH is highlighted in blue, while DHF in yellow. Figure readapted from[146].

The presence of both TMP and spermine induces an intensity damping of the Trp related vibrational bands. However, the extent of this intensity quenching results more homogeneous in the case of TMP; in fact, once the spectra were normalized to the W1+Y8 band, no remarkable differences between the spectra of DHFR and

DHFR+TMP were detected. This fact may suggest that Trp residues are not directly involved in the interaction with TMP, but they are sensitive to the structural modification induced by the binding. Additionally, band fitting results highlighted that TMP does not modify the secondary structure of the enzyme, which remains in a α-helix-like conformation. Similar conclusions on DHFR+TMP tertiary structure were reported by Couling and colleagues after the analysis of the corresponding 244 nm-UVRR spectra, where only the W7 Fermi doublet peak shows a modification of the peaks relative intensity[293]. This effect underlies only a possible average modification of Trp side chains hydrophobicity, but excludes a direct interaction between DHFR and TMP[293]. Actually, in the case of DHFR+TMP, our UVRR spectra pointed out that Phe band located at 790 cm^{-1} is actually affected by the interaction, increasing its intensity, indicating a modification of Phe microenvironments' hydrophobicity, possibly consequent to a direct interaction between Phe and TMP, as reported in other cases where protein-ligand interaction was probed[276]. Interestingly, Matthews and colleagues demonstrated that TMP binds to the active site of E. Coli DHFR with its pyrimidine ring occupying the inner part of the hydrophobic cleft[294]. Ile-5, Ala-6, Ala-7 and Ile94 form van der Waals contacts with the pyrimidine ring of TMP, while Phe31 and Leu38 establish hydrophobic contacts with the ring, in perfect agreement with our observation[294].

On the other hand, in the case of spermine, we observed that the intensity damping affects mainly peculiar bands such as W16, W7 and W3. By normalizing the spectra to the W1+Y8 band, we observed that spermine induces a strong modification of the W3 band intensity immediately after its addition. In fact, the relative intensity of W1+Y8 and W3 bands passes from 1.9 ± 0.2 in absence of spermine to 0.9± 0.1 when spermine is added. The spectral shape of the UVRR spectra as well as Trp peak relative intensity of DHFR + spermine do not vary within 3 days. Nonetheless, the binding with spermine promotes the increasing of a β-sheets population as suggested by the position of Am-I peak reported in Table 4.1. Then, after 4 days of incubation, we observed a further intensity damping of the whole spectrum and, after 5 days, the UVRR signal of DHFR was strongly reduced due to the formation

of insoluble aggregates precipitating on the bottom of the cuvette. In fact, after 4 days of incubation, TEM images reveal the presence of fibrillar formation.

Thus, while in the case of TMP, Trp bands seem to be not directly involved in the interaction, in the case of spermine, Trp residues or their torsional angles are on average perturbed by the presence of the polyamine, indicating that TMP and spermine could have different binding sites or their binding could induce different DHFR structural conformations.

Electrostatic interactions may be the leading forces for ligand binding to DHFR. Despite the enzyme displays an overall negatively charged surface at neutral pH, it also exhibits pockets characterized by a positive potential, where negatively charged DHF and the adenosine phosphate group of NADPH can bind[295]. In particular, positively charged residues are concentrated in the ligand binding site domain, while the negatively charged side chains dominate the other portion, giving rise to a strong charge anisotropy alongside the enzyme[295]. Differently to TMP or MTX, which bind to these positively charged pockets, the positively charged spermine has to bind to negatively charged side-chains, maybe inducing a strong impairment of the delicate electrostatic balance alongside the enzyme, finally leading to aggregation. A similar effect has been observed when polyamines bind to C-terminus of α-synuclein, neutralizing the negatively charged residues, thus inducing fast aggregation[51,296].

In fact, as reported by dr. De Fronzo in her PhD thesis, preliminary circular dichroism data on DHFR thermal unfolding in presence of different concentration of spermine highlighted that spermine promotes a reduction of the protein melting temperature of approximately 10 °C, accompanied by an increase of enthalpy of 50 kcal/mole, inducing the formation of a less stable structural conformer of DHFR, characterized by an increased free energy. Hence, the transition from a folded DHFR to an unfolded one is favoured by the presence of spermine. This misfolded structure has a kinetic of aggregation of days, depending on the protein-spermine relative concentrations.

Thus, the progressive partial unfolding of DHFR in presence of spermine, monitored by UVRR spectroscopy, suggests two possible scenarios: in the first one, Trp residues

are likely to be involved in the interaction with spermine and, in the second case, the structural modifications induced by the binding with spermine induce and indirect modification of Trp residues' microenvironments. As reported by Ohmae and colleagues, Trp residues and, in particular the interaction between Trp47-Trp74, play an important role in the structural stability of the enzyme[297]. Since the Trp B_b transition at ~220 nm are dominated by the contribution of Trp47-Trp74 exciton coupling as revealed by Grishina and colleagues[298], we suggest that the relevant damping of peculiar UVRR Trp bands intensity in presence of spermine can be ascribed to the impairment of this unique interaction. In particular, W3 peak, which is the most damped band, is sensitive to the torsional angle arrangements of Trp residues which is crucial for the maintenance of these kind of interactions. Interestingly, the destabilization of the native conformers of DHFR at neutral pH by the modification of the hydrophobic cluster around Trp47-Trp74 have been already reported[299].

Thus, UVRR and preliminary data suggest that during the first 3 days of incubation with spermine, DHFR seems stabilized in a high-energy misfolded conformer with a partial enzymatic activity, probably inducing a modification of the overall enzyme electrostatic potential. The loss of the native structure, of its enzymatic activity, and the consequent formation of a β-sheets-rich structure typical of protein fibrils, is a slow process occurring in few days at the concentrations explored in this project.

4.5 Conclusion

Biochemical, spectroscopic and structural studies revealed that *E. coli* DHFR is partially inhibited by the presence of spermine at pH 7.3 and 30°C. UVRR spectroscopy demonstrated to be a valuable tool to be used for the investigation of ligand-protein interactions. Trp residues vibrational peaks are sensitive reporters of the molecular interaction with both ligands, but to a different extent and depending on the ligand exploited, revealing a possible different mechanism of interaction between DHFR and spermine compared to the widely studied DHFR:TMP complex. In fact, while TMP stabilizes the DHFR in a non-evolving inactive state, the presence

of spermine drastically induces firstly the formation of misfolded high-energy conformers of DHFR, characterized by a partially inhibited enzymatic activity and, lately, the formation of fibrillar species. In light of the preliminary studies of dr. De Fronzo and the literature reported here, we speculate that the interaction between DHFR and spermine may impair the Trp47-Trp74 interaction, modifying the distribution of electrostatic charges and the structural disposition of hydrophobic an hydrophilic side chains alongside the enzyme. Further UVRR experiments, possibly using different incoming excitation wavelengths, a higher spectral resolution and employing also another positively charged ligand may help to reach a deeper knowledge of the molecular mechanism at the basis of DHFR-spermine interaction.

Chapter 5

UVRR study on α-synuclein and the effect of Thioflavin T

As reported in Chapter 1, the term "proteinopathy" was coined to indicate a class of protein misfolding-related disorders such as Alzheimer's disease and other dementias, which nowadays have been listed in the top 20 causes of death worldwide by World Health Organization (WHO). Most of the proteins responsible for proteinopathies belong to the class of intrinsically disordered proteins (IDPs). In particular, α-synuclein (aS) is a well-known IDP located in the pre-synaptic area, involved in synaptic vesicles trafficking and release, regulation of enzymes, transportation of neurotransmitters, chaperone-like activity and fatty acid binding[300]. However, due to endogenous and exogeneous causes, aS fibrillation is induced by the formation of oligomeric toxic species and, finally, to the formation of fibrils-rich proteinaceous deposits outside cells. The mechanism of their formation as well as their inhibition remain elusive. The structures of different aS fibrils have been obtained by cryo-EM and solid state NMR (ssNMR), reaching a sufficient structural resolution. X-ray diffraction techniques cannot be used for this kind of purposes, since only small ordered fragments could be investigated. As explained in Chapter 2, cryo-EM and ssNMR limitation could be overcome by a multi-technique approach based on UVRR spectroscopy. This technique offers several advantages compared to other biophysical techniques such as the possibility to work in dilute conditions, to deal with proteins in their native state, to collect spectra of proteins in aqueous

environment and without the use of cryogenic temperatures, being at the same time particularly sensitive to peptide bonds and aromatic residues vibrational modes.

In this Chapter, the structural characterization of wild-type (WT) α-synuclein (aS) and C-terminus-truncated (1-120) aS, both in the monomeric and fibrillar state, have been characterized by a UVRR spectroscopy-based approach in order to elucidate their structural arrangement. Differently to previously reported UVRR protocols in the literature, this UVRR study has been made without using an internal standard or calibrant. Despite this choice has been firstly adopted by Ramachandran and colleagues for the study of tau proteins[301], the use of internal standards is actually widely adopted for the investigation of complex systems such as biopolymers, proteins and nucleic acids[302]. Unfortunately, the high concentration of internal standard required for the assay induces relevant modifications on both fibrillation kinetics and on the final structures of protein fibrils[303]. In the next paragraphs, a brief digression on this topic is included.

Thus, an alternative method of analysis of protein spectral features has been proposed in this Chapter. This approach aims to indirectly shed light on the aS monomers and fibrils structures, roughly estimating how many tyrosine residues remain solvent-protected in both the conformers only from an intensity variation of the tyrosine bands in aS. Similarly to fluorescence spectroscopy, the intensity of the UVRR peaks is not a direct summation of contributions arisen from each resonant aromatic residues, but reflects the average Tyr microenvironments' state. For this reason, the intensity of the Tyr UVRR peaks of aS have been compared to those of insulin, that have the same amount of Tyr residues and its structure is well-characterized in literature. This alternative UVRR-based approach has been tested several times due to the high structural variability of the species investigated. The results, complemented with the literature , indirectly predict the arrangement of WT and 1-120 aS fibrils and monomers.

The investigation of the structural data is crucial for drug design and for the development also of diagnostic strategies *in vivo*. In fact, Lewy bodies, which are considered the pathological hallmark of PD, are found composed not only by aS but also by Aβ fibrils (approximately 40%), indicating that amyloid aggregates *in vivo*

are largely heterogeneous in protein composition[304]. Thus, the need to find ligands able to selectively bind to specific proteins and, in turn, distinguish overlapping pathologies *in vivo* has become crucial. To model the interaction with potential diagnostic fluorescent ligands, we investigated the interaction between aS fibrils and Thioflavin T (ThT). The use of ThT not only is directed to model the mechanism of interaction between amyloid proteins and fluorescent diagnostic dyes, but is important also to understand which are the ThT binding sites in the case of aS. UVRR, IR and nanoIR results suggest us the possible involvement of Tyr39 or of a region spatially close to it when WT and 1-120 aS are folded into fibrils.

5.1 The intrinsically disordered structure of *α*-synuclein

α-synuclein (aS) is a 14 kDa protein formed by 140 amino acids, which is encoded by the SNCA gene located in chromosome 4. aS is mainly concentrated in the presynaptic nerve terminals and nucleus, where it accounts for about 1% of the total content of cytosolic proteins. In fact, the term "synuclein" underlies the localization of the protein in SYNapses and NUCLeus[305].

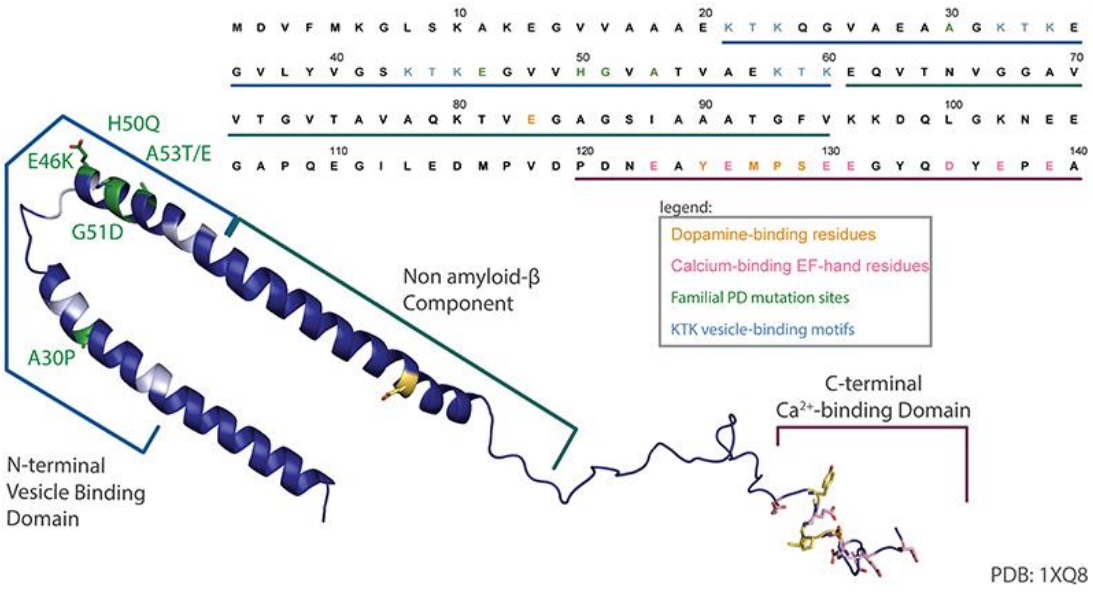


Fig. 5.1: structure of aS when bound to membranes. N-terminus adopts an α-helix-rich, while C-terminus is unordered. Familiar PD mutation sites are reported in green and the complete amino acidic sequence is reported in the upper part, where key residues are coloured as reported in the legend. The figure has been readapted from [306].

The structure of aS can be divided in three regions (Fig. 5.1): the N-terminus portion (1-60 residues), which contains four 11-amino acid imperfect repeats with a conserved KTKEGV motif, the NAC region (61-95 residues) which contain the hydrophobic and most amyloidogenic region of aS characterized also by the presence of three repeating KTKEGV regions and, finally, the highly acid, negatively charged and proline-rich C-terminus region (96-140 residues). Curiously, the NAC domain is believed to be the responsible for aS auto-assembly and all the mutations associated to the early onset of synucleinopathies lies in the preNAC region: A30P, E46K, H50Q, G51D, A53E and A53T and so on[55]. The effect of these mutations on the aS fibrillar structure will be elucidated in the next sections.

N-terminus and NAC regions are characterized by the presence of a membrane-binding domain (residues 68-72), while C-terminus is thought to be the portion accounting for the protein-protein and protein-small molecule interaction, protection of aS from aggregation and so forth. Interestingly, C-terminus remains unstructured and is target of different post-translational modifications such as Ser129 phosphorylation.

aS is an intrinsically disordered protein, since it lacks of a well-established secondary and tertiary structure. This peculiarity lets aS able to adopt multiple conformations, adapting its structure to the environmental conditions and stimuli. Despite unbound aS molecules are mostly present as unordered in the cytosol, aS adopts an α-helical structure where it is bound to intracellular membranes and vesicles, indicating that different conformational sub-ensembles of aS coexist in cells[307–311]. In presence of membranes or lipids, the N-terminus of aS is prone to adopt a helical structure while C-terminus remains soluble, unordered and completely exposed to the solvent[312]. In fact, the truncation of the N-terminus strongly reduces the binding to fatty acids[313]. Differently, when unbound in the cytosolic environment, aS molecules arrange into an unordered structure, which is actually more compact than a typical random coil[314]. This arrangement is stabilized by the establishment of long-range interactions

between C-terminus and N-terminus regions, protecting the NAC core from the solvent, thus disfavouring aS spontaneous aggregation[52,315,316]. The establishment of this long-range interactions can be explained by the electrostatic attraction between the negatively charged C-terminus and the positively charged N-terminus[52,317]. In this configuration, Phe4 and Tyr39 are mainly involved in transient interactions with cytoplasmic components[310].

Hence, the outstanding structural flexibility confers to aS the tendency to populate a highly dynamic ensemble of different conformations depending on environmental conditions[300,307]. The occurrence of post-translational modifications could impact negatively or positively to membrane binding affinity, implying the central role of these mutations regarding the delicate balance between functional and dysfunctional aS[55,318–322]. It is important to mention that the propensity of aS to fibrillate depends on the environmental conditions such as pH variation, increased salt concentration and interaction with proteins, ligands and metal ions[323,324]. Those agents have a relevant impact on aS net hydrophobicity and charge.

5.2 Biological functions of α-synuclein

Despite aS molecules have been largely studied for several decades, their normal function remain elusive. Overexpression of aS as well as the lack of a well-defined structure, let the study of aS biological function even more complicated. The localization of aS at the presynaptic region, the consequent interaction with synaptic proteins, vesicles and membranes highlight that aS molecules have a regulatory synaptic function[300]. Vesicles trafficking and neurotransmitter release, membrane remodelling, promotion of SNARE-complex assembly and many other functions are shown to be regulated by aS molecules[55,300]. Additionally, Fakhree and colleagues demonstrated that aS is able to remodel membrane curvature, modulating their physical properties[307]. Unfortunately, the overexpression of aS impairs vesicles trafficking and could inhibit the activity of the SNARE-mediated vesicles as well as provoking a concentration-dependent folding into toxic oligomers and aggregates. Additionally, the presence of Lewy bodies formed by aS fibrils strongly correlates

with a remarkable loss of dopaminergic neurons in the substantia nigra, thus indicating that aS dysfunction impairs the dopamine metabolism[325,326]. Interestingly, aS not only binds to cellular or lipid membranes, but also interacts with many proteins acting as a molecular chaperone[327].

Thus, the presynaptic localization of aS, its interaction with proteins and lipids as well as with cellular membranes, its chaperoning activity suggests that aS is involved in several mechanisms and processes fundamental for the long term operation and health condition of neurons and of synapses, especially during intense neuronal activity[300].

However, despite the central and key role of aS in the synaptic space and its fundamental role in maintaining the correct functioning of neurons and synapses, a clear understanding of the reasons underlying the onset of synucleopathies is still lacking.

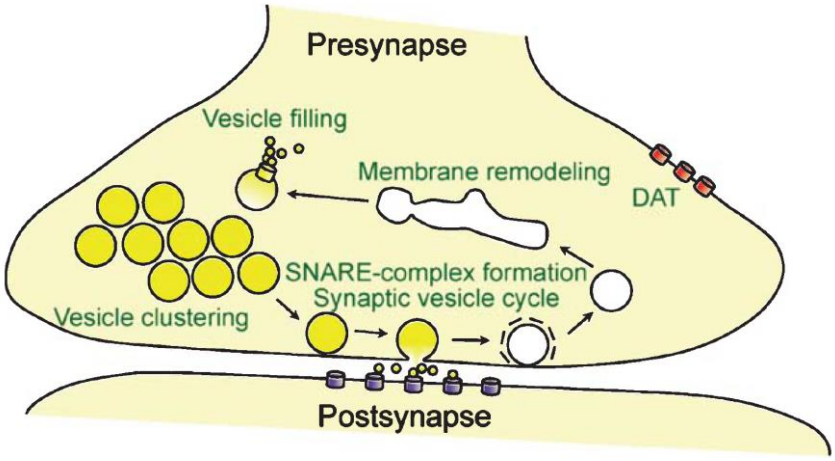


Fig. 5.2: Function of aS at the synapse includes membrane remodelling, modulation of neurotransmitters transport, clustering and trafficking of synaptic vesicles, promoting SNARE-complex assembly and modulating the release cycle of neurotransmitters at synapse. The figure is readapted from[300].

5.3 Pathways and molecular mechanism leading a-synuclein to aggregation

Despite the mechanism with which aS molecules and or its aggregated forms induce damage in neurons has not been identified yet, the kinetics of aS fibrillation has been elucidated. aS aggregation is influenced both thermodynamically and kinetically by

intrinsic and extrinsic factors such as point-mutations, environmental pH, cellular crowding, the presence of metal ions, peculiar molecules and so forth. Similarly to other proteins, it is believed that aS forms fibrils through a nucleation-dependent manner, by converting portion or its whole structure from a previously mildly unordered polypeptide into a well-defined β-sheets-rich arrangement, gradually passing through different phases (i.e. oligomers and protofibrils). Once fibrillar nuclei are formed, fibrils grow by a "dock and lock" mechanism, thus fibrils act as seeding portion for monomers inducing them to aggregate. This mechanism is called also "seeding" and it is quite common among proteins, in general. aS seeding mechanism has also been demonstrated *in vivo* after injection in rat brain[328].

In the next sections, the mechanism of formation of oligomers and fibrillar aggregates and their impact on membranes as well as the effect of post-translational modification on fibrils structure and on the onset of the diseases is elucidated.

5.4 Toxicity of α-synuclein oligomeric species and membrane disruption

Oligomers are the early intermediates of the aggregation pathway of aS and have been retained the most toxic species formed throughout the entire process. In fact, experimental studies highlighted their ability in binding to cellular and lipid membranes, disrupting and penetrating them in a pore-like manner[55,68,78]. The disruption of membranes leads to an increased amount of basal intracellular calcium and of reactive oxidative species, which impair the cellular functionalities[68,78,329]. Actually, the ability to bind and permeabilize membranes surely depends on the structural features of oligomers[68]. Being transient, unstable, intermediate species, the structural investigation of aS oligomers remains difficult and misleading, depending on many aspects such as their kinetic and thermodynamical instability, their higher structural heterogeneity in size and structure. The preparation of stable oligomers is very challenging due also to the low amount of materials available for the assays and for the possible structural modification induced by the preparation protocols.

Pioneering Trp-fluorescence spectroscopy measurements were performed on site-directed mutated aS oligomers (F4W, Y39W, A69W, A90W, A124W and A140W) in order to determine the structure of aS oligomers and to elucidate the mechanism of interaction with membranes[78]. These experiments indirectly highlighted that, in the absence of membranes, the 4W, 39W 69W and 90W are less solvent exposed compared to the monomeric forms, while 124W and 140W remains solvent exposed[78]. Hence, the N-terminus of the oligomers becomes more solvent-protected, while the C-terminus remains soluble and in contact with the solvent as lately confirmed also by Fusco and colleagues[68]. Acrylamide quenching experiment individuated in the N-terminus and in the first portion of NAC, the region of lipid membrane binding as well as the region forming the oligomers' kernel[78]. Despite C-terminus remains unstructured and solvent-exposed, the N-terminus and NAC regions remain less solvent exposed, partly forming the oligomeric hydrophobic core and partly folding into an α-helical structure, in order to promote the binding of the core to the negatively charged cellular and lipid membranes[252].

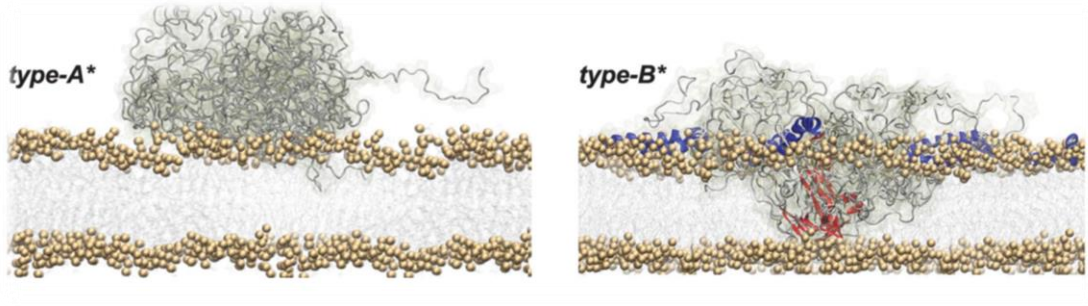


Fig. 5.3: (left) type-A* oligomers are mainly disordered and bound to the membrane while (right) type-B* oligomers are characterized by a structured α-helix N-terminal region with a rigid region rich in β-sheets structure. This structural modification confers the ability of oligomers to disrupt membranes. Figure readapted from [68].

Lately, by studying the *in vitro* interaction of two different stabilized oligomers (A* and B*) with synaptic vesicles model, and *in vivo* with human neuroblastoma SH-SY5Y cells and on rat primary cortical neurons, Fusco and colleagues discovered that due to the structural heterogeneity of oligomers, their effect on lipid and cellular membrane could be different (Fig. 5.3)[68]. Despite both oligomers showed similar

morphology and size, the interaction between type-A* oligomers and lipid membranes, cells and cortical neurons not only results in mild release of calcein molecules and reactive oxidative species (ROS) through the membrane, but also in a reduced penetration inside the lipid bilayer forming membranes. Differently, type-B* oligomers are able to penetrate in the hydrophobic core of the bilayer, promoting a release of calcein molecules and ROS 10 times more, with the consequent disruption of the membranes and the impairment of the cellular or vesicular functions[68].

The reason at the basis of this different behaviour resides in the structural properties of these oligomers: while type-B* oligomers show a highly dynamical and more accessible N-terminus composed by a highly lipophilic portion, with a lipid binding region comprises in the first 25 residues of the N-terminus, type-A* oligomers lipid binding region seems to lie in a larger portion (1-97). This structural property allows type-B* oligomers to penetrate inside lipid bilayers, while type-A* ones only bind to membranes' surface[68].

5.5 Fibrillar structure of α-synuclein

Similarly to other common amyloid fibrils, aS forms fibrils through a nucleation-dependent polymerization, thus by converting part or its whole structure into a β-sheets-rich secondary structure[55]. X ray diffraction, ssNMR and Cryo-EM revealed that aS fibrils are composed of protofilaments adopting a β-serpentine arrangement with a compact Greek-key topology rich in parallel and in-register β-sheets(Fig. 5.4)[37,56,330]. Alongside the fibrils axis, in register hydrogen bonds are present orthogonally to the hydrogen bond of the Greek-key. The innermost β-sheet of the hydrophobic core is formed by residues (71-82), while the fibril core is composed by residues 46-54 and 63-96. Residues 55-62 form a disordered structure while the rest of the polypeptide is tightly arranged in order to protect the hydrophobic core and to facilitate a backbone-backbone interaction. In particular, A69-G93 bridge the distal loops of the Greek key,while G47-A78 come close forming a hydrophobic steric zipper, establishing an intermolecular salt bridge between E46 and K80. The compact

arrangement of the Greek key topology of aS fibril core has been attributed to the presence of a large fraction of small and flexible amino acids such as glycine, alanine and serine, which account for more than 40% of the total amino acids forming the fibril core[56,330]. The characterization of aS fibrils structure as well as of their exposed regions throughout the fibrils axis is crucial for drug design.

Interestingly, both *in vitro* and *in vivo* aS fibrils have a diameter of 5-10 nm, but actually some differences in morphology and structural architecture emerged[56]. Those aspects are discussed in detail in the next paragraphs.

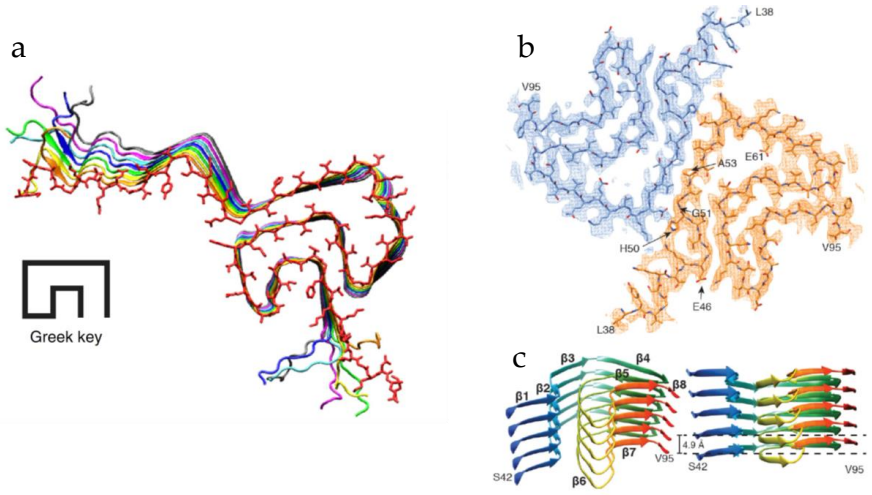


Fig. 5.4: (a) Residues 25-105 of 8 monomers, showing the β-sheet alignment alongside the perpendicular axis of the fibril and the Greek-key topology of the core; (b) cross-section of a fibril core along its axis (from L38 to V95), showing the arrangement of the protofilaments. (c) distribution of the b-strand of a single protofilament and its perpendicular view. Figure readapted from [56,331].

5.5.1 Fibrils polymorphism and the impact of mutations

Synucleinopathies, including Parkinson's disease (PD), dementia with Lewy bodies and multiple system atrophy (MSA), are all characterized by the presence of amyloid deposits in brain. However, fibrils extracted from those deposits show polymorphic structures which, in turn, have different seeding properties and cytotoxicity[318–322,330,332]. Interestingly, *in vitro* aS fibrils are also composed by an ensemble of polymorphic structures[70,319–322,332]. Additionally, both *in vivo* aS fibrils extracted form

PD and MSA patients and *in vitro* aS fibrils show an heterogeneous width (5-10 nm) and a different assembly of protofilaments forming fibrils[56,69].

Recently, both ssNMR and Cryo-EM revealed two major polymorphs of full-length aS fibrils with a rod-like and a twister-like structures, depending on the protofilaments interfaces arrangement[70,331]. They show a highly conserved fibrils core structure, characterized by a bent β-arch motif, assembled around different steric zipper interfaces. Despite the two polymorphs share the same fibril kernel structure, the twister polymorph has a well-ordered bent β-arch motif with 42 ordered amino acids (K43-E83), while rod-like structure forms a Greek-key-like structure, extending the ordered amino acids portion to 60 residues (L38-K97). Additionally, the protofilament interfaces between the two polymorphs differs: in the rod structure, residues 50-57 (preNAC) form the interface between the two protofilaments, whereas in the twister polymorphs the contact region between the protofilaments lies in the NAC region, i.e. residues 66-78 [320].

The polymorphic nature of aS fibrils, collected both *in vivo* and *in vitro*, may have different seeding efficiency and cytotoxic profile[69,70]. To this purpose, it is crucial to characterize each polymorph in order to understand its biological function and its role in the aggregation phenomenon. Nonetheless, aS fibrils polymorphism has not only been observed in wild type aS fibrils, but also in presence of point-mutated aS fibrils, the occurrence of which has been correlated to the autosomal-dominant PD inheritance[70,319,320]. In fact, the study of hereditary mutations could elucidate the link between protein structure and the severity and/or the early onset of the disease. Several aS missense mutations such as A30P, E46K, H50Q, G51D, A53V and A53T are characterized by different cytotoxic properties, kinetics of aggregation and structural arrangement[318–320,322]. Interestingly, all of them are located in the N-terminus and preNAC regions, which are crucial for lipid membrane binding and for the maintenance of the long-range interaction between N- and C-termini. Additionally, all these mutations somehow disfavour the rod-like arrangement of the fibril core, sometimes inducing the occurrence of other polymorphs[70,319,320]. To name a few, while A30P and A53T mutations do not alter significantly the structure of the aS fibrils[318], the E46K mutant determine the disruption of a key salt bridge

between E46 and K80, disturbing the fibril core structure[319]. Recently, Boyer and colleagues discovered the presence of different polymorphs dealing with H50Q aS fibrils, termed as narrow and wide fibrils due to their different widths[320]. The mutation from a histidine to an uncharged and polar glutamine could disrupt the rod polymorph protofilament interface, leading to the occurrence of polymorphs with a modified hydrogen bond network. Interestingly, H50Q mutation does not produce twister-like nor rod-like polymorphs, which could explain the enhanced aS aggregation and cytotoxicity both *in vitro* and *in vivo*.

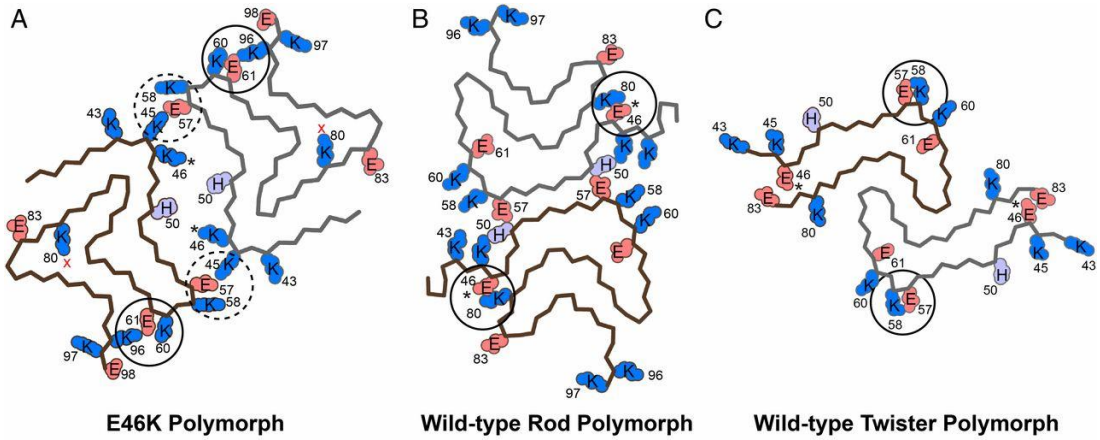


Fig. 5.5: Examples of aS polymorphism. (A) Mutation E46K induces the formation of a different aS structure; (B) and (C) are two aS polymorphs having a rod- or a twister-like structure. Figure readapted from [319].

5.5.2 Post-translational modification

Phosphorylation, tyrosine nitration, ubiquitination, glycosylation, N-terminal acetylation and truncation are some post-translational modifications (PTM) reported for aS. Each PTM has a peculiar effect on the structure and on the aggregation kinetic of aS, and each effect depends significantly also on the site where the modification occurs[55,300]. To name a few, phosphorylation at Ser129 and Ser87 are highly conserved and casein kinases CK1 and CK2 together with G-protein are believe to be the responsible for the occurrence of this PTM[333,334]. Phosphorylated aS at Ser129 comprises approximately 90% of the aS insoluble fibrils found in LBs and are highly prone to aggregate[334]. Differently, phosphorylation at Ser87 increases aS flexibility

and blocks aS *in vitro* aggregation[335]. Additionally, also Tyr125 has been found phosphorylated in brain, promoting the attenuation of aS monomers-to-oligomers conversion[336].

As listed above, tyrosine nitration takes place often both *in vitro* and in LBs from the brains of PD patients[55,337,338]. In particular, nitration of Tyr39 is the most observed event since this residue is more accessible to nitrating agents[337]. As a consequence, nitration at Tyr-39 reduces aS ability to bind to lipid vesicles[55].

Another possible PTM is represented by ubiquitination, where ubiquitin enzymatically attaches to lysine residues of aS (i.e. Lys6, Lys10, Lys12), targeting proteins for proteolytic degradation[55,339]. This process occurs after aS aggregation and ubiquitinated aS fibrils are commonly found in Lewy bodies of PD patients and in the cytoplasmatic inclusions of MSA patients. As explained before, the effect of ubiquitination on the structure and kinetics of aS aggregation depends on the specific site targeted: ubiquitinated Lys6 induces a slower aS aggregation[55], while multiple lysine ubiquitination induces the formation of cytotoxic aggregates both *in vitro* and *in vivo*[340]. However, one of the most studied PTMs is the truncation and, in particular, the C-terminus truncation which is extensively discussed in the next paragraph.

5.5.3 Truncation of α-synuclein

Truncated aS fibrils has been found in LBD, PD or MSA brain lysate in complex with full-length aS fibrils[331,341]. Several proteases which are particularly prone to partially degrade aS by removing part of its C-terminus have been identified such as 20S proteasome, calpain I, caspase I, AEP, MMP1 and 3, cathepsin D and B[341]. Most of them show an increased activity when the level of oxidative stress is high.

aS (1-119) and aS (1-122) represent the most common forms of truncated aS and, generally, approximately 15-30% of the insoluble aS fibrils extracted from PD and LBD patients brain is truncated[341]. In particular, if truncated aS molecules are present in the deposits, they are located in the inner part of the aggregate, suggesting that truncated forms of aS could play a central role or catalyse the aggregation process[341]. Sorrentino and Giasson discovered that the C-terminus truncation is highly

prevalent in particular regions of brain compared to others, suggesting that truncation could represent a useful parameter for evaluating toxicity and may have implications for the initiation and progression of the disease[341].

Differently to the N-terminus truncation, the partial or total removal of C-terminus has a strong impact on the kinetics of aggregation, making the process faster[341,342]. In fact, the loss of the electrostatic interaction with the positively charged N-terminus induces the solvent exposure of the highly amyloidogenic NAC region, no more shielded by the C-terminus. A similar enhanced kinetics of aggregation have been detected when the negative charges of C-terminus are neutralized by lowering the pH, by the binding with positively charged ligands or by mutations[341].

Recently, aS (1-121) fibrils structure and morphology have been elucidated by Cryo-EM experiments[301]. Similarly to full-length aS fibrils, the width of aS (1-121) ones varies from 5 to 10 nm but, differently from full-length aS, they tend to be longitudinally shorter. Depending on the extent of truncation, fibril widths could be larger or smaller. Morphologically, C-truncated fibrils appear to be ribbon-like, with an increased twisting of protofilaments within the fibrils compared to full-length aS[341]. This peculiar twisting confers a higher helical periodicity to full-length aS, which in turn become extreme in presence of an extensive C-truncation, making the structure more tightly packed due to the reduce steric hindrance between protofilaments[342]. However, not only the *in vivo* truncated form of aS fibrils shows an outstanding ability in self-assembly, but also the recombinant C-terminus truncated aS shows a similar aggressive aggregation kinetics. Guerrero-Ferreira and colleagues studied the structure of aS(1-121) fibrils, discovering that each protofilament forming fibrillar structure comprises eight in-register parallel β-sheets[331]. Two of them are disposed around the hydrophobic core composed mainly by alanine and valine residues, which are surrounded by two hydrophilic regions (Q79, T81 and T72, T75, T54, T59 and E61) both located within the core of the structure. Unfortunately, the entire structure of aS(1-121) fibrils is not reported, since Cryo-EM failed to reconstruct the first 37 N-terminal residues and the last 20 C-terminus residues due to their likely unordered structure, in agreement with hydrogen/deuterium exchange and ssNMR[56,331]. The fibril core is formed by β-strands that interact with

each other in a half-stacked zipper topology. The interface contact (residues G51-A56) between the two protofilaments is stabilized by an intermolecular interaction between A53 and V55 and by a salt bridge between E57 and H50. The outer surface of the fibrils is mostly hydrophilic with the exceptions of L38, V40, V82, A85, A90, F94, V95. Two highly hydrophobic regions are formed by L38-V40 and F94-V95[331]. Thus, both the full-length and the truncated (1-121) aS show a similar secondary structure arrangement. One of the main difference between the truncated and the full-length aS concerns the orientation of residue A53: in the truncated form, A53 lies in the interaction portion between the two protofilaments, whereas in the full-length seems to point towards the hydrophobic core of the fibril. Finally, cryo-EM experiments on aS (1-121) elucidates the central role of the preNAC in the interaction of the protofilaments, inducing the NAC core to be placed at the centre of a single protofilament[331].

5.5.4 Role of Tyr residues and Tyr39

As reported in the previous sections, aS is characterized by the presence of 4 tyrosine, three located in the C-terminus (Tyr125, Tyr133 and Tyr136), and another one in the N-terminus at position 39. In the monomeric form, aS adopts a non-rigid ensemble of different and flexible conformations, where long-range interactions have been shown to play a key role in maintaining a structural arrangement which disfavour the auto-assembly. Site-directed mutagenesis shows the different impact of each tyrosine residue both on the aS function and on its aggregation kinetic. As an example, the replacement of the aromatic tyrosine with aliphatic residues induces massive changes in the stabilization of aS conformers and has a strong impact on the fibrillation kinetics[343]. To name a few, cysteine substitution of Tyr39 elucidates that this portion plays a key role in the dimerization process, since this process as well as the oligomerization are enhanced. The substitution of Tyr39 and Tyr144 by alanine resulted in the inhibition of the aggregation, suggesting that fibrillation could be induced by the disruption of an aromatic cluster between C-terminus and N-terminus.

In order to understand which parameter affects considerably the fibrillation pathway, different variables have been evaluated: aromaticity, hydrophobicity, residue size and many others. Recent studies suggest that Tyr39 might play a critical role for both the normal function and the pathological dysfunction of aS, participating in a π-π self-stacking interaction[343,344]. To identify which are the main reasons underlying the fibrillation phenomenon, Tyr39 has been substituted with Phe, Leu and Ala residues, which differs in term of aromaticity, size and hydrophobicity. The results show that aromaticity and hydrophobicity at position 39 resulted in a relevant inhibition of the aS fibrillation, impacting on the inter- and intra- molecular contacts within the protein[343].

5.6 The interaction with small molecules: the case of dyes

As explained in Chapter 1, the *in vitro* and *in vivo* post-mortem detection of amyloid fibrils is usually made by staining proteinaceous deposits with specific dyes such as Congo Red or Thioflavin T. However, while they are highly sensitive to the presence of protein fibrils, these small dyes bind non-selectively to different amyloids, i.e. they lack in specificity. This aspect becomes largely important dealing with *in vivo* proteinaceous deposits and, more importantly, in the design of specific drug potentially usable for an *in vivo* detection of specific protein fibrils. In fact, proteinaceous deposits formed in brains are composed by the deposition of multiple proteins and molecules and the onset of a specific neurodegenerative disease is related to the major representative protein found in those deposits[35]. For this reason, aS fibrillation has been considered the main cause of Parkinson's disease, LB dementia and multisystemic atrophy. Anyway, aS, Aβ and tau proteins fibrils can coexist in the same deposits, with or without clinical manifestations associated with any individual pathology[304,345]. As an example, Lewy bodies are rich in aS fibrils and their presence is considered the clinical hallmark of Parkinson's disease; however, their presence have been reported also in ~40% of patients with Alzheimer's disease[304]. For this reason, the need to develop selective biomarkers aiming to

identify overlapping pathologies *in vivo* as well as to early detect a specific protein fibrils population has become crucial. However, the lack of detailed structural data about tau, aS and Aβ fibrils and how they are distributed within the deposits depending on the pathology diagnosed, provokes controversies in ligand design able to interact selectively and with high affinities to different amyloid proteins[304,345]. Away to develop biomarkers able to interact selectively with specific fibrillar proteins, the way high water-soluble benzothiazole ThT dye interacts with aS is discussed in this thesis, in order to help the design of novel derivatives that can be used in medical and clinical *in vivo* screening.

Up to now, it is still not clear which are the preferable binding sites where dye molecules interact with fibrils. Surely, these non-specific dyes recognise a shared feature of amyloid fibrils, which could be the cross-β structure or the arrangement of pockets and grooves alongside fibrils[124]. Molecular dynamics simulation highlights linear dyes bind parallel to the long axis of the fibril surface in grooves formed by side-chain ladders[124]. However, ThT does not bind uniformly to the exposed surface of proteins and has a higher propensity to bind to highly charged and aromatic moieties (in particular, Tyr and Phe residues), and in regions sterically allowed due to ThT dimension[124,346].

5.7 Experimental methods

5.7.1 Sample preparation

Sample preparation has been conducted by the prof. L. Bubacco group at the Department of Biology of the University of Padua (Italy), using the following protocol already reported in literature[199,347].

WT aS and 1-120 aS were cloned into the pET-28a plasmid and were expressed into E. Coli strain BL21(DE3). Cells were grown at 37°C in LB broth medium to an optical density OD_{600} = 0,3- 0,4 and induced with 0.1 mM isopropyl-b-thiogalactopyranoside (IPTG) for 5 h. Cells were collected by centrifugation and recombinant proteins were recovered from the periplasm by osmotic shock using an osmotic shock buffer. The

periplasmic homogenate was boiled for 15 min and the soluble fraction containing aS was subjected to a two-step (35% and 55%) ammonium sulphate precipitation. The pellet was resuspended, extensively dialyzed against 20 mM Tris-HCl pH 8.0, loaded into a 6 ml Resource Q column (Amersham Biosciences) and eluted with a 0 - 500 mM NaCl gradient. After dialysis against water, aS was lyophilised and resuspended in PBS buffer (pH 7.3) and stored at -80°C prior the measurements. The identities of all the proteins were verified by mass spectrometry.

In order to observe ThT interaction with aS fibrils, we prepared a fresh stock solution of 2mM and we added to the fibrils solution, reaching a (aS:ThT) ratio of (3,5:1), mimicking a typical fluorescence assay. After an incubation of 30 minutes, storing the solutions in a dark place, the solutions were prepared for the measurements as indicated in the following sections.

5.7.2 UV Resonance Raman spectroscopy

UV Resonance Raman measurements were carried out at the BL 10.2 – IUVS beamline at Elettra Sincrotrone Trieste using the experimental setup reported elsewhere[134]. Samples spectra were measured using two different wavelengths, namely 228 and 244 nm. Synchrotron radiation was used to get the 228 nm source, whereas the measurements taken at 244 nm were carried out using an Ar+ laser (Lexel Laser Inc., USA). Proteins spectra were acquired at room temperature. The diffused radiation was collected in backscattering geometry and energetically dispersed using a spectrometer with 1800 g/mm (Trivista, Princeton Instrument). The final spectral bandwidth is 20 cm^{-1} at 228 nm and is 8 cm^{-1} at 244 nm. Samples excited at 228 nm were irradiated for 4 hours, accumulating one spectrum every 15 minutes with an incoming radiation of 9-13 μW. Differently, with 244 nm of excitation wavelength, samples were irradiated for 10 seconds at 140 μW. The final spectrum was the average of 40 spectra. Spectra of phosphate saline buffer and Thioflavin T were carefully removed from the measured spectra of proteins. Samples were horizontally oscillated during the measurement to avoid photodamage. Sample spectra collected at 244 nm were measured in specific polarization conditions, namely with the polarization parallel and orthogonal to the direction of the electric

field. Polarized spectra were measured using a $\lambda/2$ foil spatially located before the sample, and a polarizer at the entrance of the spectrometer. Finally, the isotropic spectra were calculated by the following relation:

$$I_{iso} = I_{VV} - \frac{4}{3} I_{HV}$$

The spectral region of interest was fitted by using Gaussian curves, which in our case represent a good choice to avoid high correlation between the fit parameters. Considering the same number of curves in order to reproduce the experimental data, the use of Gaussian curves instead of Voigt ones does not induce significant variations of χ^2 but reduces the number of parameters for each peak (3 for Gaussian and 4 for Voigt curves). The high correlation of those peaks is highlighted by the analysis of the errors deriving from fitting the experimental data by using only Voigt curves. The use of Gaussian curves reduces the error of the parameters from 50% (in the case of Voigt) to 10%, which is close to the S/N ratio of the spectra collected, without altering the final result in term of peaks relative intensity. The Gaussian curves were initialized to the value corresponding to the maximum of the peaks and kept fixed for a first run. Then, the parameters were made free and the fit runs until its convergence.

5.7.3 Fourier Transform Infrared Spectroscopy (FTIR)

Infrared spectra of the samples were collected at the infrared beamline SISSI (Synchrotron Infrared Source for Spectroscopic and Imaging), Elettra Sincrotrone Trieste (Trieste, Italy)[140]. Attenuated-total-reflectance (ATR) spectra of the samples were collected using MIRacle™ Single Reflection ATR box (PIKE Technologies) with Germanium (Ge) internal reflective element (IRE). Spectra were acquired using the Vertex 70 interferometer (Bruker Optik GmbH) equipped with a MIR DLaTGS detector and a KBr beam-splitter. 5μL drop of each sample was deposited onto the IRE and the measurements were performed during drop dehydration at ambient temperature, under a gentle nitrogen flux to prevent water vapor spectral interference. To the same purpose, interferometric compartment and ATR accessory were also purged with nitrogen. For each spectrum, 25 x 128 scans-interferogram

were collected with a scanner velocity of 7 kHz and 4 cm^{-1} of resolution. The spectral region explored spanned a range within 5000 to 500 cm^{-1}. A 128-scan interferogram of the background was collected under identical external condition.

5.7.4 nanoFTIR

A total of 15μL of sample was deposited onto polished silicon substrate and it was let dehydrated. In order to remove the sample excess, the sample was washed with deionized water and let dehydrated another time. Infrared nano-spectroscopy images and spectra were taken using neaSNOM microscope (NeaSpec Gmbh, Germany). Measurements were performed with 40 THz laser excitation frequency (1333 cm^{-1}) and with 860 μW average power, by detecting the backscattered radiation interferometrically during the sample surface scanning via an AFM tip.

5.7.5 TEM

15 μl aliquots of the WT and 1-120 aS fibrils were absorbed onto a carbon-coated copper grid and a 0.05% uranyl acetate solution was used for negative staining. TEM pictures were taken on a Tecnai G2 12 Twin instrument (FEI Company, Hillsboro, OR, USA)[348].

5.8 Results

5.8.1 UVRR and FTIR probes WT aS and (1-120) aS fibrils secondary and tertiary structures.

228 nm-UVRR spectra of full-length WT aS and 1-120 aS fibrils are reported in Fig. 5.6 (a) and (b), while their 244 nm-UVRR spectra are presented in Fig. 5.6 (c) and (d), collected by using VV and HV polarizations. In order to extract only the vibrational contribution from the spectra, eliminating the rotational contributions, the isotropic

spectra of aS fibrils at 244 nm were finally calculated using the procedure reported in the Experimental Methods (see Fig. 5.6 (c) and (d)). All the spectra are cut in the region 980-1780 cm^{-1}, where the main features of tyrosine residues (indicated as Y*), phenylalanine residues (F*) and amide bands (indicated as A*) are located. Additionally, the $C_{\alpha}H$ and the CH bands are reported at 1390 cm^{-1} and 1450 cm^{-1}, respectively as clearly reported in Chapter 1 and 3. As a first remark, the variation in the relative intensity of the Raman peaks depends strongly on the excitation wavelength. In fact, as clearly visible in Fig. 5.6 (a) and (b) from the different relative intensity between amide and tyrosine bands (especially between AI and Y8), the spectra at 228 nm are mostly dominated by the signal arisen from tyrosine residues. Since the increment of the molar absorptivity of aromatic residues reduces by approaching the region 240-250 nm, protein spectra collected at 244 nm are characterized by a more intense amide bands.

Apart from the resonant effects, a common feature of all the spectra are the presence of broad and intense peak located at ~1660-1670 cm^{-1}, assigned to amide I band (AI), mainly addressed to stretching vibration of the C=O of the amide linkage in the polypeptide backbone. AI band of both fibrils probed with 228 nm of excitation wavelength at two different spectral resolutions (with 1800 g/mm and 3600 g/mm (not shown), respectively) revealed that the peak is centred in the region within 1664-1674 cm^{-1}, with a full width half maximum (FWHM) of approximately 40 cm^{-1}, suggesting that AI consists of a distribution of secondary structures (see Table 5.1). In order to appreciate these contributions, a band fitting of the 244 nm-UVRR spectrum in VV polarization were performed due to the higher resolution achieved and to the decreasing of tyrosine residues peak cross section. The analysis revealed

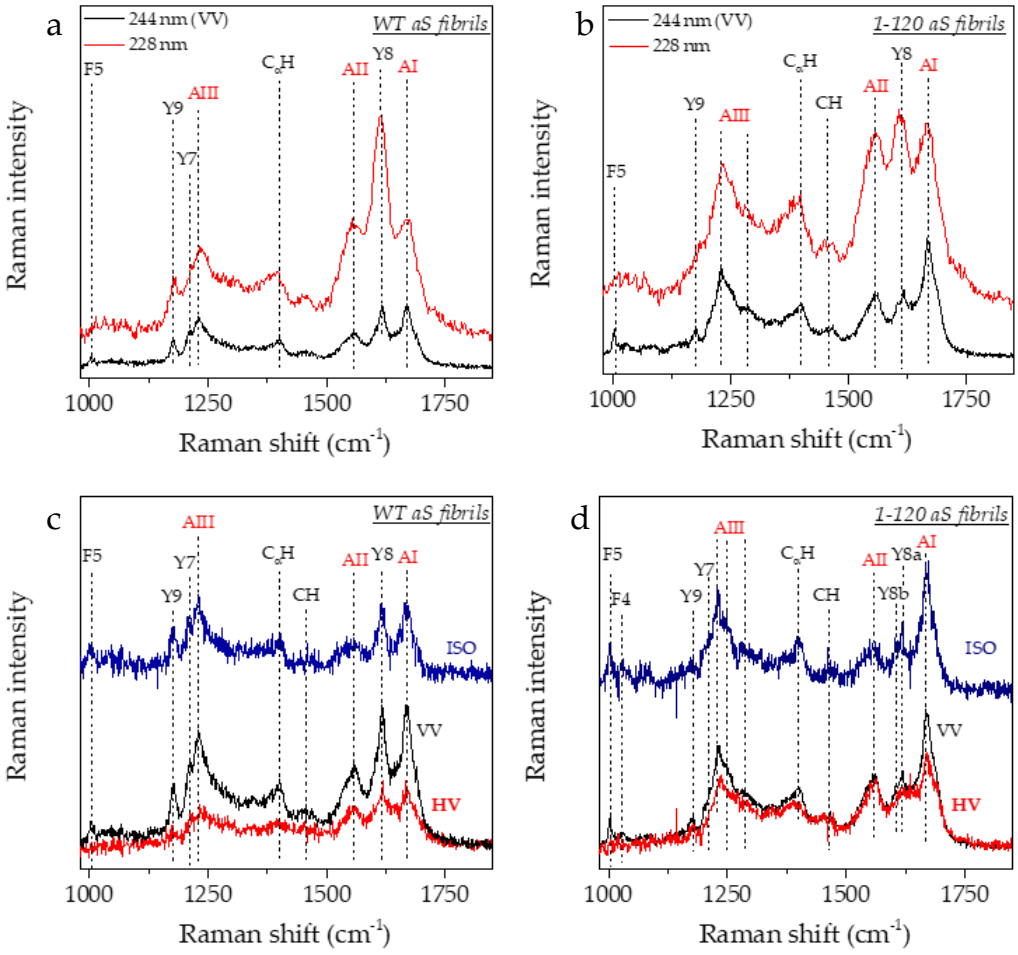


Fig. 5.6: (a) UVRR spectra of wild-type aS fibrils collected at 228 and 244 nm (VV polarization); (b) UVRR spectra of (1-120) aS fibrils collected at 228 and 244 nm; (c) UVRR spectra of wild-type aS fibrils collected at 244 nm in VV (depicted in black), in HV (in red) polarizations. The calculated ISO spectrum is depicted in blue; (d) UVRR spectra of aS (1-120) fibrils collected at 244 nm in VV (depicted in black), in HV (in red) polarizations. The calculated ISO spectrum is depicted in blue. Dashed lines indicate the corresponding peak.

that both aS WT and aS (1-120) fibrils AI bands are characterized by 3 sub-components: one shoulder near ~1640-1644 cm^{-1}, a main peak at ~1668 cm^{-1} another shoulder near ~1677-1687 cm^{-1} in agreement with a previous results provided by Maiti and colleagues (Table 5.1)[129]. As reported in literature, the band located near 1650 cm^{-1} is related to the presence of α-helix, while the component located at 1668-1674 cm^{-1} is a marker of β-sheets structure. Additionally, the higher band located above 1677 cm^{-1} is addressed to unordered structures (polyproline-like structure,

PPII)[129,221]. Thus, from band fitting of AI of the 228 nm and 244 nm UVRR spectra revealed that the secondary structure of aS WT and aS (1-120) fibrils is mostly composed by β-sheets. The population of each secondary structure population has been quantified and corresponds to the area under the fitting curve respect the total area of AI. The results are reported in Table 5.2, where it is evident that the PPII contribution is higher in the truncated form of aS.

Other reporters of protein secondary structures are AII and AIII; AII is sensitive to the out-of-phase combination of NH bending and CH stretching vibrations of the groups in the amide linkage, while AIII is sensitive to a complex vibration involving C-N stretching and NH bending. Nonetheless, in agreement with the results obtained for AI, band fitting of 228 and 244 nm- UVRR spectra reveals that AII of both fibrils is comprised in the region (1556-1562) cm^{-1}, confirming that the structure is mostly β-sheets/unordered (Table 5.1). Accordingly, AIII is formed by different main peaks reported in Table 5.1 and Fig. 5.7, confirming that both fibrils secondary structure are mainly formed by β-sheets and unordered structure (PPII). In fact, AIII of both fibrils is a multi-composed band formed by an intense peak located at 1230 cm^{-1}, which is addressed to the presence of β-sheets, whereas especially the aS 1-120 fibrils are characterized by the presence of other two shoulders located at 1249 cm^{-1} (β-sheets) and 1283 cm^{-1} (β-turns)[129,274,349].

Table 5.1:Band fitting of 228 nm and 244 nm UVRR spectra obtained with a different spectral resolution. The position of Amide I, II and III are reported for both full-length WT aS and aS (1-120) fibrils.

Amide I	*WT aS fibrils*	*aS (1-120) fibrils*
228 nm (1800 g/mm)	1668.7 ± 0.9	1664.3± 0.9
228 nm (3600 g/mm)	1674 ± 3	1674 ± 1
244 nm (3600 g/mm)	1650 ± 1; 1667.1±0.1; 1677±2	1644 ± 1; 1667.8±0.2; 1685.4±0.4
Amide II		
228 nm (1800 g/mm)	1562 ± 2	1555.8 ± 0.9
228 nm (3600 g/mm)	1554 ± 2	1566 ± 1
244 nm (3600 g/mm)	1555.9 ± 0.7	1538 ± 1; 1562 ± 0.5
Amide III		
228 nm (1800 g/mm)	1233 ± 1; 1314 ± 6	1232.7 ± 0.7; 1272 ± 4
228 nm (3600 g/mm)	1233 ± 1; 1261 ± 5; 1313 ± 5	1240 ± 7; 1323 ± 2
244 nm (3600 g/mm)	1230.1 ± 0.5; 1249 ± 2; 1283 ± 2; 1340± 2	1229.8 ± 0.4; 1244 ± 2; 1283 ± 2; 1340± 2

Table 5.2: Content of each secondary structure population characterizing WT full-length and C-truncated fibrils estimated by the UVRR amide I (AI) band, considering the total AI area as 100%.

Amide I	*WT aS fibrils*	*aS (1-120) fibrils*
α-helix	73.7 %	67 %
β-sheets	14.3 %	28 %
Unordered (PPII)	12 %	10 %

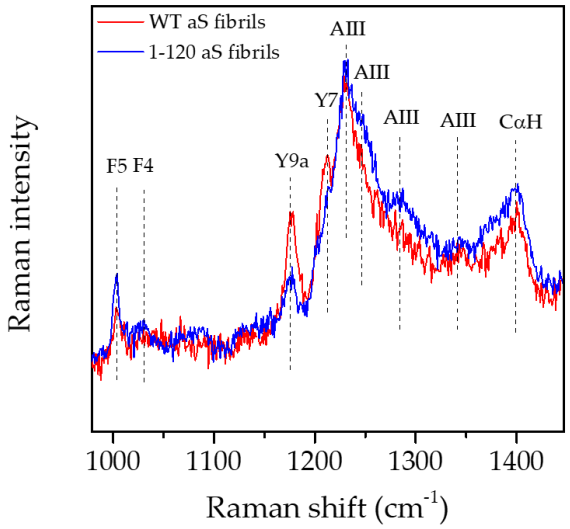


Fig. 5.7: 244 nm-UVRR spectra of WT full-length fibrils and aS 1-120 fibrils in the region (980-1450) cm^{-1}. The spectra highlight that AIII is more structured in the case of C-truncated fibrils, underlying a more prominent presence of α-helix and unordered structures.

The secondary structure elucidated trough UVRR spectroscopy and the analysis of AI, AI and AIII is in agreement with FTIR results reported in Table 5.3. In fact, the absorbance and the second derivative spectra of WT aS an 1-120 aS fibrils reported in Fig. 5.8, show that the main peak characterizing AI region (1600-1700 cm^{-1}) is located at ~1626 - 1627 cm^{-1}, where the intermolecular β-sheets typically absorb. This band is usually accompanied by the occurrence of a less intense band at higher intensity addressed to intermolecular β-sheets; in this case, this band is centred at 1693 cm^{-1}. Interestingly, the secondary structure of both fibrils is characterized also by the presence of other secondary structure contributions. In fact, the region 1648-1675 cm^{-1} of second derivative spectra is composed by an ensemble of different

minima at 1648 cm^{-1}, 1656 cm^{-1}, 1666 cm^{-1} and 1673 cm^{-1} , thus indicating the presence of unordered structure or irregular α-helix, of α-helix, and of β-turn.

Table 5.3: FTIR secondary structure determined by the analysis of full-length WT and C-truncated aS fibrils, accompanied by their assignment.

WT aS fibrils	*aS (1-120) fibrils*	*Assignment*
1514 cm^{-1}	1514 cm^{-1}	Tyr-OH
1537 cm^{-1}	1539 cm^{-1}	
1551 cm^{-1}	1548 cm^{-1}	
1562 cm^{-1}	1562 cm^{-1}	
1583 cm^{-1}	1585 cm^{-1}	Gln
1591 cm^{-1}	1593 cm^{-1}	Tyr-OH
1614 cm^{-1}		Tyr-OH
1626 cm^{-1}	1627 cm^{-1}	β-sheets
1649 cm^{-1}		Unordered
1656 cm^{-1}	1655 cm^{-1}	α-helix
1666 cm^{-1}	1666 cm^{-1}	Unordered
1673 cm^{-1}	1673 cm^{-1}	β-sheets
1692 cm^{-1}	1692 cm^{-1}	β-sheets

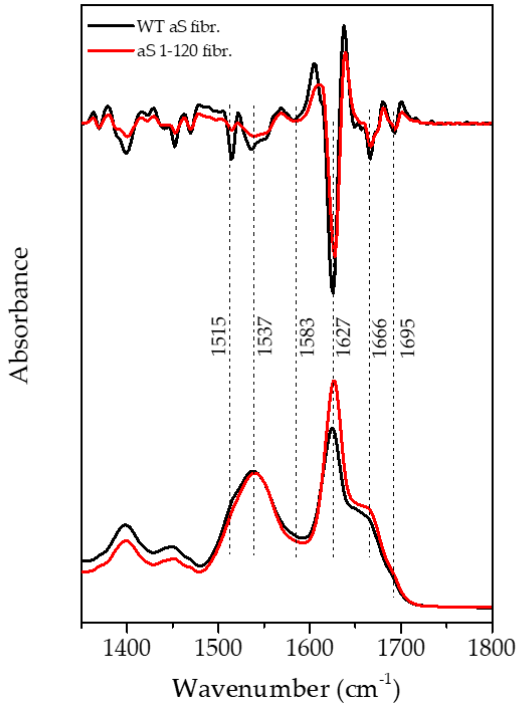


Fig. 5.8: (bottom) Absorbance spectra of full-length WT aS fibrils (depicted in black) and of C-truncated aS fibrils (in red) cut in the region 1350-1800 cm^{-1}. (top) Second derivative spectra of the corresponding absorbance spectra.

Dashed lines depicted the characteristic bands of the secondary structure. Minor contributions are reported in Table 5.3.

The tertiary structure of aS fibrils can be elucidated by the analysis of their UVRR spectra reported in Fig. 5.6, due to its sensitivity to aromatic ring vibrational modes (Tyr and Phe). The main contributions of UVRR spectra at 228 nm and 244 nm spectra arisen from Y8a, Y8b and Y9a bands. However, while in out-of-resonance condition (244 nm) this spectral region can be also affected to a minor extent by the vibrational modes of Phe residues, at 228 nm this problem could be overcome since Tyr/Phe enhancement ratio for Y8a is ~30 (see Table 5.4). Band fitting of the UVRR spectra at both wavelengths estimated the position of Y8a-Y8b and Y9a bands, taking into account that minor shifts in peak intensity are due to different spectral resolution. The results are reported in Table 5.5.

Table 5.4: Assignment of the UVRR peak commonly found in the spectra of full-length WT and C-truncates aS both at 228 nm and at 244 nm.

Peak position (cm^{-1})	*Label*
1003 cm^{-1}	F5 (Phe)
1032 cm^{-1}	F4 (Phe)
1178 cm^{-1}	Y9 (Tyr)
1207 cm^{-1}	F3 (Phe)
1208 cm^{-1}	Y7 (Tyr)
1606 cm^{-1}	F1 (Phe)
1606-1616 cm^{-1}	Y8a-Y9a (Tyr)

Table 5.5: Band fitting of the UVRR spectra at 228 and 244 nm have estimated the position of Y8a-Y8b and Y9a bands of full-length WT and C-truncated aS fibrils, with different spectral resolution.

Y8a and Y8b	*WT aS fibrils*	*aS (1-120) fibrils*
228 nm (1800 g/mm)	1615.6 ± 0.4	1612± 0.6
228 nm (3600 g/mm)	1605 ± 1; 1622.2 ± 0.5	1622.9 ± 0.8
244 nm (3600 g/mm)	1604 ± 1; 1616.5 ± 0.2	1601.1 ± 0.7; 1615.9 ± 0.5
Y9a		
228 nm (1800 g/mm)	1178.4 ± 0.7	1182 ± 2
228 nm (3600 g/mm)	1186.8± 0.7	1185 ± 1
244 nm (3600 g/mm)	1176.9± 0.3	1176 ± 0.5

As reported in Fig. 5.9, the presence of the Y8a-Y8b doublet is well-resolved in the 244 nm UVRR spectra. In fact, differently from the 244 nm spectra, where the Y8 band appears asymmetric (WT fibrils) or formed by two distinct peaks (1-120 fibrils), their contributions in the 228 nm UVRR spectra can be distinguished only after band fitting. In fact, due to the lower spectral resolution, the doublet appears as a single and broad band (see Fig. 5.9 in red).

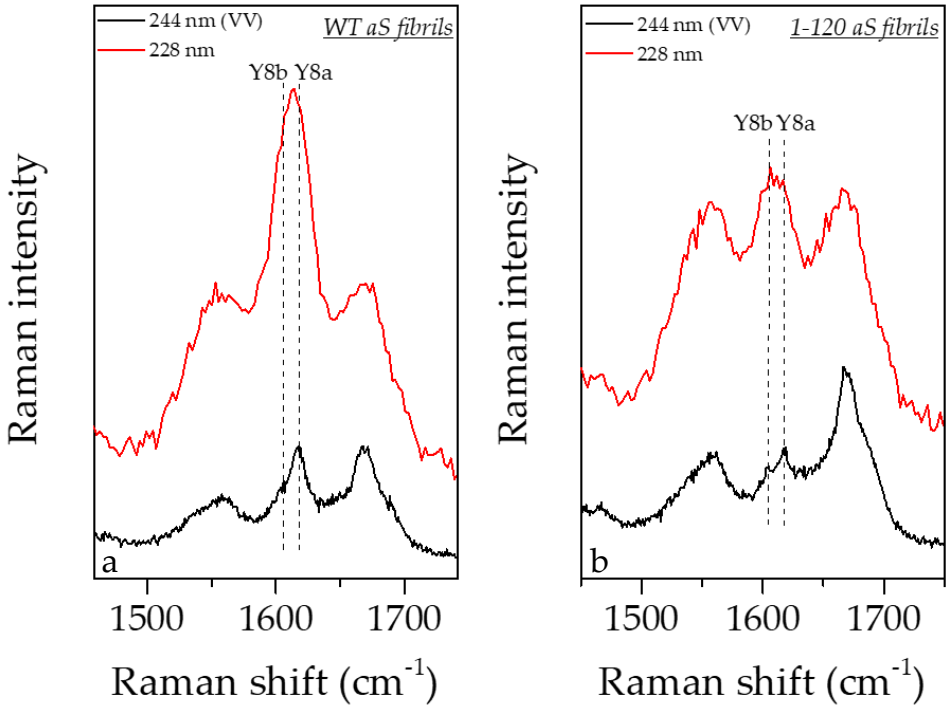


Fig. 5.9: (a) 228 nm (in red) and 244 nm (VV polarization, in black) UVRR spectra of full-length WT aS fibrils. The spectra of fibrils are zoomed in order to show AI (~1674 cm^{-1}), AII (~1560 cm^{-1}) peaks and Y8 band. The higher resolution spectrum at 244 nm shows that this peak is actually asymmetric and formed by the contribution of Y8b and Y8a peaks, whose positions are indicated by dashed lines. (b) 228 nm (in red) and 244 nm (VV polarization, in

black) UVRR spectra of C-truncated aS fibrils. Similarly to (a), the spectra were zoomed in order to show AI (~1674 cm^{-1}), AII (~1560 cm^{-1}) peaks and Y8 band. The higher resolution spectrum at 244 nm and the reduces contribution of the Y8a peak shows that Y8 peak is actually formed by two well resolved contributions (Y8b and Y8a peaks), whose positions are indicated by dashed lines.

As reported elsewhere[128], Y8a and Y8b are related to ring skeletal symmetric and asymmetric stretching vibrations (see also Table 2.4 in Chapter 2). In particular, Y8b position is sensitive to the hydrogen bonding state of the OH group: the stronger the hydrogen bonding, the lower is the vibrational frequency of this band. Also Y8a shifts upon hydrogen bonding, even if the extent is lower[233,235,350,351].

Differently to Y8a and Y8b, Y9a comprises exocyclic CH and phenolic OH in-plane bending motions, making it a good reported of the hydrogen bonding state of tyrosine residues[351]. The position of this band in both fibrils in the region 1176-1182 cm^{-1}, depending on the spectral resolution, indicates that tyrosine residues acts as H-donor or acceptor. However, due to the lack of 3 tyrosine residues in the case of (1-120) aS fibrils and to a possible modification of Tyr39 micro-environment due to the removal of C-terminus, the intensity of Y8 and Y9a bands is decreased compared to those characterizing the full-length aS. The intensity damping of Y8a, Y8b and Y9a between WT aS and 1-120 aS fibrils, estimated at each excitation energy is reported in Table 5.6.

Table 5.6: Extent of intensity damping for Y8a, Y8b and Y9a at 228 nm and 244 nm passing from full-length WT aS fibrils to C-truncated ones.

	$Y8a^{WT}/Y8a^{1-120}$	$Y8b^{WT}/Y8b^{1-120}$	$Y9a^{WT}/Y9a^{1-120}$
228 nm	2 ± 0.2		4 ± 2
244 nm	2.4 ± 0.6	2 ± 1	2.1 ± 0.5

The values reported in Table 5.6 are not statistically significant (p-value > 0.16, three replicates), thus indicating that the band intensity probed at 228 nm and 244 nm reduces by the same average factor (2.5 ± 0.7). As reported in Chapter 3 and 4, the intensity damping of a UVRR band could occur for multiple reasons, but the main

cause of quenching is the solvent-exposure of aromatic bands to water or a more polar solvent. Since Tyr39 has been reported as solvent protected by ssNMR and cryo-EM experiment and all the tyrosine of C-terminus completely exposed, we guess that the intensity of this band mostly arisen from the Tyr39 signal. Assuming that the area of the Y8 and Y9a bands is 100% addressed to a solvent-protected Tyr39, we could deduce that a two-fold decreasing of the Y8 band intensity in the (1-120) aS case could underlie that Tyr39 remains partially exposed to the solvent. This speculation will be furtherly explored in the next sections.

Differently from the 228 nm-UVRR spectra which is completely dominated by tyrosine vibrational bands, the 244 nm UVRR spectrum shows more sensitivity to Phe bands intensity modifications (see Fig. 5.10).

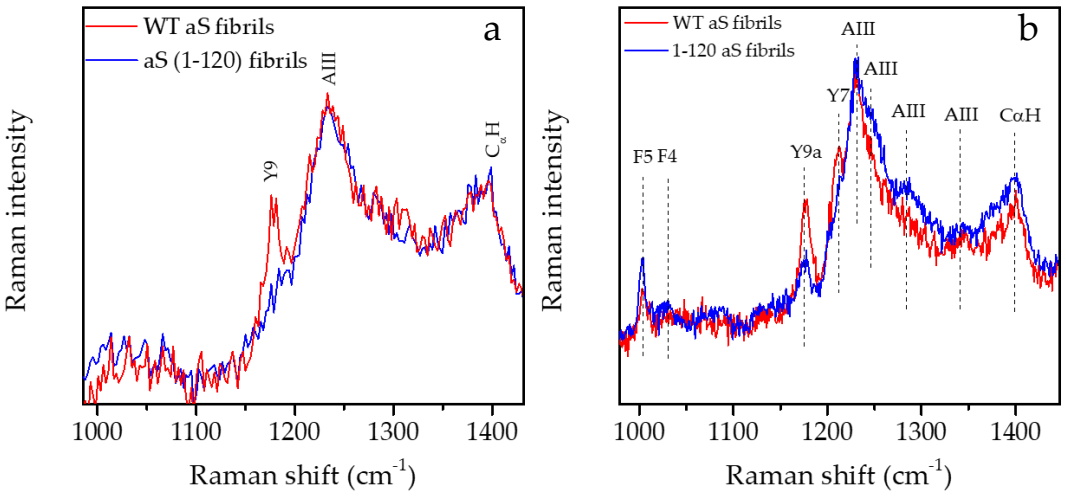


Fig. 5.10: (a) UVRR spectra of full-length WT and C-truncated aS fibrils (in red and in blue, respectively) collected at 228 nm. While AIII and CαH bands remain comparable between the fibrils, the intensity of Y9 band decreases, since C-truncated aS fibrils have only Tyr39. (b) UVRR spectra of full-length WT and C-truncated aS fibrils (in red and in blue, respectively) collected at 244 nm. Differently to (a), the reduction of Tyr/Phe ratio let Phe F5 and F4 bands visible at 1003 and 1032 cm^{-1}. Additionally, while Y9a and Y7 bands decreases their intensity upon the removal of 3 tyrosine residues in the C-terminus, AIII appears actually more structured with a main peak at 1230 cm^{-1} and other shoulders at higher energy as indicated by dashed lines. This structured peak highlights that C-truncated aS fibrils have a higher content of disordered and α-helical structures.

In particular, F4 and F5 bands located at 1003 cm^{-1} and 1030 cm^{-1}, respectively appear enhanced in the 244 nm spectra (Fig. 5.10 (b)) compared to the higher energy ones (Fig. 5.10 (a))[352]. Accordingly to Lednev and colleagues, the F5 band is dependent on

water exposure of Phe residues and this band is usually an indicator of local environment and of the arrangement of tertiary structure[136]. F5 band has been found increased when Phe residues are involved in a hydrophobic environment. Thus, in the case of (1-120) aS fibrils, Phe4 and Phe103 probe an average more hydrophobic environment compared to those in the WT structure.

5.8.2 nanoFTIR probes WT aS and (1-120) aS fibrils heterogeneity

The chemical and topographic features of WT and 1-120 aS fibrils are achieved using nanoFTIR techniques, reaching nanoscale resolution.

In Fig. 5.11, a distribution of WT aS fibrils is presented together with several images of single fibrils. The morphological features of these fibrils were complemented by the chemical maps of different regions alongside fibrils, characterized by a heterogeneous secondary structure. The distribution of fibrils is characterized by a width of ~5-7 nm and a distribution on different lengths, ranging from 150 nm to 650 nm. The estimation of fibrils length is biased due to difficulty in finding single fibrils, since they are prone to stuck together. As reported in Fig. 5.11 (e), the FTIR spectrum of the WT fibrils (in black) is compared to the average nanoFTIR spectra of the WT fibrils sampled in different positions. As previously reported, FTIR analysis reported that AI main peak is WT aS fibrils is centred at 1626 cm^{-1}, addressed to intermolecular β-sheets, while nanoFTIR analysis reported a heterogeneous distribution of the secondary structure: the sampling of single fibrils highlights that amide I main peak is shifted toward 1626 cm^{-1}, indicating the occurrence of a of β-sheets-rich structure, while amide I obtained for the random sampling of fibrils distribution (Fig. 5.11 (a)) highlights that amide I is shifted to 1656 cm^{-1}, addressed to the α-helical structures. This could be explained by the fact that monomers, dimers and intermediate species, which have a very different secondary structure compare to the fibrils one, co-precipitate with fibrillar species forming an heterogeneous matrix. Even if the β-

sheets structure is the leading component, nanoscale resolution allows to distinguish also minor secondary structure contributions.

Differently, aS (1-120) fibrils are presented as a distribution of clustered fibrils, characterized spectroscopically by a marked β-sheets secondary structure, confirming what observed by the FTIR band fitting. Several tentative of separation of aS fibrils have been conducted, but aS (1-120) fibrils seem to be highly prone to form clusters of fibrils rich in β-sheets. In fact, AI main peak is located at 1630 cm^{-1}.

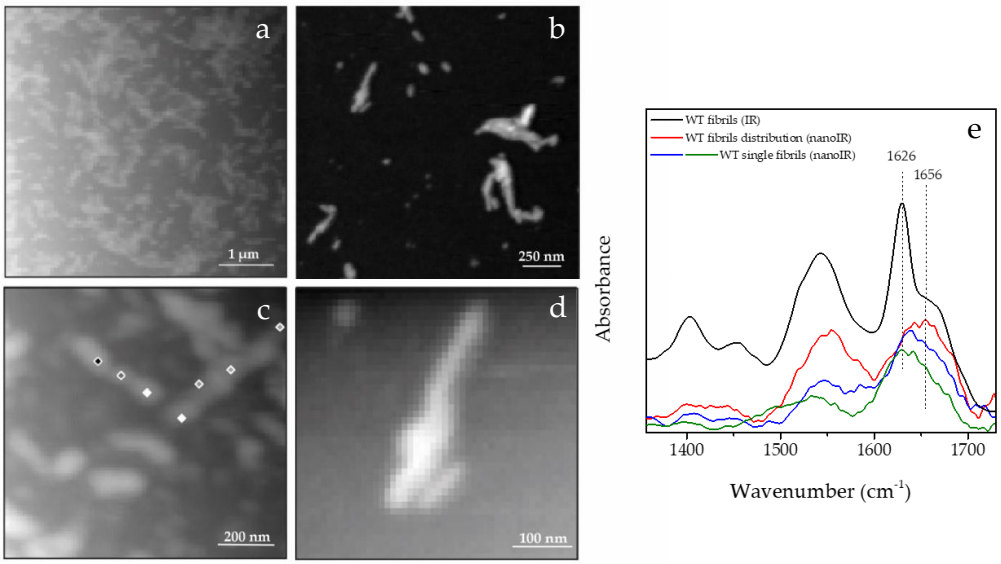


Fig. 5.11: nanoFTIR images of full-length aS fibrils. (a) Distribution of fibrils. The scale is 1000 nm. Distribution of single fibrils (b) with a scale of 250 nm and (c) of 200 nm. (d) Particular of two little fibrils stacked together (scale 100 nm). (e) nanoFTIR spectra of aS fibrils collected as single fibrils (in blue and green)or in a dense distribution (in red) and compared to the FTIR spectrum. Dashed lines indicate that the structure of full-length aS fibrils is heterogenous: mainly formed by β-sheets, it accounts also for contributions arisen from helical structure.

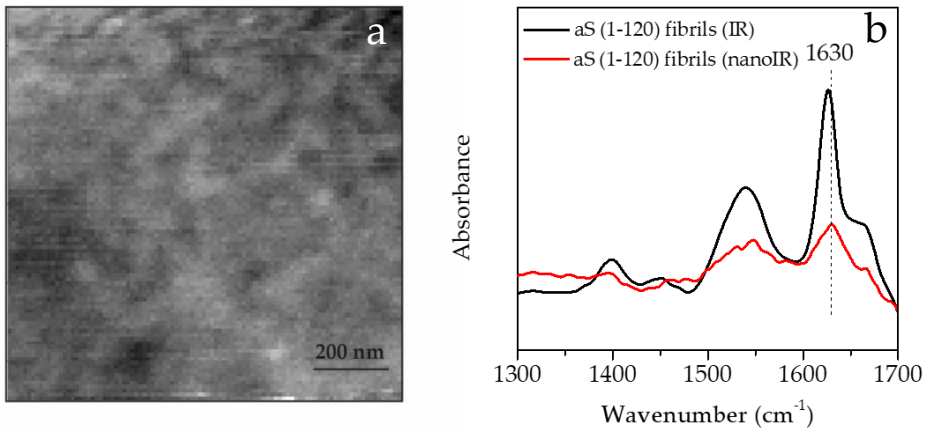


Fig. 5.12: (a) nanoFTIR images of C-truncated aS fibrils. Once washed, fibrils appear as a dense bundling of little fibrils. The scale is 200 nm. After several tentative, we could not able to find a region populated by single fibrils. (b) nanoIR spectra of C-truncated aS fibrils (in red) compared with their FTIR spectra. Apart from the low signal-to-noise ratio of the image, the chemical map suggests that the structure of C-truncated fibrils is rich in β-sheets with also minor contributions of unordered structures.

5.8.3 WT and truncated a-synuclein morphology probed by TEM

Fibrils morphology and, generally, their structural arrangement determines a different cytotoxic response both *in vivo* and *in vitro*, a modified aggregation kinetics and also a diverse response when treated with potential inhibitors, due to the diverse accessibility of specific portions. Thus, the overall morphology and distribution of full-length and C-truncated fibrils is reported in Fig. 5.13 and 5.14, respectively.

TEM images were taken at different spatial resolution in order to highlights minor morphological details. We observed that C-truncated aS fibrils have the highest tendency in forming clusters respect to the full-length fibrils, growing laterally from the major cluster (see Fig. 5.14). At comparable resolution, full-length fibrils appear more defined and longer than the C-truncated one, which actually resemble a bundle of highly interacting protofibrils.

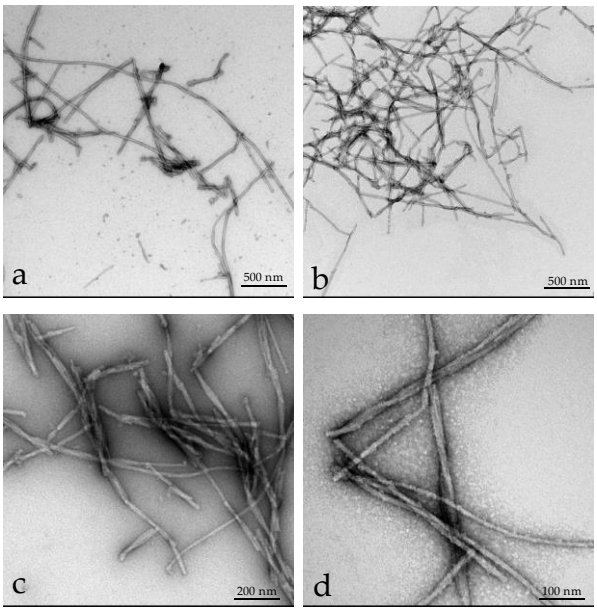


Fig. 5.13: TEM images of full-length WT aS fibrils taken at different scale. (a) and (b) with 500 nm bar scale; (c) at 200nm and (d) at 100 nm. Fibrils appear long and twisted.

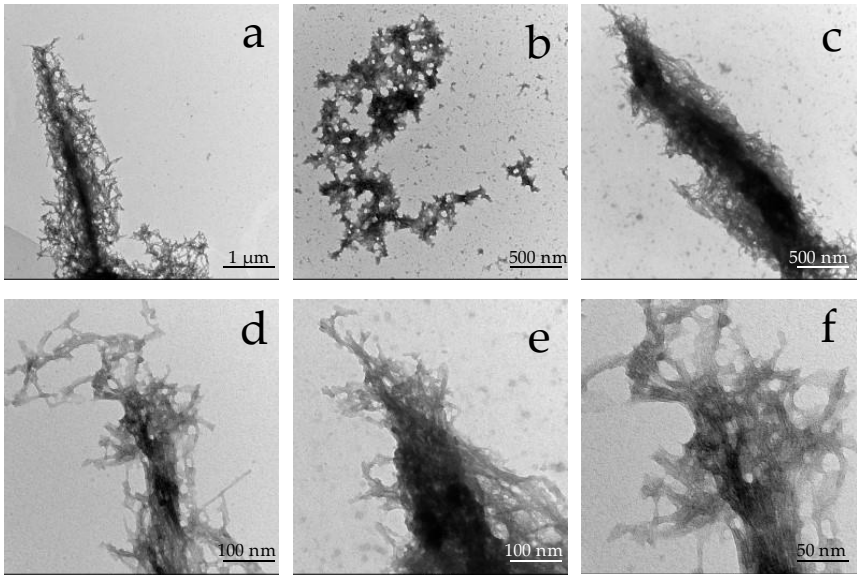


Fig. 5.14: TEM images of C-truncated aS fibrils taken at different scale. (a) 1000 nm (b) and (c) with 500 nm bar scale; (d) and (e) at 100nm and (f) at 50 nm. Fibrils appear little and tend to form clusters, they tend to grow laterally and does not have a defined structure compared to the full-length ones.

5.8.4 UVRR and FTIR probes WT aS and aS(1-120) monomers secondary and tertiary structures.

The 228 nm- and the 244 nm-UVRR (VV polarization) spectra of WT aS and (1-120) aS monomers are presented in Fig. 5.15 (a) and (b), respectively. As reported previously, the 228 nm-UVRR spectra of both monomers are dominated by the contribution of tyrosine bands. At first, the main differences highlighted by the 228 nm spectra are the intensities of Y8 and Y9a bands, that in the case of (1-120) aS are completely quenched, possibly due exclusively to the removal of the three C-terminus tyrosine residues. Differently, no remarkable differences in the amide bands location can be detected in the 228 nm-UVRR spectra.

A similar effect can be observed also in Fig. 5.15 (b), where the tyrosine residues signals appear strongly damped in the case of (1-120) aS monomers. In particular, the Y7 and the Y8a bands completely quench. Additionally, due to the higher resolution, the Y9a, Y7a and Y8a-Y8b are clearly resolved helping in the band fitting

procedure of the data (see the inset in Fig. 5.15 (b)). Moreover, AIII and AI band appear more structured and clearly visible at 244 nm.

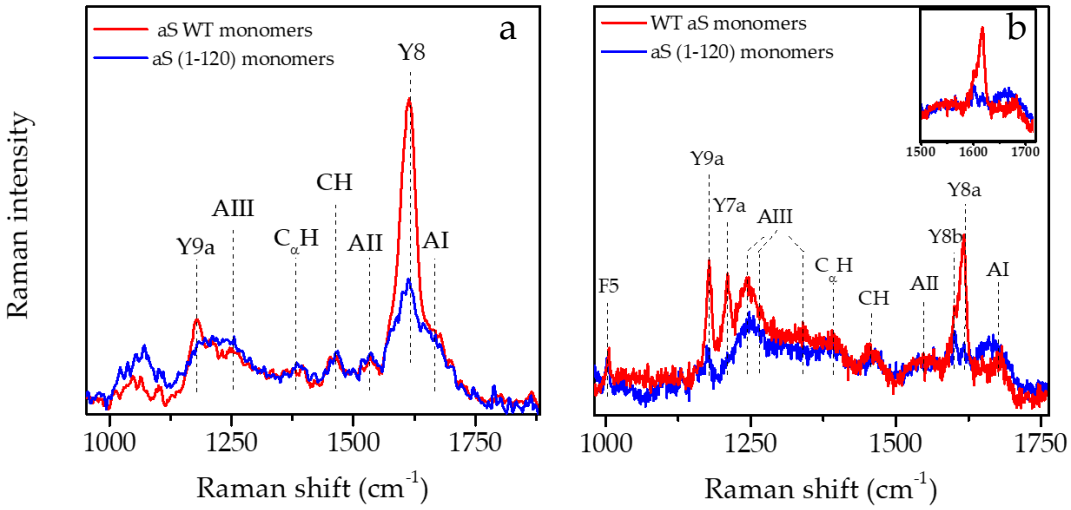


Fig. 5.15: UVRR spectra of full-length WT and C-truncated aS monomers collected at (a) 228 nm and (b) 244 nm and depicted in red and in blue, respectively. (a) The 228 nm-UVRR spectra highlights that both monomers have similar spectral features (see CH, CαH, AIII, AII and AI), but a decreased intensity of Y8 and Y9a band passing from the full-length to the truncated form. (b) similarly, the 244 nm UVRR spectra show that the secondary structure and most of the spectral features are conserved in both monomers, even if a strong modification of the Tyr-based peaks occurs upon removal of the last 20 residues of C-terminus: in fact, Y9a, Y7 and Y8b-Y8a bands appear strongly reduced. The inset shows the Y8b-Y8a doublet in both monomers, indicating that Y8a in the C-truncated monomers is strongly damped.

Thus, the determination of monomers secondary structure has been made by the band fitting of the 244 nm spectra. In particular, band fitting of VV, HV and ISO spectra of both monomers were made (reported in Fig. 5.16), finding that AI, AII and AIII do not depend strongly on the polarization selected.

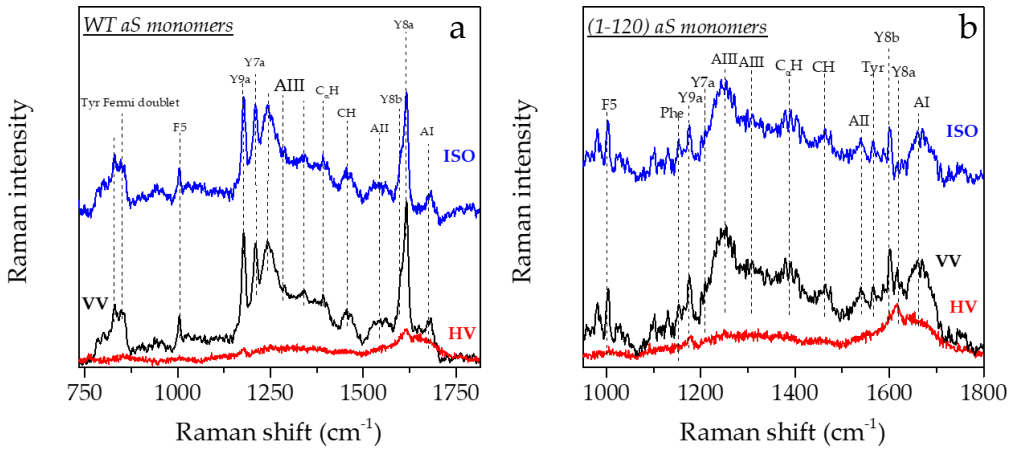


Fig. 5.16: 244 nm-UVRR spectra obtained in VV (black), HV (red) polarization and the calculated isotropic spectrum (in blue) for the full-length WT aS monomers (a) and C-truncated aS monomers (b). In both panels a complete assignment of the band of interest is reported.

The results of band fitting of both the WT aS monomers and the (1-120) aS monomers are reported in Table 5.7. AI, AII and AIII of the WT aS monomers and (1-120) aS monomers indicate that the secondary structure is multi-composed: in fact, AI presents a component located at 1649-1650 cm^{-1}, typical of an α-helix, and another one at ~1675-1680 cm^{-1}, typically found in β-sheets/unordered structure[129,221]. Accordingly, AIII shape is formed by several peaks located at ~1245 cm^{-1} (typical of β-sheets structure), at 1255 cm^{-1} (unordered structure)[221] and of α-helix (1292-1305 cm^{-1}). The secondary structure elucidated here is in agreement with previous study of Maiti and colleagues[129].

Table 5.7: The results of band fitting of both the WT aS monomers and the 1-120 aS monomers UVRR spectra taken at 244 nm. AI, AII and AIII peak position are reported to elucidate the secondary structure typical of monomers in phosphate buffer solution.

	Amide I (AI)	**Amide II (AII)**	**Amide III (AIII)**
WT aS monomers	1649 ± 2; 1681 ± 2	1550 ± 2	1245 ± 2; 1255 ± 3; 1305 ± 2
aS (1-120) monomers	1650 ± 1; 1675 ± 3	1539 ± 2; 1561 ± 4	1247 ± 2; 1292 ± 3;

Accordingly, FTIR spectra of both monomers highlight the presence of a similar secondary structure. In Fig. 5.17 the second derivative spectra of (1-120) aS and WT aS monomers taken twice are shown and those spectra share the same minima located at ~1625 cm^{-1} and 1695 cm^{-1}, typical of intermolecular β-sheets, at 1656 cm^{-1} (α-helix) and 1678 cm^{-1} (β-turn/unordered structures). The only difference between the aS monomers spectra is represented by the intensity damping of the tyrosine band at 1515 cm^{-1}, due to the loss of 3 tyrosine residues in the case of (1-120) aS.
In the 228 nm- and in the 244 nm-UVRR spectra reported in Fig. 5.16 no remarkable frequency variation of tyrosine and phenylalanine bands has been detected. However, the major difference can be observed in Fig. 5.18 and 5.19 where monomers and the corresponding fibrils are presented. Both at 228 nm and at 244

nm, UVRR spectra of aS fibrils show an enhanced contribution of AI, AII, AIII and $C_\alpha H$ compared to the corresponding aS monomers.

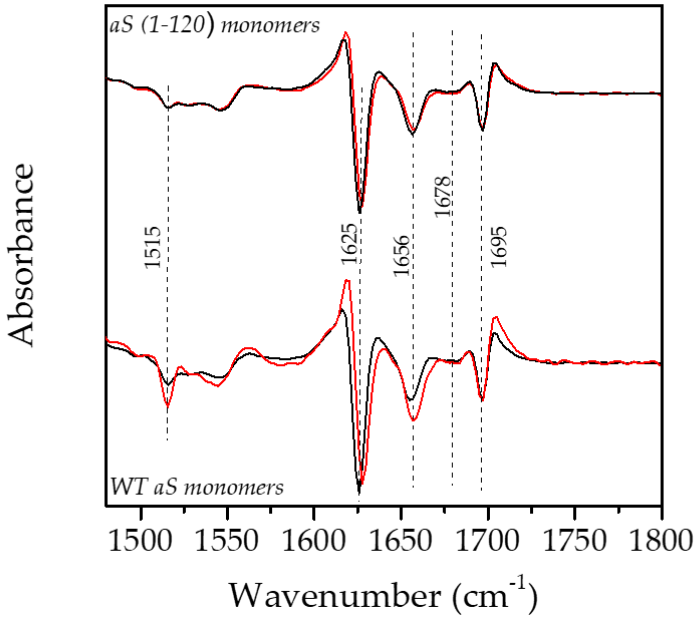


Fig. 5.17: Second derivative spectra of the FTIR absorbance spectra of full-length WT (in black) and C-truncated (in red) aS monomers. Dashed lines highlight the minima of interest indicating that amide I is multi-composed and characterized by β-sheets (1625-1695 cm^{-1}), α-helix (1656 cm^{-1}) and unordered structures/β-turn (1678 cm^{-1}).

In particular, in the case of WT aS, the formation of aS fibrils induces the slightly reduction of tyrosine bands intensities and an enhancement of AI, AII, AIII band intensities and a slightly shift of them since the secondary structure evolves into a more β-sheets-rich one.

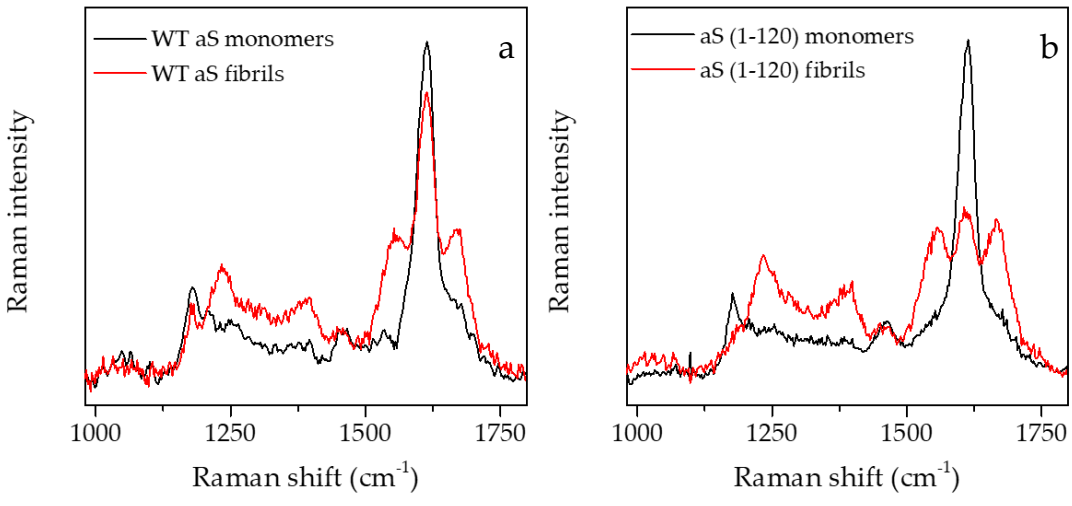


Fig. 5.18: 228 nm-UVRR spectra of (a) full-length WT aS monomers (in black) and fibrils (in red) and of (b) C-truncated aS monomers (in black) and fibrils (in red).. Both panels highlight that fibrillation induces the decreasing of tyrosine related bands, while increasing AI, AII, AIII and $C_\alpha H$ bands. This effect could be due to the orientation of fibrils respect to the polarization of the beam and to the sensitivity to the presence of a β-sheets-rich core typical of amyloid fibrils, i.e. the cross-β structure.

The comparison between the 228 nm- and the 244 nm- UVRR spectra (VV polarization) of WT and (1-120) aS actually highlights that fibrillation induces a strong enhancement of AI, AII, AIII and $C_\alpha H$ band intensity compared to the tyrosine band ones. Since fibrils are randomly oriented and the polarization angle of the electric field is constant, we could interpret this band intensity enhancement as a marker of the degree of fibrils alignment[353]. In fact, the anisotropy of these bands arisen from the preferential orientation of certain chromophores and their vibrational modes with respect to the fibril axis. AI is the most sensitive reporter to the cross-β core of amyloid fibrils, in fact it results highly enhanced in both cases due to the formation of this ordered structure. Interestingly, $C_\alpha H$ band increases its intensity due to the reduction of α-helix content upon fibrillation[354].

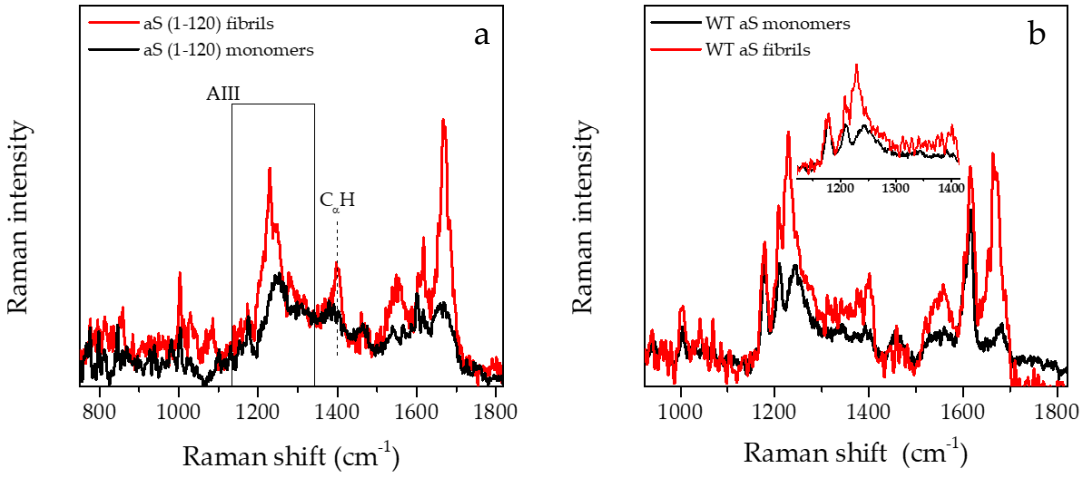


Fig. 5.19: 244 nm-UVRR spectra of (a) full-length WT aS monomers (in black) and fibrils (in red) and of (b) C-truncated aS monomers (in black) and fibrils (in red). Both panels highlight that fibrillation induces the a strong increasing of the AI, AII, AIII and $C_\alpha H$ bands. This effect could be due to the orientation of fibrils respect to the polarization of the beam to the sensitivity to the presence of a β-sheets-rich core typical of amyloid fibrils, i.e. the cross-b structure. AIII regions are highlighted in both panels in order to evidence the red-shift of the main peak upon the increasing of the β-sheets population.

A strong modification of both AI and AIII frequency is clearly reported in Fig. 5.19 (a) and (b) and Table 5.2 and 5.7: in fact, both aS monomers present a mixed secondary structure composition characterized by unordered structure and α-helix, while both fibrils are characterized also by a strong presence of β-sheets. The maximum of AIII is clearly highlighted in Fig. 5.19 (a) and (b) since in both cases results shifted to ~1230 cm^{-1}, typical identifying a β-sheets-rich population in fibrils, and it is structured in many little contributions accounting for the presence of unordered structures and a residual percentage of α-helix. However, as manifested in the enhancement of $C_\alpha H$, both fibrils are richer in β-sheets compared to the monomeric forms.

Additionally, the evident decreasing of tyrosine bands upon fibrillation is clearly manifested in Fig. 5.18 (a) and (b), indicating that the structural arrangement once aS is fibrillated contributes to the modification of tyrosine residues' microenvironments, likely promoting the solvent exposure of part of Tyr residues or their involvement in a non-covalent interaction with the solvent.

5.8.5 Alternative method of analysis: the use of a protein "external standard" and the indirect estimation of buried Tyr residues

As reported in Section 4.3, the use of internal standard is a common strategy adopted in order to quantify absolute modifications of the Raman peaks cross-sections. However, since internal standards are not resonantly enhanced in the spectral region of interest, they have to be used at higher concentration compared to the protein one, since their characteristic peak has to be clearly resolved in the spectra and cannot be hidden by the protein contributions. Typically, the concentration of an internal standard varies between 10 to 100 times that of the protein, possibly inducing modifications in protein hydration shells and also modulation of the phenomena under investigation[302]. To this purpose, several scientific papers studied the impact of these salts to the structural and morphological properties of certain biomolecules[303]. As an example, Munishkina and colleagues highlighted that anions promote fibrillation of aS by modifying the electrostatic interactions and by

screening the net charge of the protein, concluding that the basis of many endogenous and exogenous causes inducing fibrillation or the modification of protein secondary structure in general, is related to the interactions between anions, proteins and water[303]. For the same reason, Ramachandran avoided to use an internal standard while studying fibrillation of tau proteins[301].

Thus, to avoid possible misleading results mainly induced by the presence of internal standard, it could be proposed an alternative method aiming to roughly correlate the variation of peaks' intensities to a quantitative estimation, i.e. to the number of buried tyrosine residues.

This naïve method is based on the comparison between the UVRR spectrum of a reference protein, which has the same content of aromatic amino acids and a similar molar absorptivity at the wavelength used. Knowing how the tertiary structure of the reference protein is arranged and analysing the reference protein at the same protein concentration and in similar spectral conditions as for the protein studied, we are able to make a rough comparison of the structural arrangement of the protein studied.

In our case, insulin could represent a valid candidate as a reference protein, since it is formed by 4 tyrosine (Y14 and Y19 from branch A, Y16 and Y26 from branch B) and 3 phenylalanine (F1, F24 and F25 from branch B). Similarly, full-length WT aS is characterized by 4 tyrosine (Y39, Y125, Y133, Y136) and 2 phenylalanine (F4, F94). As reported in Fig. 5.20, the UVRR spectral shape and intensity of monomeric insulin and monomeric aS measured at the same concentration (250μM), at the same energy and in the same experimental conditions are highly conserved. Noteworthy, the impact of phenylalanine contribution on the UVRR spectra at 228 nm is negligeable compared to the tyrosine since the most intense band at 1003 cm^{-1} is completely absent at 228 nm (Fig. 5.20).

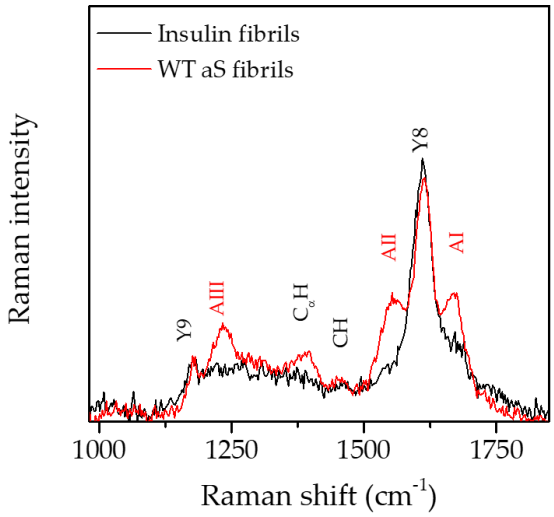


Fig. 5.20: 228 nm-UVRR spectra of insulin fibrils (in black) and full-length WT aS fibrils (in red). The peaks characterizing both spectra are highlighted. Differently to insulin, aS fibrils have higher intensity amide bands, but comparable Tyr peaks intensity.

However, Resonance Raman cross sections are usually affected by self-absorption, which determines the depth of penetration of the excitation beam and the consequent transmission of the Raman scattering out of the sample[132]. The self-absorption impact on the observed Raman radiation is low if the product $\varepsilon(\nu_0) \cdot r_{max} \cdot c\ (M)$ is small. In our case, this condition is verified; the concentration of both proteins is equal to 250 μM, their molecular absorptivity are comparable (5960 $cm^{-1}M^{-1}$ for full-length aS and 5734 $cm^{-1}M^{-1}$ for insulin) as well as their penetration depth; hence, the Raman signal intensity collected in backscattering is proportional to:

$$I_{Raman} \sim I_0 \cdot \varepsilon(\nu_0)^2 \cdot {\nu_0}^4 \cdot r_{max} \cdot c\ (M) \cdot N_A \cdot A \cdot K$$

Where I_0 is the incident radiation, $\varepsilon(\nu_0)$ is the molar absorptivity, r_{max} is the maximum depth of the sample, $c\ (M)$ is the molar concentration, N_A is the Avogadro number, A is the illuminated area on the sample and K is the spectrometer efficiency. As reported by Hong and Asher, this equation indicates that the observed Raman intensity increases with sample thickness and analyte concentration[132]. Additionally, the incoming radiation intensity, the setup performances and the illuminated area are the same since both proteins spectra have been taken in the same experiment.

Knowing the structure of both insulin monomeric and fibrillar forms, we consider the UVRR Y8 tyrosine peak intensity as a reference, since it contains information regarding the average tyrosine solvent-exposure. The Raman peak cross section of full-length aS and insulin peaks comparable and depends only on their solvent-exposure and on the possibility to form non-covalent interactions. Thus, by fitting the UVRR spectra with the protocol reported in Experimental Methods, we could have a rough estimation of the solvent-exposed tyrosine in the full-length aS.

As reported in literature, 3 tyrosine in insulin monomers are solvent protected; in fact, from Tyr autofluorescence experiments performed by Bekard and colleagues, TyrA14, TyrB16 and TyrB26 resulted resident in two hydrophobic faces of the monomers[241]. Differently, in the case of insulin fibrils, proteolysis and H/D exchange experiments as well as Tyr autofluorescence on insulin performed by Tito and colleagues and by Bekard and colleagues, respectively, showed that the region A13-A19 and B9-B19 on insulin remain solvent protected[241,246]. In particular, YA14 is solvent-protected and YB16 is mildly protected, while YA19 and YB26 of the C-terminus are exposed[246].

To start with, we have to verify that the intensity ratio of Y8 band of insulin monomers is approximately two times higher than the corresponding peak intensity in the case of insulin fibrils. Knowing that 3 tyrosine are solvent-protected in the case of insulin monomers, the 228 nm- UVRR intensity of the Y8 band of insulin monomers and fibrils were compared, obtaining an intensity damping factor of 2. Table 5.8 reports the estimation of the number of solvent-protected tyrosine in insulin fibrils, where emerges that 1.6± 0.1 tyrosine residues are solvent-protected, in agreement with the literature[246]. The difference between the number of buried Tyr in insulin fibrils and monomers is statistically significant (*p-value<0.01*).

After having roughly estimated the number of solvent-protected tyrosine residues in both monomers and fibrils of the reference protein, we apply the method on aS fibrils. The 228 nm-UVRR spectra of WT aS, (1-120) aS and insulin fibrils are reported in Fig. 5.21. Y8 an Y9a tyrosine bands of insulin fibrils and WT aS fibrils are identical, while a reduction of their intensity is visible in those of (1-120) aS fibrils.

Table 5.8: Estimation of the number of buried tyrosine in the case of insulin fibrils using the intensity variation of Y8 peak collected in two different experiments at 228 nm. The power of the beam and the resolution of the measurements are reported in the first column.

1 experiment (13,2 μW, 20 cm^{-1})	Y8 band intensity	Number of solvent-protected tyrosine
Insulin monomers	23512 ± 465	3
Insulin fibrils	12767± 300	1.6 ± 0.1
2 experiment (5 μW, 20 cm^{-1})	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Insulin monomers	10637 ± 243	3
Insulin fibrils	5422 ± 174	1.5 ± 0.1

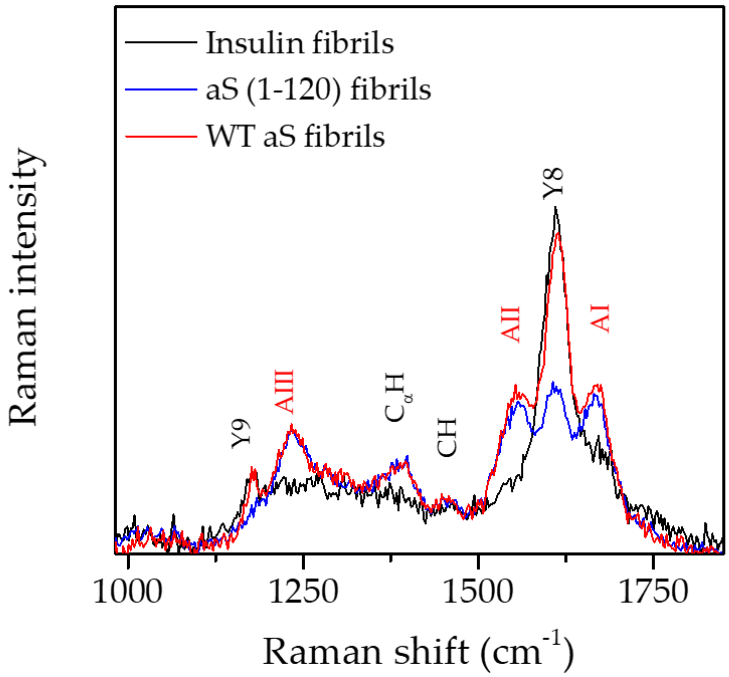


Fig. 5.21: 228 nm-UVRR spectra of full-length WT aS (in red), aS (1-120) (in blue) and insulin fibrils (in black) are shown. The spectra of aS fibrils share common spectral features even if Y8 and Y9a bands are strongly decreased in intensity by the removal of the tyrosine residues of the C-terminus.

The proposed method has been repeated three times using also different resolutions (20 cm^{-1} with the 1800g/mm grating and 10 cm^{-1} with the 3600 g/mm grating), taking each time insulin spectra and measuring the beam power on the sample. Table 5.9 summarized the results of each experiment and the ratio between Y8 band intensity of insulin fibrils and that of aS fibrils, knowing that approximately 1 tyrosine in insulin fibrils is buried and 1 is partially exposed. The data reported in Table 5.9

show that on average 1.2 ± 0.1 tyrosine residues are buried in the cross-β spine of the WT aS, which is in agreement with previous reported literature[57,330,331], suggesting that Tyr39 is likely part of the hydrophobic core of aS. The difference between the number of buried Tyr in insulin and full-length aS fibrils is statistically significant (*p-value<0.05*).

Table 5.9: the external standard approach has been repeated 3 times with different spectral resolution and beam power in order to reduce the structural variability affecting aS fibrils. Knowing the structure of the insulin fibrils, we roughly estimated the number of buried tyrosine in full-length WT aS fibrils. Beam power and spectral resolution are indicated in the first column.

1 experiment (13,2 μW, 20 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Insulin fibrils	12767 ± 300	1.6 ± 0.3
Full-length aS fibrils	10555 ± 1768	1.3 ± 0.4
2 experiment (12,6 μW, 20 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Insulin fibrils	12464 ± 174	1.6 ± 0.2
Full-length aS fibrils	7790 ± 530	1.0 ± 0.2
2 experiment (9 μW, 15 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Insulin fibrils	3552± 291	1.6 ± 0.1
Full-length aS fibrils	4372 ± 317	1.3 ± 0.3

In the following, we extend the analysis on (1-120) aS fibrils, whose spectra were collected together with the spectra of WT aS fibrils. The comparison between the Y8 band of WT aS fibrils and of 1-120 aS fibrils and the estimation of the buried tyrosine are reported in Table 5.10. From the data presented in Table 5.10, we could estimate that Tyr39 in (1-120) aS fibrils is partially exposed, since the average number of buried tyrosine residue is 0.6 ± 0.3. Unexpectedly, the twofold intensity decreasing reported here in the case of (1-120) aS fibrils is in agreement with our preliminary results reported in Section 5.8.1. In fact, from the quenching of Y8 and Y9a bands intensities passing from the full-length to the truncated species, we estimated that both the ratios $Y8^{WT}/_{Y8^{1-120}}$ and $Y9a^{WT}/_{Y9a^{1-120}}$ have a value close to 2. This indicates not only that Y8 band intensity is particularly sensitive to the solvent-

exposure of tyrosine, but that Tyr39 is partially solvent exposed in the C-truncated aS. Noteworthy, the use of a protein external standard somehow confirms this value: in fact, WT aS fibrils have 1.2±0.3 buried tyrosine residues and (1-120) aS fibrils have approximately 0.6 ± 0.3 tyrosine residues buried and the difference between the values is statistically significant (*p-value<0.01*).

Table 5.10: the external standard approach has been repeated 5 times with different spectral resolution and beam power in order to reduce the structural variability affecting aS fibrils. Knowing the structure of the full-length aS fibrils, we roughly estimated if Tyr39 is buried in the C-truncated aS fibrils. Beam power and spectral resolution are indicated in the first column.

1 experiment (13,2 μW, 20 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Full-length aS fibrils	10555 ± 1768	1.3 ± 0.4
aS (1-120) fibrils	5360 ± 453	0.7 ± 0.3
2 experiment (12,6 μW, 20 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Full-length aS fibrils	7790 ± 530	1.0 ± 0.2
aS (1-120) fibrils	4123 ± 528	0.5 ± 0.2
3 experiment (9 μW, 15 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Full-length aS fibrils	3552 ± 291	1.3 ± 0.3
aS (1-120) fibrils	3641 ± 181	0.5 ± 0.2
4 experiment (9,6 μW, 15 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Full-length aS fibrils	3852 ± 291	1.3 ± 0.3
aS (1-120) fibrils	2351 ± 566	0.7 ± 0.4
5 experiment (9 μW, 15 cm⁻¹)	**Y8 band intensity**	**Number of solvent-protected tyrosine**
Full-length aS fibrils	3339 ± 159	1.2 ± 0.2
aS (1-120) fibrils	1272 ± 132	0.5 ± 0.2

Additionally, this result is in agreement with previously reported Cryo-EM, H/D exchange ssNMR and proteolysis data exploiting the structure of aS (1-120)[331,343,355]. Those data evidenced that, actually, Tyr39 is located in a boundary region, thus it is located between the disordered portion formed by the first 37 residues of the N-terminus and the hydrophobic fibril core[331,355]. Additionally, while the outer surface of fibrils is mostly hydrophilic, Tyr39 is located within one of the 2 hydrophobic

clusters of aS, which is L38-V40[331]. Thus, in light also of the published literature, our data suggest that Tyr39 can be partially exposed to solvent, since it is located in a boundary region between a disordered and solvent-exposed portion of N-terminus and the hydrophobic core, closed to L38-V40 hydrophobic pocket.

The same calculations could be performed also with aS monomers using the 228 nm- UVRR spectra (Fig. 5.22). The comparison between the Y8 band of insulin monomers and that of WT an (1-120) aS monomers is reported in Table 5.11.

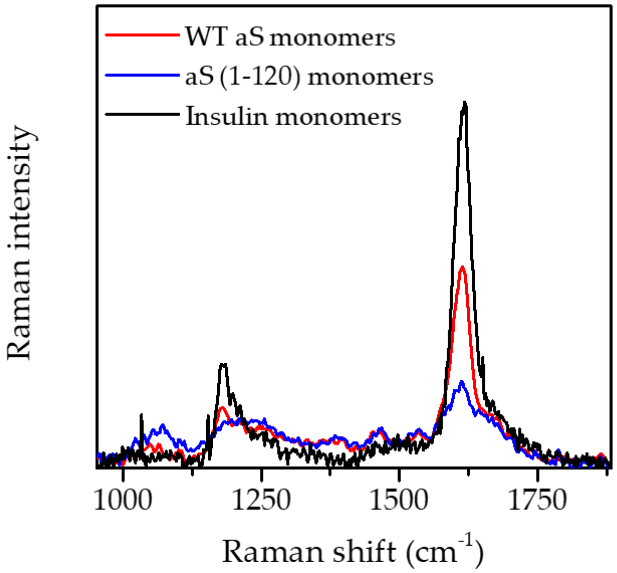


Fig. 5.22: 228 nm-UVRR spectra of insulin monomers (in black), full-length WT aS monomers (in red) and C-truncated aS monomers (in blue). The comparison highlights that the major differences regard the Y8 and the Y9a band of Tyr residues, which damped in the cased of aS monomers.

Table 5.11: the external standard approach has been repeated 2 times with different spectral resolution and beam power in order to compare the monomeric structure of insulin and aS. Knowing the structure of the insulin monomers, we roughly estimated if Tyr39 is buried in the C-truncated aS fibrils and how many tyrosine residues are buried in the case of full-length aS monomers. Beam power and spectral resolution are indicated in the first column.

1 experiment (13,2 μW, 20 cm⁻¹)	Y8 band intensity	Number of solvent-protected tyrosine
Insulin monomers	23512 ± 465	3
Full-length aS monomers	13890± 1768	1.8 ± 0.3
1-120 aS monomers	1788± 453	0.22 ± 0.05

2 experiment (9 μW, 15 cm⁻¹)	Y8 band intensity	Number of solvent-protected tyrosine
Insulin monomers	6876 ± 231	3
Full-length aS monomers	3555 ± 300	1.6 ± 0.3
1-120 aS monomers	907± 107	0.39 ± 0.08

Band fitting of the 228 nm-UVRR spectra shows that approximately 1.7±0.1 tyrosine are solvent protected in the case of full-length aS monomers, while in the case of (1-120) aS monomers Tyr39 is completely solvent exposed (0.31 ± 0.06). The difference between the two values is statistically significant (*p-value< 0.005*). The results obtained are summarized in Table 5.12 and Fig. 5.23:

Table 5.12: The results obtained by the "external standard" approach are summarized. The number of buried Tyr residues are indicated for the full-length and for the C-truncated aS monomers and fibrils.

	Full-length WT aS		*aS (1-120)*	
	Monomers	*Fibrils*	*Monomers*	*Fibrils*
Number of solvent-protected tyrosine	1.7 ± 0.3	1.2 ± 0.3	0.31 ± 0.06	0.6 ± 0.3

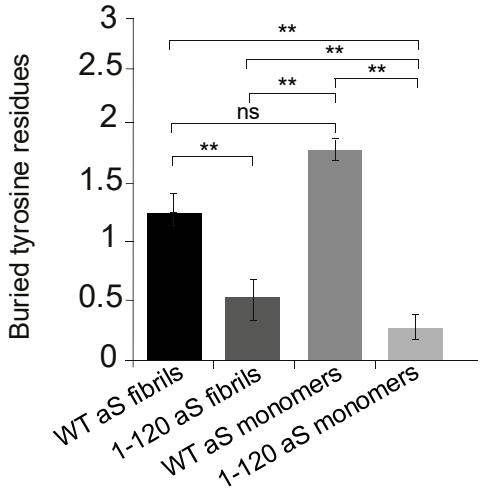


Fig. 5.23: graphical view of the number of buried tyrosine in the monomeric and fibrils phase of full-length WT and C-truncated aS, complemented by a statistical analysis. **indicates a p-value≤0.01 (statistically significant) and ns indicates a p-value >0.05 (not statistically significant).

5.8.6 Interaction between WT and aS (1-120) fibrils and Thioflavin T

In Fig. 5.24, the FTIR spectra of WT and 1-120 aS fibrils are presented and compared to those collected in presence of ThT.

The absorbance spectra and the second derivative spectra do not highlight relevant secondary structure modification in both cases; however, from the second derivative spectra, it is possible to identify that the interaction with ThT could induces a slight modification in the 1602 cm^{-1} and in the 1498 cm^{-1} band, possibly linked to the Tyr-O^- vibrational mode. However, the removal of the ThT contribution from the FTIR spectra of the protein mixed with the dye, indicates that the contribution highlighted by the dashed line actually results from the linear combination of the aS fibrils spectra and ThT spectra.

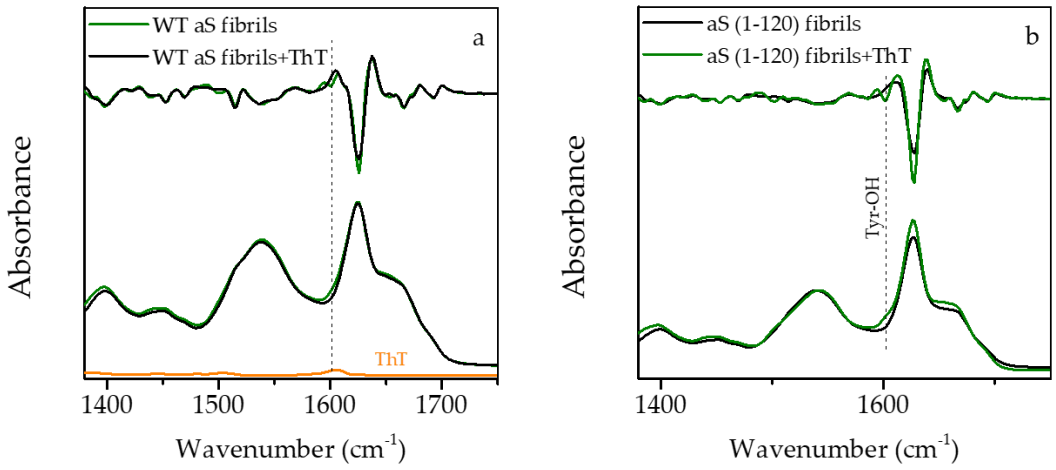


Fig. 5.24: (a) FTIR spectra of full-length WT fibrils obtained in presence (in green) and in absence of ThT (in black), accompanied by the second derivative spectra. The contribution of the ThT is highlighted in orange. (b)FTIR spectra of C-truncated aS fibrils obtained in presence (in green) and in absence of ThT (in black), accompanied by the second derivative spectra.

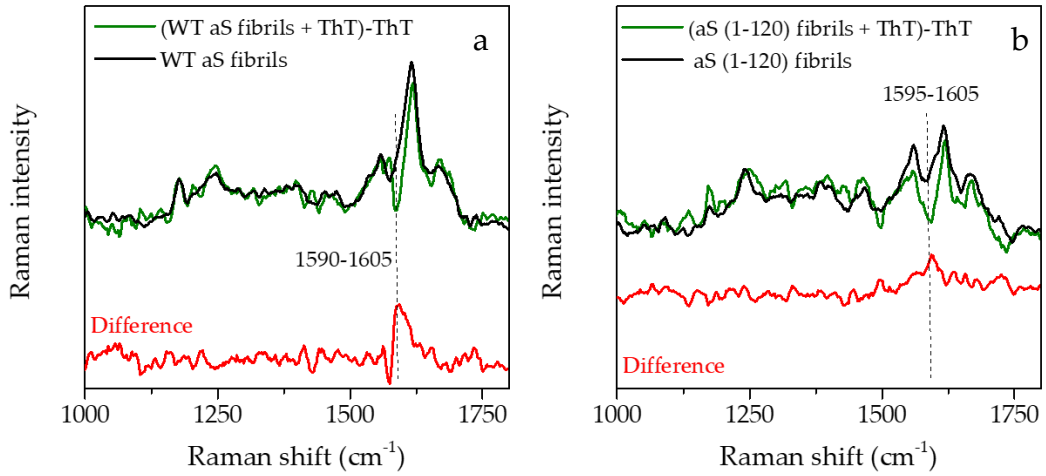


Fig. 5.25: 228 nm-UVRR spectra of (a) full-length WT aS fibrils in absence (black) or in presence (green) of ThT, and of (b) C-truncated aS fibrils in absence (black) or in presence (green) of ThT . The contribution of ThT and of the buffer have been properly removed by the protein spectra. The difference spectra between proteins incubated with or without ThT is reported in red, identifying the presence of a damped contribution at around 1600 cm^{-1} once ThT is present.

At the same time, the interaction between ThT and aS fibrils has been exploited by UVRR using a 228 nm-excitation wavelength. In Fig. 5.25, we compare the UVRR spectra of WT and (1-120) aS fibrils obtained in absence of ThT and in presence of the dye, after the removal of its contribution.

UVRR data indicates that the protein-dye interaction induces a modification of the spectral region comprises within 1590 and 1605 cm^{-1}; more specifically, a Raman peak of Phe at 1595 cm^{-1} and another the Y8b band of Tyr residues lie. In light of the FTIR results and due to the fact that Phe bands are not resonantly enhanced at this energy, we could suggest that the spectral modification induced in the UVRR spectra by the presence of the dye is related to a modification of the Tyr microenvironments.

Since the removal of the last 20 residues of the C-terminus does not induce remarkable modifications in the UVRR spectra, we deduce that ThT binds to the region (1-120). Being positively charged, ThT demonstrates low binding propensity to highly positively charged amyloid surfaces. Additionally, ThT binding is affected by ionic strength and pH of the buffer, since both factors could produce different polymorphs and a different distribution of ThT binding sites. Thus, it is likely that ThT benzothiazole ring is stabilized by the electrostatic interactions with negatively

charged residues[356,357]. At neutral pH, the N-terminus of aS has a net charge of +4, the NAC region of -1 and the C-terminus of -12[358]. However, it is unlikely that ThT binds to an unordered structure such as C-terminus since ThT recognizes the cross-β spine of protein fibrils and does not show interactions with disordered structures typical of monomeric proteins[56]. Despite they make part of the Greek key which is highly inaccessible also by proteinase K, residues 46-67 and 82-86 are retained the binding sites of phenotiazine-based PET radiotracers[56]. The net charge of the former portion is -1, while the portion 82-86 contains a positive and a negative charges, letting it neutral. Thus, residues 46-67 could be considered a possible binding region for the positively charged ThT, which is spatially closed to Y39 at the protofilaments interface. Since the UVRR vibrational modes of tyrosine residues are sensitive to the presence of ThT, we could ascribed that Tyr39 does not participate directly to the interaction, but can be considered closer to the ThT binding sites.

5.9 Discussion

The study of the structure and of the morphology of both full-length WT and C-truncated aS fibrils is of crucial importance since structurally diverse fibrils can results in a different cytotoxic response both *in vitro* and *in vivo*, in a different aggregation kinetics, inducing also a modified impact of inhibitors or PET tracers depending on their structural arrangement. At the same time, investigating the structure of full-length WT monomers and their C-truncated counterpart could provide unique data on the long-range interactions stabilizing both structures, indirectly indicating the extent of solvent-protection of Tyr39.

The need to find an alternative way able to measure aS monomers and fibrils structure, aiming to measure them in a less-biased manner, provides less-informative results but it ensures to measure proteins in a more native-like condition. Thus, aS fibrils and monomers structure has been elucidated through a multi-technique approach based on UVRR spectroscopy, FTIR, nanoFTIR spectroscopy and Trasmission Electron Microscopy (TEM). While UVRR spectroscopy has been used at two different wavelengths due to its pronounced sensitivity to different

molecular chromophores, providing important insight of the secondary and tertiary structure arrangement, FTIR spectroscopy has been employed for its outstanding sensitivity to protein secondary structure, complementing the results obtained by UVRR spectroscopy. nanoFTIR and TEM, instead, provided both a chemical and a morphological mapping of the both fibrils.

The study of aS fibrils by this multi-technique approach highlights that both WT aS and 1-120 aS fibrils are mainly arranged in a typical β-sheets-rich structure, typical of amyloid fibrils. The enhancement of $C_\alpha H$ band and the energy shift of AI, AII and AIII upon fibrillation to spectral position typically characterizing the presence of a β-sheets-rich structure is reported in the 228 nm-UVRR spectra and, more specifically in the 244 nm-ones which are more sensitive to the backbone vibrational modes. In particular, the 244 nm- and the 228 nm- UVRR spectra of aS fibrils are characterized by an increasing of the $C_\alpha H$ band intensity upon fibrillation, since the α-helical content present in the monomeric state reduces furtherly in favour an increased β-sheets population. In fact, the reorientation from helix to sheets could alter the Ramachandran angles, hence inhibiting the coupling between the $C_\alpha H$ bending and the C-N stretch of AIII[354]. Additionally, as already observed by Maiti and colleagues, AI and AIII of aS are multi-composed, presenting a secondary structure formed mainly by β-sheets, but with minor contributions arisen from α-helix and unordered structures[129]. The multi-composed secondary structure is revealed by both FTIR and nanoFTIR spectroscopy, since AI is mainly dominated by the ~1626-1630 cm^{-1} band addressed to the occurrence of intermolecular β-sheets, typical of amyloid fibrils. However, both the techniques highlight the presence of α-helical and disordered structures.

nanoFTIR and TEM elucidate also the structure of these fibrils. The morphology of aS fibrils strongly depends on several parameters such as buffer ionic force, on the extent of C- and N- termini truncation, on the occurrence of PD familiar mutations and so on. In fact, full-length and C-truncated fibrils appears radically diverse in morphology: full-length proteins are longer and more disperse, while the C-truncated ones tends to form dense bundling of protofibrils. The different morphology could explain why C-truncated aS fibrils show a higher tendency to self-

assembly compared to the full-length ones. The removal of part of the negatively and acid residues of C-terminus impairs the long-range interactions between N- and C-termini, destabilizing the charge distribution of the protein and inducing faster self-assembly, hence increasing the seeding ability of (1-120) aS fibrils compared to the full-length one *in vitro*[341,359]. Interestingly, the enhanced seeding activity of C-truncated aS fibrils compared to the full-length ones can partially explain why the majority of the nucleus of aS deposits are formed by truncated fibrils[341].

The emergence of different polymorphs upon C-terminus truncation underlies a different structural arrangement of residues onto fibrils surface and also a modification of the protofilaments interface, as occur in presence of point mutations[70,319]. This information is actually crucial for the design of potential fibrils disassembling agents or radioligands aiming to treat or detect specifically aS fibrils *in vivo*. In fact, a recent study evidenced that the binding between the amyloid inhibitor phthalocyanine tetrasulfonate (PcTS) to Tyr39, mediated by an aromatic interaction, could inhibit the aggregation process[97]. Thus, Tyr39 seems to be essential for aS functions[343]. The solvent accessibility of this residue has been evaluated by an alternative, internal standard-free UVRR approach. Due to its sensitivity to aromatic residues and to the lack of fluorescence background, UVRR becomes a unique tool to be used for the investigation of the arrangement of tyrosine residues and their microenvironments' in both fibrillar and monomeric proteins, opening the possibility to extend and optimize this approach also for other intrinsically disordered proteins lacking of Trp residues.

By the use of monomeric and fibrillar insulin as a reference, we investigate whether Tyr39 is solvent inaccessible, making part of the hydrophobic core. This approach has been tested many times since protein fibrils are affected by a high structural variability and this could partially lead to misleading information. The Y8 band located at 1615 cm^{-1} has been used for the estimation of the number of buried tyrosine due to its high intensity and sensitivity to local environment and residues solvent exposure. Knowing the extent of buried Tyr residues in both the monomeric and fibrillar insulin, we made several experiments measuring insulin and aS samples at

the same spectral resolution, with the same beam power ad at the same concentration.

From those experiments, we estimate that Tyr39 remains solvent-protected in the full-length aS, while it remains partially exposed in the C-truncated form in agreement with published literature[56,57,330,331,360]. In fact, Tyr39 makes part of the hydrophobic core of WT aS fibrils by reconstruction of electron density maps using cryo-EM, ssNMR, limited proteolysis experiments and H/D exchange NMR experiments[56,56,57,70,331,360,361], while the three Tyr residues of C-terminus have been found excluded from the core and completely solvent exposed. Thanks to the proline-rich nature of C-terminus, it remains unordered and behaves like a fuzzy tail in contact with the solvent. As previously reported, C-terminus covers an important role in aggregation since it disfavours aS tendency to aggregate by establishing long-range interactions with N-terminus, stabilizing aS in a compact random coil conformation and avoiding the solvent-exposure of the highly amyloidogenic NAC region. In fact, when the negative charge of C-terminus is screened by lowering the pH or by the presence of ligands such as polyamines, calcium or metal ions[55,296,361], these contacts are progressively lost and aS aggregation evolves faster. Noteworthy, not only (1-120) aS fibrillates faster, but as previously shown by TEM images, C-terminus truncation favours also the establishment of a different polymorphic structure compared to the full-length WT aS. In fact, the region close to Tyr39 is partially exposed in 1-120 aS fibrils and it is likely that the interface region between the two protofilaments is partially modified[331]. A partial exposure of Tyr39 in (1-121) aS has been also suggested by cryo-EM data published recently, where the first 37 residues of the N-terminal are shown to remain unstructured, while Tyr39 is located at the boundary of the core, close to a small hydrophobic pocket formed by L38 and V40 and could be partially exposed to the solvent[331].

The fact that a hydrophobic and aromatic amino acid at residue 39 remains solvent exposed could be harmful since it is able to form non-covalent interactions with other proteins or small hydrophobic ligands[97].

The β-sheets-rich secondary structure of aS fibrils is strongly reduced in aS monomers, which exist in a large ensemble of different conformations, namely with a high percentage of unordered structures. However, the secondary structure of full-length and (1-120) aS monomers is comparable. Generally, monomeric secondary structure strongly depends on the way it is measured, on the ionic strength of the buffer, on the pH of the solution, on the presence of membranes and ligands and so on. We used the approach of external standard also to indirectly elucidate the structure of full-length and C-truncated aS monomers.

While in the case of (1-120) aS monomers, Tyr39 seems to be completely solvent-exposed, in the case of full-length aS monomers we estimated that approximately 2 tyrosine residues are solvent-protected. Interestingly, HDX-MS experiments made on full-length aS monomers found the highest degree of protection at C-terminus (residues 115-125 and 135-140, where Tyr125 and Tyr136 reside) and a little bit lower protection in two portions of N-terminus (5-26 and 40-55)[361]. Additionally, NMR experiments identified two hydrophobic regions of aS monomers at residues 115−119 and 125−129 within the C-terminus[51]. Due to electrostatic interactions between C- and N- termini, full-length aS monomers result somehow folded in a more compact structure than a typical random coil[362]. This peculiar folding promotes the solvent shielding of more than a single Tyr residues of the C-terminus, presumably Tyr125 and Ty136. Complementing our UVRR results on full-length aS monomers with the data reported on literature, we could ascribed that Tyr39 remains exposed in the monomeric state. According to this UVRR approach, in the case of (1-120) aS monomers, Tyr39 seems to have the lowest degree of protection. Unfortunately, no data are reported in literature for comparison.

Overall, the alternative approach used to quantify the extent of solvent exposure of tyrosine residues in both WT and 1-120 aS monomers and fibrils provided experimental evidences in line with recent findings in literature and could be considered as valuable in the UVRR study of IDPs. Additionally, the information extracted by this approach could be also used for drug design purposes. As an example, phthalocyanines, which are amyloid inhibitor molecules, seem to bind in

the region 31-39 of aS, which is accessible in the monomeric and oligomeric phases, but not at when mature fibrils are finally formed[330,363].

The modification of aS monomeric structure upon fibrillation has to be taken into account dealing with drug design, aiming to develop drugs for the amyloidosis treatment and also radiotracers for the *in vivo* detection of aS fibrils into proteinaceous deposits in brain. PET radiotracers are usually aromatic-based compounds, smaller in size and able to come across the blood brain barrier. Even if the last requirement is not fulfilled, Thioflavin T can be used as an aromatic-based dye able to recognize amyloid fibrils. Despite ThT is not specifically designed for aS, it is able to detect the presence of amyloid fibrils in general. Thus, the interaction with aS could be investigated from a basic point of view, in order to elucidate ThT probable binding sites alongside the β-spine of the fibrils.

The interaction between ThT and aS fibrils, both full-length and C-truncated ones, has been investigated by FTIR and UVRR spectroscopies, showing that the interaction does not modify fibrils secondary structure, slightly affecting the Y8 band intensity of Tyr residues. After the removal of the UVRR signal of ThT, both UVRR spectra of aS fibrils present a decreased Y8 band intensity within the range 1595-1605 cm^{-1}. Since the removal of the last 20 residues of aS does not have an impact on the interaction with aS fibrils, and ThT does not bind to unordered structure and oligomeric species[364], we deduce that ThT does not bind to Tyr125, Ty133 and Tyr126 located in the C-terminus. In fact, ThT recognizes the β-sheets structure of the cross-β structure and, being positively charged, it is stabilized by the electrostatic interactions with negatively charged surfaces[356,357]. Apart from the fact that N-terminus has a net charge of +4 at neutral pH, two β-sheets making part of the Greek-key motif are more exposed respect to the central region, forming two negatively surfaces around cross-β spine[56]. These two portions are represented by residues 46-67 and 82-86, which are also retained as phenothiazine-based PET radiotracers binding sites[56]. Interestingly, almost all the PD familiar mutation of aS are found in the region 46-67, where certain substitutions such as E46K or G51D could impair the electrostatic interactions at the protofilaments interfaces, promoting the emergence of different polymorphs characterized by a modified cytotoxicity and aggregation

kinetic[319,322]. Noteworthy, recent studies indicate that the number of ThT binding sites and its affinity to aS fibrils is strongly polymorph-dependent[332,365]. As an example, ThT shows a higher binding affinity and fluorescent signal when bound to A53T-mutant aS compared to the A30P one and N-terminally acetylated aS, typically found in cytosol[41,335]. In fact, despite non-acetylated and N-terminally acetylated aS fibrils structures are highly similar, the latest one is characterized by the lack of a salt-bridge in the region K58-E61 enclosed in the 46-67 portion, letting K58 solvent exposed[335]. In light of this consideration we could speculate about the possible interaction between the positively ThT and the negatively β-sheets surface formed by residues 46-67. This could explain why PD familiar mutations, which impair the arrangement of protofibrils interfaces, could modulate the extent of ThT interaction[41]. From a structural point of view, the protofilaments arrangement into a single fibril lets the region 46-67 spatially closed to Y39 at the interface of the protofilaments. This implies a possible indirect perturbation of Tyr39 due to the interaction with ThT, since it is spatially close to the region 46-67. Obviously, the interaction between aS fibrils and ThT needs further experiments, especially with PD familiar mutated species which induces the occurrence of different aS polymorphs and modify the interface cavity between aS protofilaments.

5.10 Conclusions

The biophysical approach used for the investigation of the full-length and C-truncated aS monomers and fibrils allows to indirectly investigate the structural arrangement of both species, while evaluating also the morphological modifications. UVRR spectroscopy demonstrate to be a valuable tool for the investigation of the structural arrangement of peculiar resonant chromophores such as tyrosine residues. Both full-length and C-truncated fibrils present a β-sheets-rich structure, even if those fibrils present the occurrence of α-helical and unordered structures. They appear morphologically different, with C-truncated fibrils showing a higher tendency to form fibrillar cluster and bundling of shorter fibrils/protofibrils. The increased tendency to form fibrils cluster could be explained by the impairment of

the electrostatic forces upon C-terminus truncation. The post-translational mutation implies that (1-120) aS aggregates faster and show a higher seeding ability.

The alternative UVRR approach based on external standard not only offer the possibility to avoid the use of internal standard, able to modify the fibrillation kinetics and fibrils morphology, but it also helps estimating the number of solvent-protected tyrosine residues in both full-length and C-terminus monomeric and fibrillar species. The solvent accessibility of Tyr39 has been investigated by this approach, revealing that Tyr39 is actually buried in the hydrophobic core of the fibrils in the full-length case, while remaining partially exposed in the C-truncated case. Differently, in the monomeric phase, both full-length and C-truncated aS are characterized by a solvent accessible Tyr39. Additionally, this residue seems to be indirectly sensitive to the binding of ThT to fibrillar species, even if further experiments need to confirm the results obtained, eventually investigating other aS fibrillar polymorphs.

The structural arrangement and the solvent accessibility of potential binding sites such as Tyr39 is of fundamental importance, aiming to develop new drugs able to disassembled fibrils and to design specific and non-toxic radiotracers able to detect the presence of aS fibrils in brain proteinaceous deposits.

Conclusions of *Part 1*

The results reported in this Ph.D thesis highlight the potential of UV Resonance Raman scattering technique for the investigation of protein aggregation, of the effect of protein-ligand interaction and for the structural investigation of the structure of a well-known intrinsically disordered protein model, which is α-synuclein. Thus, we can summarize the main results as it follows:

- Chapter 3: the combination of UVRR spectroscopy with FTIR, fluorescence spectroscopy and AFM has been used to explore modifications occurring in protein secondary and tertiary structure upon fibrillation and ligand interaction. In particular, the use of UVRR spectroscopy provides clear markers of both fibrillation and interaction with ligands, being more informative than intrinsic fluorescence spectroscopy in the case of Trp-lacking proteins. The interaction between HEWL and insulin with resveratrol actually promotes a remodeling of protein fibrils and modifies the overall spectroscopic information. Thus, the characterization of fibrillar molecular architecture of model proteins by UVRR spectroscopy is important not only to answer fundamental questions related to protein folding, but it is also crucial for the development of therapeutics interventions and for diagnostics goals.

- Chapter 4: *E. coli* DHFR has been characterized by UVRR spectroscopy, in absence and in presence of two different molecular ligands such as trimethoprim (TMP) and spermine. We demonstrate that while TMP is able to inhibit the enzymatic activity, by blocking the DHFR structure in a low energy conformation without modifying significantly its 3D arrangement, spermine actually partially inhibits the activity of DHFR in a concentration-dependent manner and the way it binds to the enzyme induces a significant

modification in the Trp residues dihedral angle. Additionally, the structural inhibition is stable for 3 days, then the binding with spermine induces the precipitation of DHFR into fibrils as highlighted by TEM. This study reports the first evidence of DHFR induced fibrillation at neutral pH.

- Chapter 5: the structural arrangements of both fibrillar and monomeric structures of full-length and C-truncated aS have been provided by UVRR spectroscopy, using a novel approach based on the use of external protein reference. In particular, the solvent accessibility of Tyr39 has been investigated by this approach and complemented with the data already reported in literature, indicating that Tyr39 is actually buried in the fibrillar core in the case of full-length aS, while it remains partially exposed in the C-truncated aS fibrils. Differently, the monomeric phase of both full-length and C-truncated aS is characterized by a solvent-accessible Tyr39.

  Additionally, preliminary results indicates that Tyr39 seems to be indirectly sensitive to the binding of ThT, elucidating the mechanism of interaction between the protein and the ligand also looking at different polymorphs.

All these results indicate the potential of UVRR spectroscopy-based approach for proteins' structural information and for their molecular interactions with ligands.

List of figures – Part 1

(D) insulin. in the spectral region (200-328) nm. Shaded contours represent the variability of the absorption spectra in 3 different experiments.

3.6 (left) Normalized Trp fluorescence intensity of HEWL freshly prepared and fibrils solutions (in black and in red, respectively). Trp peak maximum red-shifts from 342 nm towards 345 nm during fibrillation. (right) Normalized Tyr fluorescence intensity of insulin freshly prepared and fibrils solutions (in black and in red, respectively). Tyr maximum peak does not shift in energy upon fibrillation.

3.7 (a) quenching experiment on HEWL+resveratrol at λ_{ex} = 280nm; (b) Stern–Volmer plot and (c) modified Stern Volmer plot for fluorescence quenching of HEWL; (d) Double logarithmic curve of HEWL-resveratrol solution. (e) quenching experiment on insulin + resveratrol at λ_{ex} = 276nm; (f) Stern–Volmer plot and (g) modified Stern-Volmer plot for fluorescence quenching of insulin; (h) double logarithmic curve of insulin-resveratrol solution.

3.8 (left) 10x10 μm^2 AFM topographic images of HEWL (HEWL in black, HR1 in red and HR2 in blue) and insulin (IE in black, IR1 in red, IR2 in blue) fibrils. False colour scales represent the 3D topographic height. (right) the distributions of fibrils contour length are presented. The most probable value for contour lengths as well as the 25^{th} and 75^{th} percentiles (in squared brackets) are summarized in the table.

3.9 FT-IR spectra of amide I and II bands (1450-1750 cm^{-1}) of (a) insulin fibrils in absence (in black) and presence of resveratrol (IR1 in red and IR2 in blue) and (b) of HEWL fibrils in absence (in black) and presence of resveratrol (HR1 in red and HR2 in blue). In (b) and (c) the respective second-derivative spectra presented.

3.10 (a) Trp fluorescence of HEWL fibrils solution in absence (in black) and in presence of resveratrol, i.e. HR1 (in red) and HR2 (in blue). (b) Tyr fluorescence of freshly prepared solution of insulin in absence (in black) and in presence of resveratrol, i.e. IR1 (in red) and IR2 (in blue). (c) *trans-* (in black) and *cis-* (in red) resveratrol spectra. *Trans-* isoform is characterized by a low quantum yield symmetric peak with a maximum at 400 nm, while *cis-* isoform by a high

in black), in HV (in red) polarizations. The calculated ISO spectrum is depicted in blue. Dashed lines indicate the corresponding peak.

5.7 244 nm-UVRR spectra of WT full-length fibrils and aS 1-120 fibrils in the region (980-1450) cm^{-1}. The spectra highlight that AIII is more structured in the case of C-truncated fibrils, underlying a more prominent presence of α-helix and unordered structures.

5.8 (bottom) Absorbance spectra of full-length WT aS fibrils (depicted in black) and of C-truncated aS fibrils (in red) cut in the region 1350-1800 cm^{-1}. (top) Second derivative spectra of the corresponding absorbance spectra. Dashed lines depicted the characteristic bands of the secondary structure. Minor contributions are reported in Table 5.3.

5.9 (a) 228 nm (in red) and 244 nm (VV polarization, in black) UVRR spectra of full-length WT aS fibrils. The spectra of fibrils are zoomed in order to show AI (~1674 cm^{-1}), AII (~1560 cm^{-1}) peaks and Y8 band. The higher resolution spectrum at 244 nm shows that this peak is actually asymmetric and formed by the contribution of Y8b and Y8a peaks, whose positions are indicated by dashed lines. (b) 228 nm (in red) and 244 nm (VV polarization, in black) UVRR spectra of C-truncated aS fibrils. Similarly to (a), the spectra were zoomed in order to show AI (~1674 cm^{-1}), AII (~1560 cm^{-1}) peaks and Y8 band. The higher resolution spectrum at 244 nm and the reduces contribution of the Y8a peak shows that Y8 peak is actually formed by two well resolved contributions (Y8b and Y8a peaks), whose positions are indicated by dashed lines.

5.10 (a) UVRR spectra of full-length WT and C-truncated aS fibrils (in red and in blue, respectively) collected at 228 nm. While AIII and $C_\alpha H$ bands remain comparable between the fibrils, the intensity of Y9 band decreases, since C-truncated aS fibrils have only Tyr39. (b) UVRR spectra of full-length WT and C-truncated aS fibrils (in red and in blue, respectively) collected at 244 nm. Differently to (a), the reduction of Tyr/Phe ratio let Phe F5 and F4 bands visible at 1003 and 1032 cm^{-1}. Additionally, while Y9a and Y7 bands decreases their intensity upon the removal of 3 tyrosine residues in the C-terminus, AIII appears actually more structured with a main peak at 1230 cm^{-1} and other shoulders at higher energy as indicated

by dashed lines. This structured peak highlights that C-truncated aS fibrils have a higher content of disordered and α-helical structures.

5.11 nanoFTIR images of full-length aS fibrils. (a) Distribution of fibrils. The scale is 1000 nm. Distribution of single fibrils (b) with a scale of 250 nm and (c) of 200 nm. (d) Particular of two little fibrils stacked together (scale 100 nm). (e) nanoFTIR spectra of aS fibrils collected as single fibrils (in blue and green) or in a dense distribution (in red) and compared to the FTIR spectrum. Dashed lines indicate that the structure of full-length aS fibrils is heterogenous: mainly formed by β-sheets, it accounts also for contributions arisen from helical structure.

5.12 (a) nanoFTIR images of C-truncated aS fibrils. Once washed, fibrils appear as a dense bundling of little fibrils. The scale is 200 nm. After several tentative, we could not able to find a region populated by single fibrils. (b) nanoIR spectra of C-truncated aS fibrils (in red) compared with their FTIR spectra. Apart from the low signal-to-noise ratio of the image, the chemical map suggests that the structure of C-truncated fibrils is rich in β-sheets with also minor contributions of unordered structures

5.13 TEM images of full-length WT aS fibris taken at different scale. (a) and (b) with 500 nm bar scale; (c) at 200nm and (d) at 100 nm. Fibrils appear long and twisted.

5.14 TEM images of C-truncated aS fibris taken at different scale. (a) 1000 nm (b) and (c) with 500 nm bar scale; (d) and (e) at 100nm and (f) at 50 nm. Fibrils appear little and tend to form clusters, they tend to grow laterally and does not have a defined structure compared to the full-length ones.

5.15 UVRR spectra of full-length WT and C-truncated aS monomers collected at (a) 228 nm and (b) 244 nm and depicted in red and in blue, respectively. (a) The 228 nm-UVRR spectra highlights that both monomers have similar spectral features (see CH, $C_\alpha H$, AIII, AII and AI), but a decreased intensity of Y8 and Y9a band passing from the full-length to the truncated form. (b) similarly, the 244 nm UVRR spectra show that the secondary structure and most of the spectral features are conserved in both monomers, even if a strong modification of the Tyr-based peaks occurs upon removal of the last 20 residues of C-terminus: in fact, Y9a, Y7 and Y8b-Y8a bands appear strongly reduced. The inset shows the Y8b-Y8a

doublet in both monomers, indicating that Y8a in the C-truncated monomers is strongly damped.

5.16 244 nm-UVRR spectra obtained in VV (black), HV (red) polarization and the calculated isotropic spectrum (in blue) for the full-length WT aS monomers (a) and C-truncated aS monomers (b). In both panels a complete assignment of the band of interest is reported.

5.17 Second derivative spectra of the FTIR absorbance spectra of full-length WT (in black) and C-truncated (in red) aS monomers. Dashed lines highlight the minima of interest indicating that amide I is multi-composed and characterized by β-sheets (1625-1695 cm^{-1}), α-helix (1656 cm^{-1}) and unordered structures/β-turn (1678 cm^{-1}).

5.18 228 nm-UVRR spectra of (a) full-length WT aS monomers (in black) and fibrils (in red) and of (b) C-truncated aS monomers (in black) and fibrils (in red).. Both panels highlight that fibrillation induces the decreasing of tyrosine related bands, while increasing AI, AII, AIII and $C_\alpha H$ bands. This effect could be due to the orientation of fibrils respect to the polarization of the beam and to the sensitivity to the presence of a β-sheets-rich core typical of amyloid fibrils, i.e. the cross-β structure.

5.19 244 nm-UVRR spectra of (a) full-length WT aS monomers (in black) and fibrils (in red) and of (b) C-truncated aS monomers (in black) and fibrils (in red). Both panels highlight that fibrillation induces the a strong increasing of the AI, AII, AIII and $C_\alpha H$ bands. This effect could be due to the orientation of fibrils respect to the polarization of the beam to the sensitivity to the presence of a β-sheets-rich core typical of amyloid fibrils, i.e. the cross-β structure. AIII regions are highlighted in both panels in order to evidence the red-shift of the main peak upon the increasing of the β-sheets population.

5.20 228 nm-UVRR spectra of insulin fibrils (in black) and full-length WT aS fibrils (in red). The peaks characterizing both spectra are highlighted. Differently to insulin, aS fibrils have higher intensity amide bands, but comparable Tyr peaks intensity.

5.21 228 nm-UVRR spectra of full-length WT aS (in red), aS (1-120) (in blue) and insulin fibrils (in black) are shown. The spectra of aS fibrils share common spectral features even if Y8 and Y9a bands are strongly decreased in intensity by the removal of the tyrosine residues of the C-terminus.

5.22 228 nm-UVRR spectra of insulin monomers (in black), full-length WT aS monomers (in red) and C-truncated aS monomers (in blue). The comparison highlights that the major differences regard the Y8 and the Y9a band of Tyr residues, which damped in the cased of aS monomers.

5.23 graphical view of the number of buried tyrosine in the monomeric and fibrils phase of full-length WT and C-truncated aS, complemented by a statistical analysis. ** indicates a p-value≤0.01 (statistically significant) and ns indicates a p-value >0.05 (statistically unsignificant).

5.24 (a) FTIR spectra of full-length WT fibrils obtained in presence (in green) and in absence of ThT (in black), accompanied by the second derivative spectra. The contribution of the ThT is highlighted in orange. (b)FTIR spectra of C-truncated aS fibrils obtained in presence (in green) and in absence of ThT (in black), accompanied by the second derivative spectra.

5.25 228 nm-UVRR spectra of (a) full-length WT aS fibrils in absence (black) or in presence (green) of ThT, and of (b) C-truncated aS fibrils in absence (black) or in presence (green) of ThT. The contribution of ThT and of the buffer have been properly removed by the protein spectra. The difference spectra between proteins incubated with or without ThT is reported in red, identifying the presence of a damped contribution at around 1600 cm^{-1} once ThT is present.

List of Tables – Part 1

reported to elucidate the secondary structure typical of monomers in phosphate buffer solution.

5.8 Estimation of the number of buried tyrosine in the case of insulin fibrils using the intensity variation of Y8 peak collected in two different experiments at 228 nm. The power of the beam and the resolution of the measurements are reported in the first column.

5.9 The external standard approach has been repeated 3 times with different spectral resolution and beam power in order to reduce the structural variability affecting aS fibrils. Knowing the structure of the insulin fibrils, we roughly estimated the number of buried tyrosine in full-length WT aS fibrils. Beam power and spectral resolution are indicated in the first column.

5.10 The external standard approach has been repeated 5 times with different spectral resolution and beam power in order to reduce the structural variability affecting aS fibrils. Knowing the structure of the full-length aS fibrils, we roughly estimated if Tyr39 is buried in the C-truncated aS fibrils. Beam power and spectral resolution are indicated in the first column.

5.11 The external standard approach has been repeated 2 times with different spectral resolution and beam power in order to compare the monomeric structure of insulin and aS. Knowing the structure of the insulin monomers, we roughly estimated if Tyr39 is buried in the C-truncated aS fibrils and how many tyrosine residues are buried in the case of full-length aS monomers. Beam power and spectral resolution are indicated in the first column.

5.12 The results obtained by the “external standard” approach are summarized. The number of buried Tyr residues are indicated for the full-length and for the C-truncated aS monomers and fibrils.

Part 2

Introduction

During the last six months of Ph.D, my experimental activity has been totally stopped due to the SARS-CoV-2 pandemic. During this period, I dedicated part of my time for the study of SARS-CoV-2 biology and genome. Thanks to the collaboration with Ulisse BioMed S.p.A, with Elettra Sincrotrone and with the Institute of Human Virology based in Baltimore (US), I provided important contributes in the field of Human Coronavirus[367–370].

RNA viruses such as SARS-CoV-2 are biological entities characterized by a high mutation rate. This peculiarity significantly contribute to their evolution and spread, improving their viral adaptation and fitness. To this purpose, we published a first work titled "*Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant*", in order to characterize the evolution of SARS-CoV-2 mutation pattern over the first months of the pandemic, aiming to elucidate the potential biological meaning of mutations. Interestingly, we found that SARS-CoV-2 is mutating in specific genomic hot-spots and that the RNA-dependent RNA polymerase (RdRp) started mutating once the virus penetrated in Europe. We also evidenced the emergence of a pattern of recurrent mutations, comprising the mutation in RdRp and the widely reported D614G spike protein mutation, which identify a new strain of SARS-CoV-2 virus (i.e. the G-clade).

A second work have been centered on the study of SARS-CoV-2 case fatality rate (CFR), highlighting that it is significantly affected by many parameters. We took into account the diagnostic test capability of each analysed country and

the lockdown strategies which have been adopted to reduce the spread of the virus, elucidating the effect they exert on the CFR. Thanks to a novel mathematical model we were able to demonstrated that strong lockdown strategies, complemented with a large testing campaign, strongly reduce the CFR. Countries adopting less rigid lockdown strategies or making a less aggressive diagnostic test campaign have a higher CFR, indicating that both aspects are crucial for containing pandemic. In the same work we characterized the emergence of other viral mutations in Europe and US.

Brief overview

Coronaviruses are a peculiar class of viruses able to infect many animals, causing mild-to-severe respiratory infections. Among all the reported cases, two coronaviruses of zoonotic origin have caused severe respiratory syndromes, namely severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and Middle East respiratory syndrome coronavirus (MERS- CoV) in 2012. However, at the end of 2019, a novel coronavirus provoking a new severe acute respiratory syndrome was identified firstly in the city of Wuhan (China) and rapidly it spread all around the world, causing the ongoing pandemic[371].

The novel coronavirus has been named SARS-CoV-2 since it shares 79 % genome sequence identity with SARS-CoV and it is similarly organized. It is a single-stranded RNA-enveloped virus, counting for more than 29800 nitrogenous base pairs (see Fig. 2_0). SARS-CoV-2 genome is composed by gene fragments expressing structural and nonstructural proteins[371]. The replicase gene occupies the first 22000 bps of the whole genome and encodes a large polyprotein, which is proteolytically cleaved into 16 non-structural proteins once the virus has entered into cells. This important gene is responsible for the replication and transcription of the virus and this gene shares more than 85% identity with the replicase gene of SARS-CoV[371]. Among the most important nonstructural proteins, it is important to mention the RNA-dependent RNA polymerase, the papain-like protease and the 3-chymotrypsin-like protease.

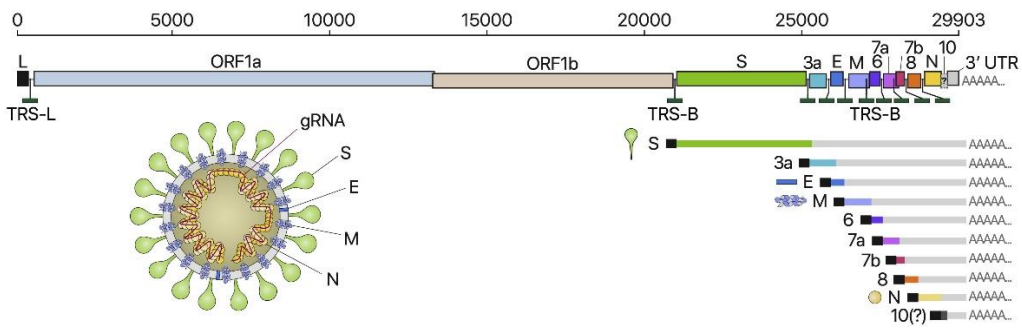


Fig. 2_0: Schematic representation of the SARS-CoV-2 genome architecture and of the virions structure (bottom left). ORF1a and ORF1b are translated into a polyprotein from the full-length genomic RNA (29,903 bps), that also serves as an mRNA. S, E, M and N genes are highlighted in green, blue, lavender and yellow. The other boxes (3a, 6, 7a, 7b, 8, 10) are accessory genes encoding for accessory proteins. The figure has been readapted from[372].

Differently, the last portion of the SARS-CoV-2 viral genome is composed by four structural genes: N (nucleocapsid), E (envelope), M (membrane) and S (spike). The structural genes of SARS-CoV-2 share more than 90% identity with SARS-CoV, with the exception of the S gene. S gene encodes the spike protein, which covers the outer corona of the virus and mediates the entry of the virus into cells. In fact, spike proteins bind to the cell receptor angiotensin-converting enzyme 2 (ACE2), inducing the entry of the virus inside the cell, thus exploiting the cellular machinery for its replication. Once the virus enters cells, the viral RNA is released and is traduced into polyproteins in order to replicate the RNA viral genome and to produce novel virions in order to infect other cells.

1 Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant

1.1 Introduction

The recent emergence of the novel, human pathogen Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in China and its rapid national and international spread poses a global health emergency. On March 11th 2020, WHO publicly declared the SARSCoV- 2 outbreak as a pandemic. In a few weeks, the virus caused thousands of deaths worldwide, strongly impacting the global economy and human habits. SARS-CoV-2 is an enveloped, +ssRNA virus, belonging to the *Betacoronavirus* genus which includes two other RNA viruses that have caused recent important epidemics: Severe Acute Respiratory Syndrome (SARS) caused by SARS-CoV, and the Middle East Respiratory Syndrome (MERS) by MERS-CoV. Noteworthy, some evidence has been recently provided, supporting that SARS-CoV-2 mortality can significantly differ depending on the geographic area. For example, Baud and colleagues reported that mortality rate is three times higher out of China (15.2% [95% CI 12.5–17.9] out of China, compared to 5.6% [95% CI 5.4–5.8] in China)[373]. This rate has been re-estimated by dividing the number of deaths on a given day by the number of patients with confirmed SARS-CoV-2 infection 14 days before, considering the WHO data relative to the cumulative number of deaths to March 1st, 2020[373]. Differences in viral infection rates can be due to a combination of factors, including different national strategies adopted for people movement restrictions, isolation and quarantine, different genetic population herd immunity. Mortality differences are to understand, but viral mutations and evolution capability over time may be important.

RNA viruses mutation rate is dramatically high, up to a million times higher than that of their hosts and this high rate is correlated with virulence modulation and evolvability, traits considered beneficial for viral adaptation[374]. Wang and coworkers have recently characterized 13 variation sites in SARS-CoV-2 ORF1ab, S, ORF3a, ORF8 and N regions, among which positions 28144 in ORF8 and 8782 in ORF1a showed a mutation rate of 30.53% and 9.47%, respectively[375]. Prior reported results show that SARS-CoV-2 is rapidly moving across countries and genomes with new mutation hotspots are emerging. RNA virus mutation rate contributes to viral adaptation creating a balance between the integrity of genetic information and genome variability[376–378]. Biological characterization of viral mutations can provide precious insights for assessing viral drug resistance, immune escape and pathogenesis related mechanisms. Additionally, viral mutation studies can be crucial for designing new vaccines, antiviral drugs and diagnostic assays. The viral genome mutagenic process depends on the viral enzymes that replicate the nucleic acids, influenced by few or no proofreading capability and/or post-replicative nucleic acid repair. Other mutation-generating processes include: host enzymes, spontaneous nucleic acid damages due to physical and chemical mutagens, recombination events and also particular genetic elements responsible for production of new variants. Mutation rates are modulated by other factors such as determinants of the template sequence and structure involved in viral replication. RNA-dependent RNA polymerases (RdRps) are multi-domain proteins able to catalyze RNA-template dependent formation of phosphodiester bonds between ribonucleotides in the presence of divalent metal ion[379–381]. In most viruses, RNA polymerase lacks proofreading capability, with some exceptions such as *Nidovirales* order (to which the *Coronavirus* genus belongs), that stands out for having the largest RNA genomes. *Nidoviruses* are characterized by a complex machinery dedicated to RNA synthesis, that is operated by non-structural proteins (nsps), being produced as cleavage products of the ORF1a and ORF1b viral polyproteins[382] to facilitate virus replication and transcription.

The SARS-CoV-2 RdRp (also named nsp12) is a key component of the replication/transcription machinery. SARS-CoV-2 shares a high homology for nsp12 compared to SARS-CoV, suggesting that its function and mechanism of action might be well conserved[383]. This has been confirmed by a recent cryo-EM structural study obtained for SARS-CoV-2 nsp12[384]. In SARS-CoV, an exonuclease activity with proofreading function has been reported for the nsp14 (ExoN), and a homologue nsp14 protein is found in the SARS-CoV-2 as well[383,385]. ExoN increases the fidelity of RNA synthesis by correcting nucleotide incorporation errors made by RdRp[386]. Genetic inactivation of the coronavirus ExoN results in a 21-fold decrease in replication fidelity compared to wild type SARS-CoV[387]. Moreover, Kirchdoerfer and colleagues showed the involvement of nsp7 and nsp8 in the formation of a supercomplex with RdRp in SARS-CoV[388], and this has been confirmed also for SARS-CoV-2 in a recent study unveiling the structure of SARS-CoV-2 RdRp/nsp7/nsp8 complex[384]. This complex ensures RdRp processivity, becoming fundamental in the transcription fidelity. Nevertheless, the critical SARS-CoV RdRp residues involved in ExoN, nsp7 and nsp8 interaction have still to be identified. RdRps are considered among primary targets for antiviral drug development, against a wide variety of viruses. Some RdRp inhibitors have been considered to target SARS-CoV-2: Favipiravir[389], Galidesivir[390], Remdesivir[391] and Ribavirin[392]. Interestingly, the docking site is not located in proximity to the catalytic domain of the RdRp[393]. In addition, other possible drugs such as Filibuvir, Cepharanthine, Simeprevir and Tegobuvir, are predicted to be potential inhibitors of RdRp[394]. Naturally occurring mutations in critical residues for drug efficacy can lead to drug resistance phenomena, with a significant loss in the binding affinity of these molecules to the RdRp.

We focused our study on SARS-CoV-2 mutations in order to assess if new viral variants were spreading across the Countries. This characterization of SARS-CoV-2 variants could lead to better therapeutics treatments, vaccines design and diagnostics approaches.

1.2. Methods

SARS-CoV-2 virus reference sequence used for the analysis was deposited in January 2020 by Wu and coworkers[383] formerly called "Wuhan seafood market pneumonia virus" (WSM, NC_045512) (https://www.ncbi.nlm.nih.gov/nucco re/NC_04551 2**)**. GISAID database (https://www.gisaid.org/) filtered from December 2019 up to March 13th, 2020 was used to collect 220 SARS-CoV-2 complete genomes of different patients all around the world (i.e. China, USA, Canada, Australia, United Kingdom, Germany, France, Japan, Italy, Switzerland, Singapore,Luxembourg, Netherlands, Spain, Portugal, Sweden, Czech Republic, Thailand, India, Cambodia, Hong Kong, Finland, Singapore, and Ireland) taking into particular consideration those deposited during the development of European outbreaks. Only complete genomes (28000–30,000 bps) were analysed. Clustal Omega, Serial Cloner and Blast tools were used to conduct multiple sequence alignment, comparing WSM sequence to sequences isolated from patients, whereas Swiss Model and Ez-mol were used for protein modelling. The statistical analysis was performed by R software. We first checked the normality of data distribution with the Shapiro–Wilk test, expressing the continuous variables as median and range (min–max). Categorical variables were expressed as absolute frequency and percentages. Nonparametric Mann–Whitney and Fisher-Exact tests were used to compare the number of mutations per genome with at least one of the selected mutations with respect to the group of genomes that do not present the specific mutation analysed. All p-values were calculated from 2-sided tests using 0.05 as the significance level.

1.3 Results

1.3.1 Identification of recurrent mutation hotspots in different geographic areas

A database of 220 complete SARS-CoV-2 patient-isolated genome sequences randomly collected from the GISAID database were aligned and compared to the WSM SARS-CoV- 2 reference genome. In particular, 5 patient-isolated genomes were submitted to the GISAID database in December 2019 (2.3%), 67 in January 2020 (30.45%), 67 in February 2020 (30.45%) and 81 (36.8%) up to the 13th of March 2020. About 33.6% of complete genomes belong to patients aged less than 44 years old, which is the average age of the patients included in the database. The majority of patients are men (55.5%). We divided our dataset into 4 geographic areas: Asia, Oceania, Europe, North America (Fig. 2_1). Within each area we performed alignment analysis comparing patients' genomes with the reference sequence. The Asian group comprises genomes obtained from patients located in China, Japan, South-East-Asia and India. The Oceanian group comprises genomes from Australian patients, whereas the European one includes every genome obtained from patients located in each one of the European states (Spain, Portugal, United Kingdom, Netherlands, Italy, Germany, Switzerland, France, Luxemburg, Sweden, Finland, Denmark and Belgium). Finally, the North America group contains genomes from US and Canadian patients.

We evaluated the distribution of SARS-CoV-2 mutations through different geographic areas (see Fig. 2_1), calculating the mutation frequency within these 4 geographic areas, by normalizing the number of genomes carrying a given mutation per geographic area. We confirmed the occurrence of mutations located at positions 3037 8782, 11083, 28144 and 26143[395–397]. Moreover, we highlighted the presence of additional "conserved mutations" in all the geographic areas, taking into account only those occurring more than 10 times in our database. Those with a lower occurrence were not reported. These mutations were found in position 1397, 2891, 14408, 17746, 17857, 18060, 23403, 28881, belonging to ORF1ab (1397 nsp2, 2891 nsp3, 14408 RdRp, 17746 and 17857 nsp143, 18060 nsp14), S (23403, spike protein) and ORF9a (28881, nucleocapsid phosphoprotein) sequences, respectively.

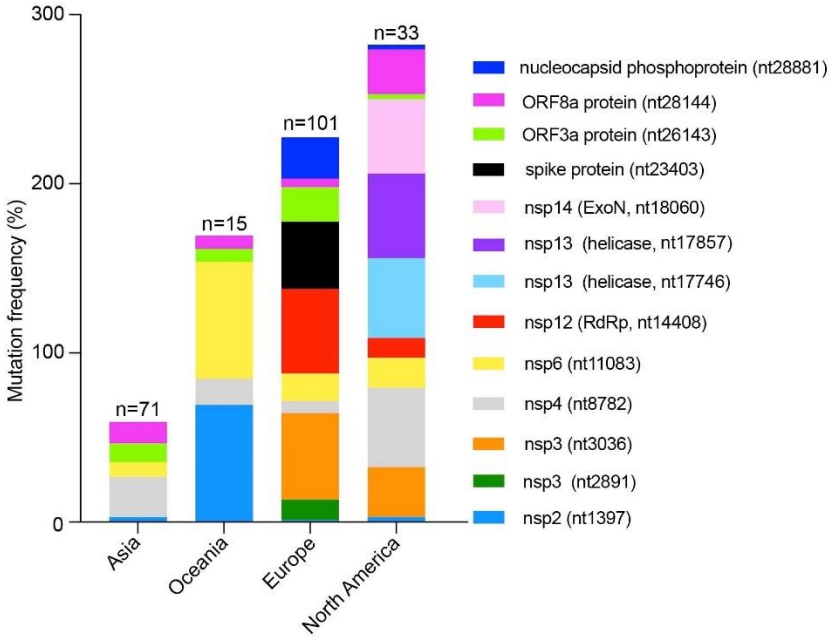


Fig. 2_1: SARS-CoV-2 mutation frequency in different geographic areas. Eight novel recurrent hotspots mutations (namely 1397, 2891, 14408, 17746, 17857, 18060, 23403 and 28881) and 5 hotspots already reported in literature (namely 3036, 8782, 11083, 28144 and 26143) were subdivided into 4 geographic areas: Asia (n = 71), Oceania (n = 15), Europe (n = 101), North America (n = 33). The mutation frequency was estimated for each of them, by normalizing the number of genomes carrying a given mutation in a geographic area, by the overall number of retrieved genomes per geographic area; the graph shows the cumulative mutation frequency of all given mutations present in each geographic area. Mutation locations in viral genes are reported in the legend as well as the proteins (i.e. non-structural protein, nsp) presenting these mutations. The figure shows that genomes from European and North American patients present an increase in mutation frequency compared to Asia. It is also possible to observe that Europe and North America show a differential pattern of mutations: mutation 14408 (red), 23403 (black), 28881 (electric blue) and 26143 (light green) are present mostly in Europe, whereas 18060 (pink), 17857 (purple) and 17746 (light blue) are present mostly in North America.

We found that 3 out of the 12 most frequent mutations (positions 3036, 8782 and 18060) were silent, whereas one mutation (position 11083) was outside the ORF sequence. On the other hand, mutations 1397, 2891, 14408, 17746, 17857, 23403, 26143, 28144 and 28881 resulted in amino acid changes as follows: 1397 (V to I),14408 (P to L), 17746 (P to L), 17857 (C to Y), 23403 (D to G), 26143 (G to V), 28144 (L to S). Mutation located t position 28881 is related to a double codon mutation, inducing the substitution of two amino acids, namely 28881 (R to K) and (G to R). The new amino acid present in 1397 (V to I), 14408 (P to L), 17746 (P to L), 17857 (C to Y), 26143 (G to V) and 28144 (L to S) had a similar isoelectric

point compared to the original amino acid present in the reference protein sequences, with the exception of the mutations at positions 23403 (D to G), 28881 (R to K) and 28881 (G to R), where the mutated amino acid has a significantly different isoelectric point.

Further studies are needed to determine whether these mutations have an impact on proteins' function and structure. We noted that the number and the occurrence of each mutation increase in genomes found out of Asia, reaching a maximum in genomes found in Europe and North America. We also noted that the viral strains found in Europe and North America are derived from the L-"strain" originated in Asia[395].

1.3.2 Characterization of geographically distinct hotspots over time

In order to determine the appearance of each mutation, we analysed each genome from each geographic area over time, by classifying them according to the timing of sample collection, as indicated in the GISAID database. According to this analysis, 6 time subgroups were defined, namely December 2019 (genomes from 5 patients), 1st–15th Jan. 2020 (genomes from 15 patients), 16th–31st Jan. 2020 (genomes from 52 patients), 1st–15th Feb. 2020 (genomes from 13 patients), 16th–29th Feb 2020 (genomes from 55 patients) and 1st–13th Mar 2020 (genomes from 80 patients). The number of mutations (normalized by the population taken into account for each period of time) increases over time during viral spread out of Asia (see Fig. 2_2). No mutations were observed in the Asian genomes analysed in December 2019. Interestingly, a different pattern of mutations was observed in Europe between January and February, when a new mutation, at position 14408, emerged (depicted in red). This mutation is located in the RdRp gene. Also starting from February 2020, the emergence of additional new mutations (i.e. 23403, 28881 and 2891–black, electric blue, dark green, respectively) is observed. Over time, we also noted an increase in the frequency of mutation 3036 (orange), already present in mid-January (2.2%).

Moreover, a different pattern of hotspot mutations is clearly distinguishable in viral genomes detected in North American patients starting from March 2020, when the outbreak of positive cases was reported in the US and Canada. In this group, three novel mutations (17746, 17857 and 18060–light blue, purple and light pink, respectively) were reported. Interestingly, viral genomes present in North American patients carrying RdRp mutation (14%) do not carry any of the European specific mutations.

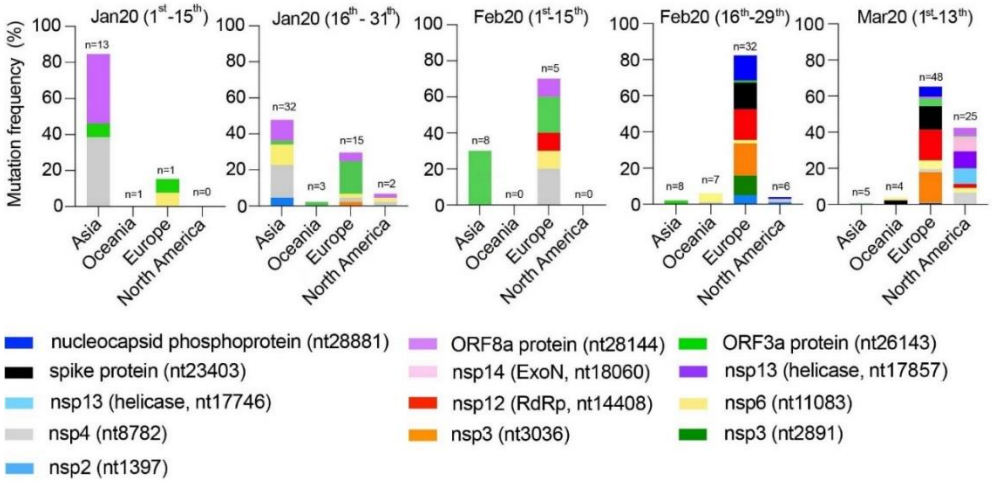


Fig. 2_2: SARS-CoV-2 Mutation occurrence over time divided per geographic area. Eight novel recurrent hotspots mutations (namely 1397, 2891, 14408, 17746, 17857, 18060, 23403 and 28881) and 5 hotspots already reported in literature (namely 3036, 8782, 11083, 28144 and 26143) were subdivided first into 5 period subgroups: December 2019 (n = 5), 1st–15th Jan. 2020 (n = 15), 16th–31st Jan 2020 (n = 52), 1st–15th Feb 2020 (n = 13), 16th–29th Feb 2020 (n = 55) and 1st–13th Mar 2020 (n = 80). Next, for each time group, a further subclassification per geographic area (Asia, Oceania, Europe and North America) was performed (number of genomes in each area are reported in the figure inset). The number of mutations in each area was normalized by the number of genomes analysed for each period of time. This figure shows that mutation frequency increases over time during viral spread out of Asia. No mutations were observed in the Asian genomes analysed in December 2019. In the time group of February 16th–29th, a defined cluster of mutations emerged in Europe; in March 1st–13th, a different cluster of mutations emerged in North America.

1.3.3 Mutations hotspots pattern after February 9th, 2020

Given the importance of RdRp for viability and replication of RNA viruses, mutations in this gene are statistically less likely to occur. However, in some cases, such as in poliovirus, episodes of drug-resistance induced by a point mutation in RdRp have been reported[398]. In our database, the first appearance of a silent RdRp mutation (nt 14804) is manifested on February 9th, 2020 in UK (England), while a different RdRp mutation (nt 14408, amino acid P to L) is observed for the first time in Italy (Lombardy) on February 20th, 2020, when a dramatic increase of the number of European infected patients was reported from the WHO website (https://www.who.int/). We evaluated the increase/decrease of each mutation frequency before and after February 9th, 2020 across the different geographic areas (Fig. 2_3). In particular, we observed a strong increase (+60.5%) of genomes carrying the 14408 mutation affecting RdRp) in Europe, together with an increase of genomes carrying the 3036 mutation (+61.7%), the 23403 mutation (48.1%) and the 28881 mutation (+29.6%) (see upper table Fig.2_3).

	n pts before	n pts after	1397	2891	3036	8782	11083	14408	17746	17857	18060	23403	26143	28144	28881
Europe	17	81	+2.5%	+14.8%	+61.7%	-12.7%	-22.9%	+60.5%	0%	0%	0%	48.1%	-46.5%	-15.2%	29.6%
Asia	59	12	-3.4%	0%	0%	11.3%	-0.14%	0%	0%	0%	0%	0%	+26.6.%	-15.2%	0%
North America	0	28	0%	0%	+35.7%	+42.9%	+17.9%	+14.3%	+57.2%	+60.7%	+53.4%	0%	+3.6%	+21.4%	+3.6%

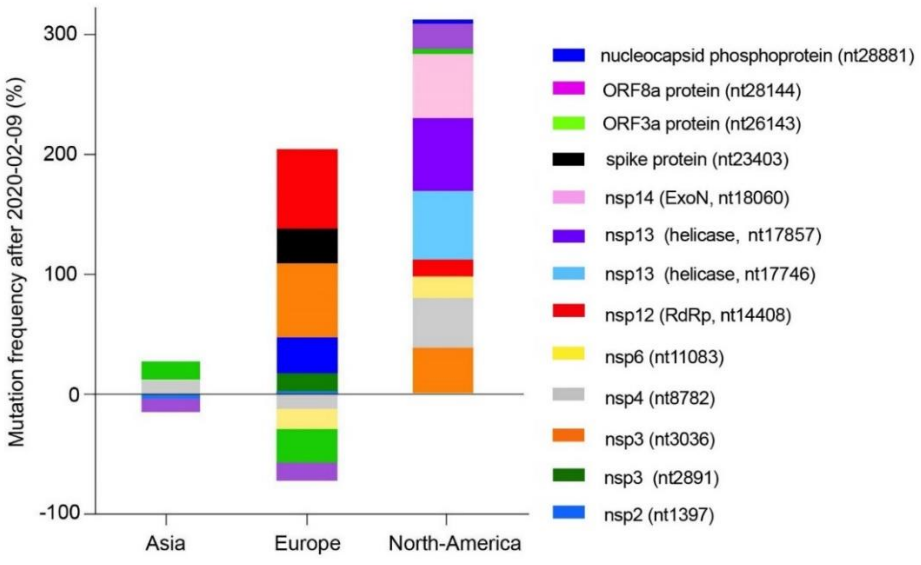


Fig. 2_3: Increment of SARS-CoV-2 mutation frequency after RdRp mutation appearance per geographic area. The increment of mutation frequency before and after February 9th, 2020 across the different geographic areas (Asia, Europe, North America) is shown. The figure shows a diminishment of Asian mutations (i.e. 1397, 8782, 11083, 26143 and 28144) that is simultaneous with the appearance of new mutations such as 2891, 3036, 23403 and 28881, when RdRp novel European mutation located at 14408 (in red) occurred. The upper table shows the increment or decrement for each single mutation, per geographic area.

1.3.4 Simultaneous occurrence of RdRp mutation with other mutations

Next, we analysed genomes collected after February 9th 2020, when mutation in RdRp gene was reported in the database for the first time. For the purpose of analysis, we divided the genomes into two groups: group 1 contains genomes with mutation in position 14408 (RdRp) (n = 53, 4 North America and 49 European), and group 2 without RdRp mutation (n = 84). Genomes in group 1 showed an increased number of mutations compared to group 2. In particular, group 1 shows 6 genomes with two mutations (11.3%), 25 genomes with three mutations (47.2%), 21 genomes with four mutations (39.6%), and 1 genome with 5 mutations (1.9%). In group 1, the most reported mutations are the ones in positions 3036, 14408, 23403 and 28881. Regarding genomes in group 2, 20 do not carry any mutations (23.8%), 25 genomes have a single mutation (29.8%), 19 genomes have two mutations (22.6%), 6 genomes have three mutations (7.1%), 9 genomes have four mutations (10.7%), 2 and 3 genomes have five and six mutations respectively (2.4% and 3.6%). In group 2, the most reported mutations are located at positions 8782, 11083, 17746 and 17857.The distribution between the two groups in terms of number of mutations is statistically relevant (Fisher-Exact test, p value < 0.001). In particular, group 1 and 2 are significantly different in terms of the distribution of genomes having 0, 1, 3 and 4 numbers of mutations (Fisher-Exact test, $p < 0.001$) (Fig. 2_4). This difference, instead, is insignificant when the number of mutations is 2, 5 or 6. We found that viral strains with RdRp mutation have a median of 3 point mutations [range: 2–5], whereas viral strains with no RdRp mutation have a median of 1 mutation [range: 0–3] (p value < 0.001, Mann–Whitney test). The different distribution

between the two groups relative to the number of mutations is statistically significant (Fig. 2_4).

We also analysed the most frequent mutations detected: the ones in positions 3036, 23403 and 28881 (in Europe), and the ones in positions 17746, 17857 and 18060 (in North America). Viral genomes carrying each one of these mutations were compared with viral genomes without mutations, by using Mann–Whitney test for paired-groups comparison analysis. Genomes carrying mutations in positions 3036, 23403, 28881, 17746, 17857 and 18060 show a median of 3–4 mutations (range [2:5]), whereas genomes carrying none of them have a median of 1 or 2 mutations (range [0:3], p-value < 0.001, Mann–Whitney test). This difference is statistically significant and implies that if one of those mutations is present, other mutations are more likely to occur.

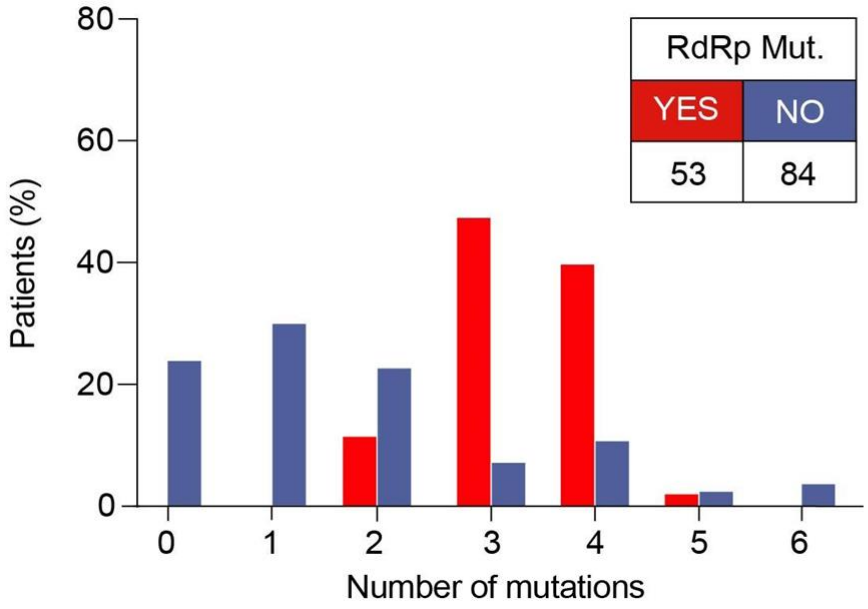


Fig.2_4: Number of SARS-CoV-2 mutations associated with the RdRp mutation. Genomes were subdivided into two groups: group 1 contains genomes with mutation in position 14408 (RdRp) (n = 53, 4 North America and 49 European), and group 2 without RdRp mutation (n = 84). We further subdivided group 1 and 2 by the number of mutations present in the genome. Genomes in group 1 (red bars) showed an increased number of mutations compared to group 2 (grey bars). Most genomes of groups 1 (86.8%) have at least 3 or 4 mutations, whereas 76.2% of genomes of group 2 have less than 2 mutations. We found that viral strains with RdRp mutation have a median of 3 point mutations [range: 2–5], whereas viral strains with no RdRp mutation have a median of 1 mutation [range: 0–3] (p value < 0.001, Mann–Whitney test).

1.3.5 Homology study of mutant RdRp protein

Among all mutation sites analysed, RdRp mutant is particularly interesting given that the enzyme is directly involved in viral replication and its fidelity determines the mutagenic capabilities of SARS-CoV-2. Due to the high homology between RdRps of SARS-CoV and SARS - CoV- 2, we aligned SARS-CoV-2 RdRp reference sequence with the reported catalytic site sequence of SARS-CoV RdRp.

The amino acid substitution 323 (P to L) (due to nucleotide mutation 14408) falls outside the catalytic site, in a region that in SARS-CoV is reported to be an Interface Domain, a still poorly characterized surface structure, supposedly implicated in the interaction with other proteins which may regulate the activity of RdRp[388]. To this regard, it is well-known that SARS-CoV RdRp forms a hollow cylinder-like supercomplex with nsp7 and nsp8, which confer processivity to RdRp[399]. Additionally, replication supercomplex interacts with nsp14, an exonuclease having the Nidovirales-typical proofreading capability. This activity is important in the context of the mutation rate and for controlling the fidelity in RNA replication. However, critical RdRp residues involved in this interaction are still to be identified, and for this reason further studies are needed to assess the possible role of mutation 14408 concerning RdRp fidelity.

1.4 Discussion

In the present work we have compared the SARS-CoV-2 reference genome to those exported from the GISAID database with the aim of gaining important insights into virus mutations, their occurrence over time and within different geographic areas.

We observed that after February 2020, when the first locally transmitted SARS-CoV-2 cases out of Asia were reported, viral genomes presented different point mutations, clearly distinguishable within different geographic areas. Over time, we were able to identify three recurrent mutations in Europe (in positions 3036,

14408 and 23403) and 3 other different mutations in North America (in positions 17746, 17857 and 18060). So far, these mutations have not been detected in Asia. The number and the occurrence, as well as the median value of virus point mutations registered out of Asia, increase over time. In our study, we found that the RdRp mutation, located at position 14408, which is present in European viral genomes starting from February 20th, 2020, is associated with a higher number of point mutations compared to viral genomes from Asia. Given that RdRp works in a complex machinery that includes proofreading activities (in cooperation with other viral cofactors, like ExoN, nsp7 and nsp8), it is tempting to speculate that this mutation has contributed in impairing its proofreading capability. One possible mechanism could involve a minor change in the RdRp structure, without affecting its catalytic activity, that might nonetheless alter its binding capability with other cofactors such as ExoN, nsp7 or nsp8, thus altering the mutation rate. This could explain the increased number of mutations that we observed in Europe since February 2020. Further studies are needed to determine whether the observed cluster mutations originate from the same molecular mechanism. Further studies are also needed to determine whether the mutation in RdRp results in increased viral replication. Some polymerase inhibitors[400,401] are currently being tested in clinical studies to target SARS-CoV-2 RdRp, including Favipiravir[389,391], Galidesivir[390], Remdesivir[391], Ribavirin[392], Penciclovir[402], and Ponatinib[403]. Additionally, other drugs such as Simeprevir (FDA approved HCV protease inhibitor), as well as Filibuvir and Tegobuvir (both RdRp inhibitors)[394], are predicted to bind RdRp by molecular docking studies. In particular, a putative docking site was identified in a hydrophobic cleft very close to the mutated site 323 (P to L), corresponding to mutation 14408 identified in our study[394]. Naturally occurring mutations in RdRp can potentially lead to drug-resistance phenomena, as already observed previously[391,404,405]. Alternatively, it might induce a significant decrease in drug-RdRp complex binding affinity. This could lead to different effectiveness of antiviral treatments where mutation 14408 is present. For this reason, due to the high frequency of RdRp mutation in the infected population, it is important to

characterize the impact of 14408 mutation on the activity of RdRp and its susceptibility to antiviral drugs.

1.5 Conclusions

We identified novel mutation hotspots in the SARS-CoV-2 genome sequences. Interestingly, some appeared after February 2020, only in European patients. Among these hotspots, one mutation in position 14408 is located within the RdRp protein and is associated with an overall increased mutation rate. An in silico analysis comparing annotated functional domains of SARS-CoV and SARS-CoV-2 sequences, showed that this particular mutation occurs in the so-called RdRp interface domain, a still poorly characterized surface structure, involved in protein–protein interactions[388]. The role for the RdRp interface domain requires further investigations, and in particular the effect of mutation in position 14408, its interaction with other cofactors (such as ExoN, nsp7 and nsp8), possibly affecting its proofreading activity and potentially altering its mutation rate. It is also essential to understand if the described mutations could result in the emergence of drug-resistance viral phenotypes. Our data may help the development of diagnostic and therapeutic strategies and the study of potential drug resistance mechanisms.

2. Impact of lockdown on Covid-19 case fatality and viral mutations spread in 7 countries in Europe and North America

2.1 Introduction

Up to April 30th, 2020 there were approximately 3.1 M confirmed cases of COVID-19 worldwide and more than 217.000 infection-related deaths. SARS-CoV and MERS-CoV have caused more than 10,000 cumulative cases in the past two decades, with mortality rates of 9,6% for SARS-CoV and 37% for MERS-CoV, respectively[406–409].

Although SARS-CoV-2 is less lethal than MERS-CoV, up to 20% of the infected people develop rapidly a severe disease characterized by interstitial pneumonia and acute respiratory distress syndrome that can ultimately lead to death. This is particularly reported in elderly and in people with underlying medical conditions. However, most of the patients remain asymptomatic or develop mild symptoms, like fever and dry cough, followed then by breathing difficulties (dyspnea), and bilateral ground-glass opacities on chest CT scans, indicating that the target cells are located in the lower airways[410].

Nowadays, the main goal is to identify an effective treatment and a vaccine against SARS-CoV-2 and to found effective diagnostics, sociological and public health strategies to reduce the spread of the virus, ensuring a faster economic recovery.

This study aims to compare the effectiveness of the different lockdown strategies, the need of a considerable diagnostic PCR testing capability as well as the impact of the representative viral strains isolated in each country

presented here. To this purpose, we focused our study on Italy, Spain, France, Germany, UK, Sweden and United States, broadening our previous analysis of SARS-CoV-2 variants[367]. The characterization of SARS-CoV-2 variants might also significantly contribute to the design of effective therapies, vaccines and novel diagnostics tools.

2.2 Methods

Case fatality rate (CFR) represents the proportion of cases who eventually die from a disease. Once an epidemic has ended, CFR is calculated as (deaths cases/infected cases). However, while an epidemic is still ongoing, as it is the case with the current novel coronavirus outbreak, this formula does not represent the true case fatality rate and might be off by orders of magnitude. Diagnosis of viral infection will precede recovery or deaths by days to weeks and the number of death should therefore be compared to the past case counts – accounting for this delay increasing the estimate of the case fatality rate[411]. To calculate CFR, we used the following formula:

$$\text{CFR} = \text{deaths at day } x \,/\, \text{cases at day } x\text{-}\{T\}$$

(where T = average time period from case confirmation to death)

Therefore, in our study, mortality rate was calculated as the ratio between the death cases due to COVID-19, over the total number of SARS-CoV-2 reported cases 14 days before, according also to the method described[373]. We normalized the mortality rates among different countries, considering the different policies in terms of number of testing/million inhabitants, and at the same time considering the different incidence of the infection taking into account the number of cases/million inhabitants. A corrective country-specific ϱ factor was defined as the ratio between the number of PCR tests/1 million inhabitants and the number of reported cases/1 million inhabitants (data obtained from https://www.worldometers.info/coronavirus/#countries). Standard mortality

rate values were normalized by the country-specific ϱ factor. Case-fatality rates between countries were compared using proportion test. Post-hoc analysis in the case of more than two groups was performed using pairwise comparison of proportions and p-value was adjusted using Holm method. We also analysed 487 full-length genomic sequences of SARS-CoV-2 from GISAID database. Sequenced specimens were collected from December 2019 to April 2020, from the following countries: Germany, Italy, Spain, France, UK, Sweden and USA. NC_045512.2 genome deposited in the GenBank has been used as SARS-CoV-2 reference genome. Muscle and Jalview software were used for genomes alignment and analysis.

2.3 Results and discussion

2.3.1 CFR comparison in different countries

Mortality calculations during the epidemics are difficult, mostly due to calculation biases: during the initial period of the epidemic, many patients were diagnosed with COVID-19 only after developing critical illness or even at the time of death, whereas asymptomatic or paucisymptomatic patients were untested, leading to an underestimation of the denominator[373]. Additional significant biases affect mortality curves: to name a few, the parameters used for death counting, the rigidity of lockdown measures, population age. Over time Countries started adopting better policies for diagnostic PCR testing and lockdown strategies, and consequently the spread of the virus was better monitored and the data were more carefully determined. We chose to analyze the Country-specific data relative to the number of COVID19 deaths in April 2020, when some of the initial biases were likely attenuated, using the method described[373]. The number of deaths of a specific day was divided by the total number of infected cases reported 14 days before. This method considers the fact that 14 days are the average lag time estimated between the first symptoms to death[412]. The data analysed for Italy, France, Germany, Spain, UK, Sweden

and USA are reported in Fig. 2_4a. For all Countries we observed a decrease in the CFR values over time, with the exception of Germany (that maintains a very low value overall) and Sweden (where no decrease is observed). We identified two critical elements that might affect CFR among these Countries: (a) the number of PCR tests made and (b) the total number of positive cases for each Country. Since the second parameter (b) depends on the first parameter (a), we introduced a corrective Country-specific factor ϱ=a/b, that was later used to normalize the CFR previously calculated (Table 2_1). Data obtained through this normalization model are reported in Fig. 2_4b.

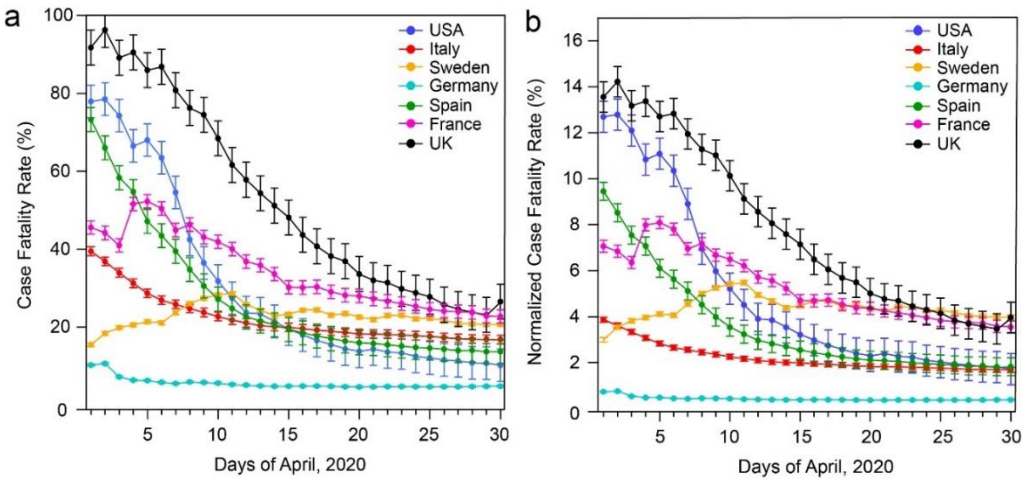


Figure 2_4. (a) calculated case fatality rate curves for USA, Italy, Sweden, Germany, Spain, France and UK as explained by Baud and colleagues[373]. Bars indicate the 95% of confidence interval. (b) case fatality rate of panel (a) normalized by the ϱ factor, i.e. by the number of PCR tests performed per 1M population over positive cases per 1M population up to the 30th of April, 2020. Bars indicate the 95% of confidence interval. The normalization leads to the formation of three main groups: group 1 includes Germany, group 2 includes Italy, USA and Spain and group 3 includes UK, France and Sweden.

Table 2_ 1. Country-specific data showing number of PCR tests and cases per million inhabitants and corrective factor ϱ=a/b.

	Italy	*Spain*	*France*	*Germany*	*UK*	*Sweden*	*USA*
PCR tests/1M inhabitans (a)	36244	41332	16856	30400	19026	11833	22545
Cases/1M inhabitans (b)	3505	5311	2595	1984	2807	2250	3665
Corrective factor (ρ)	10.34	7.78	6.49	15.32	6.78	5.26	6.15

By taking only the data calculated on the 30th of April and representing them in a bubble plot (Fig. 2_5), we clearly identify the presence of three clusters of Countries. Group 1 includes Germany and has a very low normalized CFR

(0.31% CI (95%) [0.29 : 0.33] on April 30th 2020). Group 2 includes Italy, USA and Spain and has an intermediate value of normalized CFR (1.62% CI (95%) [1.51:1.72]; 1.65% CI (95%) [0.97:2.33]; 1.76% CI (95%) [1.36:2.15], respectively, on April 30th 2020). Group 3 includes France, Sweden and UK (3.49% CI (95%) [3.23:3.76]; 3.92% CI (95%) [3.83:4.02]; 3.90% CI (95%) [3.25:4.27], on April 30th 2020). The difference among cluster's CFR (respectively 0.31% vs 1.68% vs 3.78%) was statistically significant (70 y.o.) have a lower lag time[412] compared to others.

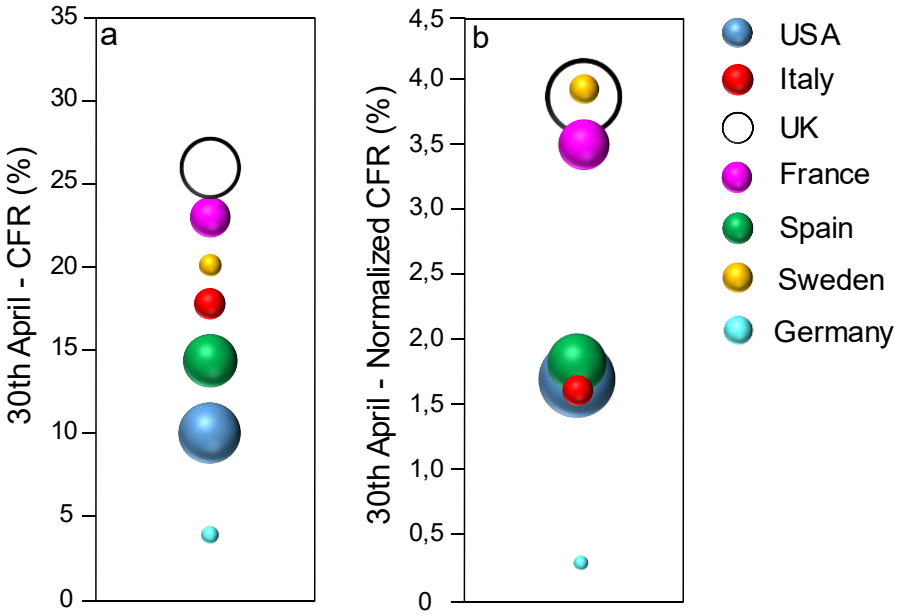


Fig.2_5: Bubble plot representation of the 30th of April Case Fatality Rate (a) and of the 30th of April Normalized Case Fatality Rate of USA (in light blue), Italy (in red), of UK (black line), France (in magenta), Spain (in green), Sweden (in yellow) and Germany (in cyan). In a, the CFRs are distributed within a large range of values, whereas in b the normalized CFRs values are clustered in three well-distinct groups: Germany forms the first group, Italy, Spain and USA the second group and, finally, Sweden, UK and France the third group with the higher normalized CFR value.

However, even if the daily number of death patients divided per age is available for each Country, we could not provide in this study a further normalization of the CFR taking into account patients age, since a similar daily database of infected people divided per age is not publicly available. Anyway, since the infection mostly leads to death older people or those that have ongoing severe illnesses (i.e. cardiovascular diseases, diabetes, cancer), we can speculate that the overall estimation of the CFR is driven by this class of patients. Therefore,

the observed CFR curves observed among different Countries through the introduction of an innovative corrective factor ϱ, might be explained mainly by the different policies that were enacted by each Country. To further support this hypothesis, we note that in Countries of group 3 where lockdown was not put in place (i.e. Sweden) or it was adopted late, and less SARS-CoV-2 PCR tests were executed (i.e. in UK and France), normalized CFR is higher than in the other groups. Although further data are needed to refine the CFR estimation, we improved the CFR estimate by using a new corrective factor which considers two important variables (number of positives and number of PCR tests performed). In fact, several sources of variability affect CFR but for modifiable confounding factors, a standardization process could help to reduce the biases, improving the interpretability and comparability of CFR across Countries.

2.3.2 Lockdown impact on viral mutation spread

A database of 487 genome sequences isolated from patients infected with SARS-CoV-2 in Italy, Spain, Germany, France, UK, Sweden and USA has been randomly collected from the GISAID database, aligned and compared to the SARS-CoV-2 reference genome. A total of 27 genomes were considered in January 2020, 91 in February 2020, 210 in March 2020 and, finally, 159 genomes in April 2020. We analysed 54 genome samples collected in Italy, 61 in Spain, 62 in Germany, 52 in France, 80 in UK, 50 in Sweden and 128 in the United States (Table 2_2).

Table 2_2: Sequenced genomes selection for different geographic areas and time of collection

	Italy	*Spain*	*France*	*Germany*	*UK*	*Sweden*	*USA*	**Tot**
January 2020	3	0	8	1	2	0	13	**27**
February 2020	5	7	11	12	25	2	29	**91**
March 2020	36	35	24	23	27	25	40	**210**
April 2020	10	19	9	26	26	23	46	**159**
Tot	**54**	**61**	**52**	**62**	**80**	**50**	**128**	**487**

We studied the evolution of the mutation patterns in the selected Countries from January to April 2020, and we reported only the recurrent mutations occurring more than 10 times in the time range considered, as described

elsewhere[367]. The occurrence of each mutation in a specific Country has been normalized by the number of genomes collected in that geographic area for each timeframe, dividing the silent by the non-silent mutations (Fig. 2_6). Interestingly, the number of nonsynonymous mutations increases over time during the spread out of Asia, and appears to stabilize in April (Fig. 2_6, top panel) The pattern of nonsynonymous mutations changes quite dramatically from January to February, when such mutations appeared for the first time. More in detail, part of the genomes analysed in January 2020 belong to patients infected in China or to patients in close contact to those travelling or coming back from Asia. In February, most Countries decided to suspend fights at first from and to China and, after, only few communications were maintained between nations and during that month locally transmitted outbreak cases occurred. We observed a pattern of recurrent mutations which reached a homogeneous

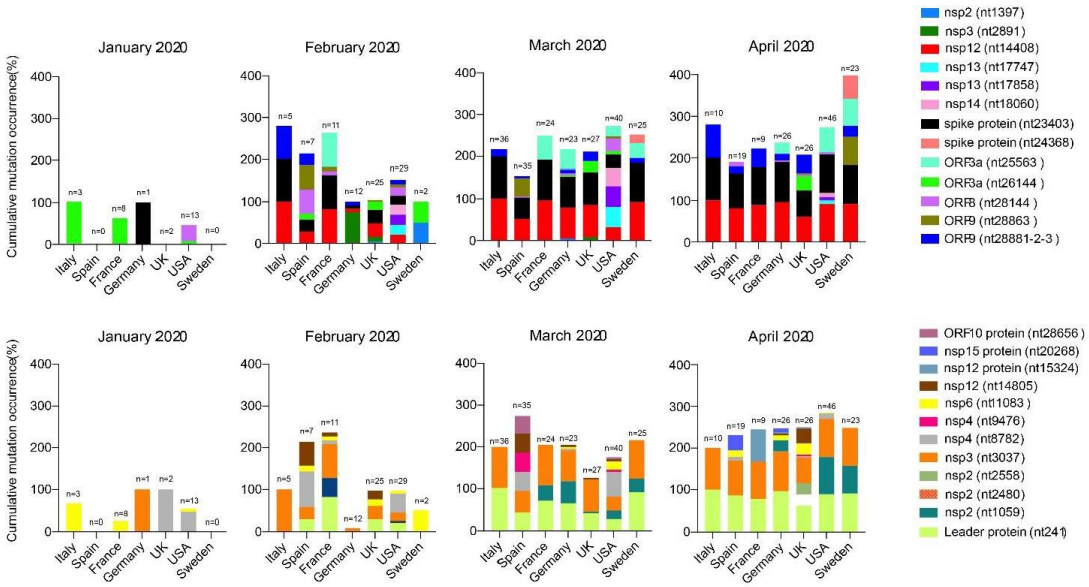


Figure 2_6: SARS-CoV-2 recurrent mutations occurrence over time, divided per geographic area. The graph reports evolution of nonsynonymous (top) and silent (bottom) mutation patterns from January 2020 to April 2020 in Italy, Spain, France, Germany, UK, Sweden and in the USA. The frequency of each mutation in each country and for each month was normalized to 100%, i.e. to the total number of genomes analysed in that frameshift and collected in that specific country. Recurrent nonsynonymous mutation pattern is characterized by the occurrence of mutations at nt14408,nt 23403 and nt28881-28882-28883 (RdRp and spike protein, respectively), while the most found silent mutations are at nt241 and nt3037 (localized in the leader protein and in the nsp3).

distribution across the different Countries in March 2020. This observation is confirmed also in April 2020 in all the analysed Countries. It is likely that lockdown policies implemented in this period greatly reduced further viral spread from Asia and hampered mixing of SARS-CoV-2 strains among Countries. We observed a similar pattern for silent mutations (Fig. 2_6, bottom panel). Overall, our data show a number of silent mutations (nt241, nt3037) and nonsynonymous mutations (nt14408, nt23403 and nt28881–28882–28883) (Fig. 2_6). Among the nonsynonymous mutations, we note the occurrence of an already observed mutation at position 14408, which is located in the viral RNA-dependent RNA-polymerase (RdRp gene), a key component of the replication/transcription machinery[367]. This mutation (Fig. 2_6, depicted in red) emerged in February 2020 and is quite homogeneously distributed across all the Countries analysed. Tis is also observed for a mutation occurring in the spike protein (nt23403, Fig. 2_6, depicted in black) and to a minor extent for a mutation in the nucleocapsid phosphoprotein (nt28881–28882–28883, Fig. 2_6, depicted in blue). The occurrence of the mutation in the RdRp (nt14408) is always associated with that of the spike protein (nt23403), of the nsp3 mutation (nt3037) and of the mutation in the leader protein (nt241). A different pattern of hotspot mutations characterized viral genomes detected in patients from the United States. In February we initially detected three novel mutations (in position 17747, 17858 and 18060), that were not found elsewhere. These mutations were found predominantly in the viral genomes sequenced in Washington State (USA). The occurrence of this isolated pattern over time reflects the viral spreading of a more "European-like" strain (nt241, nt3037, nt14408 and nt23403) in the rest of the US. Overall, the occurrence of this "European-like" group varies from 32.5% of analysed genomes (in USA) to 100% (in Italy). Our data confirm the previous observations made by Korber et al.[413], when the authors hypothesized that this mutation group, associated with the G clade, could enhance viral fitness, possibly due to the nt23403 mutation that triggers a significant amino acid substitution in a strongly

immunogenic linear epitope of Spike protein, which might affect neutralizing antibodies sensitivity.

2.3.3 Emerging of new mutations

We noted the emergence of other recurrent mutation sites over time, both nonsynonymous (nt25563, nt28863) and silent (nt2480, nt2558, nt9476, nt15324, nt20268 and nt28656). The nonsynonymous mutations occur in the ORF3a and ORF9 (nucleocapsid phosphoprotein), causing the amino acid mutation Q56H (glutamine to histidine) and S197L (serine to leucine). All these mutations are found in most Countries and they are not exclusively reported in a specific geographic area. An additional recurrent mutation has been detected exclusively in genomes from Swedish at nt24368 (G to T transition); this mutation, which is located in the spike protein sequence, appeared in March (carried by 20% of genomes analysed) and its frequency more than doubled in April (52% of genomes analysed). Tis mutation triggers an amino acid substitution at position 936, from an aspartic acid to a tyrosine, with a significant shift in terms of isoelectric point from 2.85 to 5.64. D936 residue in SARS-CoV-2 Spike protein corresponds to the E918 residue of the homologue protein of SARS-CoV, and it is located in the heptad repeat 1 (HR1) domain[414,415]. Heptad repeat 1 interacts with heptad repeat 2 (HR2) domain and form a six-helix bundle fusion core, able to bring viral and cellular membranes in close proximity, promoting fusion and infection of host cell[416,417]. Tis makes HR1 and HR2 good target candidates for drug design. Recently, D936 (site of the recurrent mutation) has been proved to bind to R1185 of the heptad repeat 2 (HR2) domain through a salt bridge. Additional studies are required to further characterize if G936 mutant, present in April in more than half of Swedish genomes analysed, could provide some beneficial advantages in terms of viral fitness, as observed for mutation nt23403[413]. Among the Countries in the different groups there are no significant differences in the distribution of mutations, since the recurrent mutation pattern is comparable among different

Countries (Fig. 2_6, top panel). The only significant difference is the newly emerged mutation nt24368, that in our database was detected only in the genomes analysed in Sweden.

2.4 Conclusions

By normalizing the CFR by the ϱ factor, we divided the analysed Countries in three groups with an increased estimated CFR: group 1 is represented by Germany, group 2 by Italy, Spain and USA and group 3 by Sweden, France and UK. Groups 1 and 2 include Countries that adopted strict lockdown strategies and/or have a wide testing capability, whereas group 3 is formed by Countries that have adopted lockdown restrictions later (or have not at all) and/or did not perform an extensive diagnostic PCR testing. A decreasing trend of case fatality rate has been observed among most Countries. There are several direct factors that might contribute to this decline, such as health service's ability to cope with COVID-19 patients, increased and improved viral testing and tracing, efficacy of the different lockdown strategies, herd immunity development, influence of age on the affected population, variation in viral contagiousness and lethality. We observe that, after the rapid emergence and diffusion of recurrent mutations in February and March, a specific mutation pattern has stabilized by April 2020 in all the Countries analysed. Tis pattern is comprised of mutations nt241, nt3037, nt14408 and nt23403. In Sweden we report the occurrence of a unique nonsynonymous mutation in the spike protein (nt24368) which has been found in more than 50% of genomes. The emergence of specific patterns of mutations concomitant with the decline in case fatality rate needs further confirmation and the biological significance of such mutations remains unclear.

List of Figures – Part 2

literature (namely 3036, 8782, 11083, 28144 and 26143) were subdivided first into 5 period subgroups: December 2019 (n = 5), 1st–15th Jan. 2020 (n = 15), 16th–31st Jan 2020 (n = 52), 1st–15th Feb 2020 (n = 13), 16th–29th Feb 2020 (n = 55) and 1st–13th Mar 2020 (n = 80). Next, for each time group, a further subclassification per geographic area (Asia, Oceania, Europe and North America) was performed (number of genomes in each area are reported in the figure inset). The number of mutations in each area was normalized by the number of genomes analysed for each period of time. This figure shows that mutation frequency increases over time during viral spread out of Asia. No mutations were observed in the Asian genomes analysed in December 2019. In the time group of February 16th–29th, a defined cluster of mutations emerged in Europe; in March 1st–13th, a different cluster of mutations emerged in North America.

2_3 Increment of SARS-CoV-2 mutation frequency after RdRp mutation appearance per geographic area. The increment of mutation frequency before and after February 9th, 2020 across the different geographic areas (Asia, Europe, North America) is shown. The figure shows a diminishment of Asian mutations (i.e. 1397, 8782, 11083, 26143 and 28144) that is simultaneous with the appearance of new mutations such as 2891, 3036, 23403 and 28881, when RdRp novel European mutation located at 14408 (in red) occurred. The upper table shows the increment or decrement for each single mutation, per geographic area.

2_4 Number of SARS-CoV-2 mutations associated with the RdRp mutation. Genomes were subdivided into two groups: group 1 contains genomes with mutation in position 14408 (RdRp) (n = 53, 4 North America and 49 European), and group 2 without RdRp mutation (n = 84). We further subdivided group 1 and 2 by the number of mutations present in the genome. Genomes in group 1 (red bars) showed an increased number of mutations compared to group 2 (grey bars). Most genomes of groups 1 (86.8%) have at least 3 or 4 mutations, whereas 76.2% of genomes of group

2 have less than 2 mutations. We found that viral strains with RdRp mutation have a median of 3 point mutations [range: 2–5], whereas viral strains with no RdRp mutation have a median of 1 mutation [range: 0–3] (p value < 0.001, Mann–Whitney test).

2_5 Bubble plot representation of the 30th of April Case Fatality Rate (a) and of the 30th of April Normalized Case Fatality Rate of USA (in light blue), Italy (in red), of UK (black line), France (in magenta), Spain (in green), Sweden (in yellow) and Germany (in cyan). In a, the CFRs are distributed within a large range of values, whereas in b the normalized CFRs values are clustered in three well-distinct groups: Germany forms the first group, Italy, Spain and USA the second group and, finally, Sweden, UK and France the third group with the higher normalized CFR value.

2_6 SARS-CoV-2 recurrent mutations occurrence over time, divided per geographic area. The graph reports evolution of nonsynonymous (top) and silent (bottom) mutation patterns from January 2020 to April 2020 in Italy, Spain, France, Germany, UK, Sweden and in the USA. The frequency of each mutation in each country and for each month was normalized to 100%, i.e. to the total number of genomes analysed in that frameshift and collected in that specific country. Recurrent nonsynonymous mutation pattern is characterized by the occurrence of mutations at nt14408,nt 23403 and nt28881-28882-28883 (RdRp and spike protein, respectively), while the most found silent mutations are at nt241 and nt3037 (localized in the leader protein and in the nsp3).

List of tables – part 2

Bibliography

(1) Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. The Amyloid State and Its Association with Protein Misfolding Diseases. *Nat. Rev. Mol. Cell. Biol.* **2014**, *15* (6), 384–396. https://doi.org/10.1038/nrm3810.

(2) Berrondo, M. Predicting the structure and function of protein mutants, dissertation, **2010**.

(3) Frauenfelder, H. *The Physics of Proteins*; Biological and Medical Physics, Biomedical Engineering; Springer New York: New York, NY, **2010**. https://doi.org/10.1007/978-1-4419-1044-8..

(4) Mustafa, M. I.; Murshed, N. S.; Abdelmoneim, A. H.; Abdelmageed, M. I.; Elfadol, N. M.; Makhawi, A. M. Extensive In Silico Analysis of ATL1 Gene : Discovered Five Mutations That May Cause Hereditary Spastic Paraplegia Type 3A. *Scientifica* 2020, **2020**, 1-13, https://doi.org/10.1155/2020/8329286.

(5) Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The Energy Landscapes and Motions of Proteins. *Science*. **1991**, 254 (5038), 1598–1603. https://doi.org/10.1126/science.1749933..

(6) Chiti, F.; Dobson, C. M. Amyloid Formation by Globular Proteins under Native Conditions. *Nat. Chem. Biol.* **2009**, 5 (1), 15–22. https://doi.org/10.1038/nchembio.131.

(7) Saibil, H. Chaperone Machines for Protein Folding, Unfolding and Disaggregation. *Nat. Rev. Mol. Cell. Biol.* **2013**, *14* (10), 630–642. https://doi.org/10.1038/nrm3658.

(8) Singh, J.; Srivastava, A.; Sharma, P.; Pradhan, P.; Kundu, B. DNA Intercalators as Amyloid Assembly Modulators: Mechanistic Insights. *RSC Adv.* **2017**, *7* (1), 493–506. https://doi.org/10.1039/C6RA26313E.

(9) Armstrong, E. F. Enzymes. *J. Soc. Chem. Ind.* **1930**, 49 (44), 919–920. https://doi.org/10.1002/jctb.5000494433.

(10) Pauling, L. Nature of Forces between Large Molecules of Biological Interest. *Nature* **1948**, *161*, 707–709. https://doi.org/10.1038/161707a0.

(11) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.;

Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically Disordered Protein. *J. Mol. Graph. Model.* **2001**, 19, 26–59. https://doi.org/10.1016/S1093-3263(00)00138-8.

(12) Redfern, O. C.; Dessailly, B.; Orengo, C. A. Exploring the Structure and Function Paradigm. *Curr. Opin. Struct. Biol.* **2008**, *18*, 394–402. https://doi.org/10.1016/j.sbi.2008.05.007.

(13) Comstock, M. J.; Whitley, K. D.; Jia, H.; Sokoloski, J.; Lohman, T. M.; Ha, T.; Chemla, Y. R. Direct Observation of Structure-Function Relationship in a Nucleic Acid–Processing Enzyme. *Science* **2015**, *348*, 352–354. https://doi.org/10.1126/science.aaa0130.

(14) Pegg, S. C.-H.; Brown, S. D.; Ojha, S.; Seffernick, J.; Meng, E. C.; Morris, J. H.; Chang, P. J.; Huang, C. C.; Ferrin, T. E.; Babbitt, P. C. Leveraging Enzyme Structure−Function Relationships for Functional Inference and Experimental Design: The Structure−Function Linkage Database. *Biochemistry* **2006**, *45*, 2545–2555. https://doi.org/10.1021/bi0521011.

(15) Briggs, G. E.; Haldane, J. B. S. A Note on the Kinetics of Enzyme Action. *Biochem. J.* **1925**, *19*, 338–339. https://doi.org/10.1016/j.febslet.2013.07.028

(16) Papoian, G. A. Proteins with Weakly Funneled Energy Landscapes Challenge the Classical Structure–Function Paradigm. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14237–14238. https://doi.org/10.1073/pnas.0807977105.

(17) Tompa, P. The Interplay between Structure and Function in Intrinsically Unstructured Proteins. *FEBS Letters* **2005**, *579*, 3346–3354. https://doi.org/10.1016/j.febslet.2005.03.072.

(18) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901.

(19) Dobson, C. M. Protein Folding and Misfolding. *Nature* **2003**, *426*, 884–890. https://doi.org/10.1038/nature02261.

(20) Walker, L.C.; Diamond, M.I.; Duff, K.E.; Hyman, B.T. Mechanisms of Protein Seeding in Neurodegenerative Diseases | Dementia and Cognitive Impairment

JAMA Neurol. 2013, 70, 1-12. https://doi:10.1001/jamaneurol.2013.1453.

(21) Ono, K.; Takahashi, R.; Ikeda, T.; Yamada, M. Cross-Seeding Effects of Amyloid β-Protein and α-Synuclein. *J. Neurochem.* **2012**, 122, 883–890. https://doi.org/10.1111/j.1471-4159.2012.07847.x.

(22) Moreno-Gonzalez, I.; Edwards I., G.; Salvadores, N.; Shahnawaz, M.; Diaz-Espinoza, R.; Soto, C. Molecular Interaction between Type 2 Diabetes and Alzheimer's Disease through Cross-Seeding of Protein Misfolding. *Mol. Psychiatry* **2017**, *22*, 1327–1334. https://doi.org/10.1038/mp.2016.230.

(23) Morales, R.; Moreno-Gonzalez, I.; Soto, C. Cross-Seeding of Misfolded Proteins: Implications for Etiology and Pathogenesis of Protein Misfolding Diseases. *PLoS Pathogens* **2013**, *9*, e1003537. https://doi.org/10.1371/journal.ppat.1003537.

(24) Amm, I.; Sommer, T.; Wolf, D. H. Protein Quality Control and Elimination of Protein Waste: The Role of the Ubiquitin–Proteasome System. *Biochim. Biophys. Acta* **2014**, *1843* , 182–196. https://doi.org/10.1016/j.bbamcr.2013.06.031.

(25) Goldberg, A. L. Protein Degradation and Protection against Misfolded or Damaged Proteins. *Nature* **2003**, *426*, 895–899. https://doi.org/10.1038/nature02263.

(26) Hipp, M.S; Kasturi, P; Hartl, F.U. The proteostasis network and its decline in ageing. *Nat. Rev. Mol. Cell. Biol.* **2019**, 20, 421–435. https://doi.org/10.1038/s41580-019-0101-y.

(27) Iadanza, M. G.; Jackson, M. P.; Hewitt, E. W.; Ranson, N. A.; Radford, S. E. A New Era for Understanding Amyloid Structures and Disease. *Nat. Rev. Mol. Cell. Biol.* **2018**, *19*, 755–773. https://doi.org/10.1038/s41580-018-0060-8.

(28) Sipe, J. D.; Cohen, A. S., History of the Amyloid Fibril. *J. Struct. Biol.* **2000**, *130*, 88–98. https://doi.org/10.1006/jsbi.2000.4221.

(29) Cohen, A. S. General Introduction and a Brief History of Amyloidosis. In *Amyloidosis* Springer Netherlands, **1986**, 3–19. https://doi.org/10.1007/978-94-009-4309-4_1.

(30) Alzheimer, A. Über Eine Eigenartige Erkrankung Der Hirnrinde. *Neurol. Central* **1906**, 25, 1134.

(31) Alzheimer, A. Über Einen Eigenartigen Schweren Erkrankungsprozeβ Der Hirnrincle. *Allg. Z. Psychiatr. Psych.-Gerichtl. Med* **1907**, *641*, 46–48.

(32) Cohen, A. S.; Calkins, E. Electron Microscopic Observations on a Fibrous Component in Amyloid of Diverse Origins. *Nature* **1959**, *183*, 1202–1203. https://doi.org/10.1038/1831202a0.

(33) Geddes, A. J.; Parker, K. D.; Atkins, E. D. T.; Beighton, E. "Cross-β" Conformation in Proteins. *J. Mol. Biol.* **1968**, *32*, 343–358. https://doi.org/10.1016/0022-2836(68)90014-4.

(34) Shirahama, T.; Cohen, A. S. High-resolution electron microscopic analysis of the amyloid fibril. *J. Cell. Biol.* **1967**, 33 (3), 679–708.http://doi.org/10.1083/jcb.33.3.679

(35) Merlini, G.; Bellotti V. Molecular Mechanisms of Amyloidosis. *New. Engl. J. Med.* **2003**, 14.https://doi.org/ 10.1056/NEJMra023144

(36) Prusiner, S. B.; McKinley, M. P.; Bowman, K. A.; Bolton, D. C.; Bendheim, P. E.; Groth, D. F.; Glenner, G. G. Scrapie Prions Aggregate to Form Amyloid-like Birefringent Rods. *Cell* **1983**, *35*, 349–358. https://doi.org/10.1016/0092-8674(83)90168-X.

(37) Jahn, T. R.; Makin, O. S.; Morris, K. L.; Marshall, K. E.; Tian, P.; Sikorski, P.; Serpell, L. C. The Common Architecture of Cross-β Amyloid. *J. Mol. Biol.* **2010**, *395* (4), 717–727. https://doi.org/10.1016/j.jmb.2009.09.039.

(38) Eisenberg, D.; Jucker, M. The Amyloid State of Proteins in Human Diseases. *Cell* **2012**, *148* , 1188–1203. https://doi.org/10.1016/j.cell.2012.02.022.

(39) Westermark, P.; Benson, M. D.; Buxbaum, J. N.; Cohen, A. S.; Frangione, B.; Ikeda, S.-I.; Masters, C. L.; Merlini, G.; Saraiva, M. J.; Sipe, J. D. Amyloid: Toward Terminology Clarification Report from the Nomenclature Committee of the International Society of Amyloidosis. *Amyloid* **2005**, *12*, 1–4. https://doi.org/10.1080/13506120500032196.

(40) Gosal, W. S.; Clark, A. H.; Ross-Murphy, S. B. Fibrillar β-Lactoglobulin Gels: Part 1. Fibril Formation and Structure. *Biomacromolecules* **2004**, *5* (6), 2408–2419. https://doi.org/10.1021/bm049659d.

(41) Sidhu, A.; Segers-Nolten, I.; Raussens, V.; Claessens, M. M. A. E.; Subramaniam, V. Distinct Mechanisms Determine α-Synuclein Fibril Morphology during

Growth and Maturation. *ACS Chem. Neurosci.* **2017**, *8* (3), 538–547. https://doi.org/10.1021/acschemneuro.6b00287.

(42) Eichner, T.; Radford, S. E. A Diversity of Assembly Mechanisms of a Generic Amyloid Fold. *Mol. Cell* **2011**, *43* (1), 8–18. https://doi.org/10.1016/j.molcel.2011.05.012.

(43) Iadanza, M. G.; Silvers, R.; Boardman, J.; Smith, H. I.; Karamanos, T. K.; Debelouchina, G. T.; Su, Y.; Griffin, R. G.; Ranson, N. A.; Radford, S. E. The Structure of a β2-Microglobulin Fibril Suggests a Molecular Basis for Its Amyloid Polymorphism. *Nat. Commun.* **2018**, *9* (1), 4517. https://doi.org/10.1038/s41467-018-06761-6.

(44) Watt, B.; van Niel, G.; Raposo, G.; Marks, M. S. PMEL: A Pigment Cell-Specific Model for Functional Amyloid Formation. *Pigment Cell Melanoma Res.* **2013**, *26* (3), 300–315. https://doi.org/10.1111/pcmr.12067.

(45) Dobson, C. M.; Dobson, C. M. Protein Misfolding, Evolution and Disease. *Trends Biochem. Sci.* **1999**, *24* (9), 329–332. https://doi.org/10.1016/S0968-0004(99)01445-0.

(46) Krebs, M. R. H.; Wilkins, D. K.; Chung, E. W.; Pitkeathly, M. C.; Chamberlain, A. K.; Zurdo, J.; Robinson, C. V.; Dobson, C. M. Formation and Seeding of Amyloid Fibrils from Wild-Type Hen Lysozyme and a Peptide Fragment from the β-Domain. *J. Mol. Biol.* **2000**, *300* (3), 541–549. https://doi.org/10.1006/jmbi.2000.3862.

(47) Xing, Y.; Higuchi, K. Amyloid Fibril Proteins. *Mech. Ageing Dev.* **2002**, *123* (12), 1625–1636. https://doi.org/10.1016/S0047-6374(02)00098-2.

(48) Morrissey, M. P.; Shakhnovich, E. I. Evidence for the Role of PrPC Helix 1 in the Hydrophilic Seeding of Prion Aggregates. *Proc. Natl. Acad. Sci. USA*, **1999**, *96* (20), 11293–11298. https://doi.org/10.1073/pnas.96.20.11293.

(49) Jarrett, J. T.; Lansbury, P. T. Seeding "One-Dimensional Crystallization" of Amyloid: A Pathogenic Mechanism in Alzheimer's Disease and Scrapie? *Cell* **1993**, *73* (6), 1055–1058. https://doi.org/10.1016/0092-8674(93)90635-4.

(50) Perez, C.; Miti, T.; Hasecke, F.; Meisl, G.; Hoyer, W.; Muschol, M.; Ullah, G. Mechanism of Fibril and Soluble Oligomer Formation in Amyloid Beta and Hen

Egg White Lysozyme Proteins. *J. Phys. Chem. B* **2019**, *123* (27), 5678–5689. https://doi.org/10.1021/acs.jpcb.9b02338.

(51) Bertoncini, C. W.; Fernandez, C. O.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. Familial Mutants of α-Synuclein with Increased Neurotoxicity Have a Destabilized Conformation. *J. Biol. Chem.* **2005**, *280* (35), 30649–30652. https://doi.org/10.1074/jbc.C500288200.

(52) Hoyer, W.; Cherny, D.; Subramaniam, V.; Jovin, T. M. Impact of the Acidic C-Terminal Region Comprising Amino Acids 109−140 on α-Synuclein Aggregation in Vitro. *Biochemistry* **2004**, *43* (51), 16233–16242. https://doi.org/10.1021/bi048453u.

(53) Cohen, S. I. A.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. From Macroscopic Measurements to Microscopic Mechanisms of Protein Aggregation. *J. Mol. Bio.* **2012**, *421* (2), 160–171. https://doi.org/10.1016/j.jmb.2012.02.031.

(54) Knowles, T. P. J.; Waudby, C. A.; Devlin, G. L.; Cohen, S. I. A.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **2009**, *326* (5959), 1533–1537. https://doi.org/10.1126/science.1178250.

(55) Breydo, L.; Wu, J. W.; Uversky, V. N. α-Synuclein Misfolding and Parkinson's Disease. *Biochim. Biophys. Acta* **2012**, *1822* (2), 261–285. https://doi.org/10.1016/j.bbadis.2011.10.002.

(56) Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan, W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M. Y.; George, J. M.; Rienstra, C. M. Solid-State NMR Structure of a Pathogenic Fibril of Full-Length Human α-Synuclein. *Nat. Struct. Mol. Biol.* **2016**, *23* (5), 409–415. https://doi.org/10.1038/nsmb.3194.

(57) Li, Y.; Zhao, C.; Luo, F.; Liu, Z.; Gui, X.; Luo, Z.; Zhang, X.; Li, D.; Liu, C.; Li, X. Amyloid Fibril Structure of α-Synuclein Determined by Cryo-Electron Microscopy. *Cell Res.* **2018**, *28* (9), 897–903. https://doi.org/10.1038/s41422-018-0075-x.

(58) Chen, G.; Xu, T.; Yan, Y.; Zhou, Y.; Jiang, Y.; Melcher, K.; Xu, H. E. Amyloid Beta: Structure, Biology and Structure-Based Therapeutic Development. *Acta Pharmacol. Sin.* **2017**, *38* (9), 1205–1235. https://doi.org/10.1038/aps.2017.28.

(59) Fan, H.-C.; Ho, L.-I.; Chi, C.-S.; Chen, S.-J.; Peng, G.-S.; Chan, T.-M.; Lin, S.-Z.; Harn, H.-J. Polyglutamine (PolyQ) Diseases: Genetics to Treatments. *Cell Transplant.* **2014**, *23* (4–5), 441–458. https://doi.org/10.3727/096368914X678454.

(60) Renton, A. E.; Majounie, E.; Waite, A.; Simón-Sánchez, J.; Rollinson, S.; Gibbs, J. R.; Schymick, J. C.; Laaksovirta, H.; van Swieten, J. C.; Myllykangas, L.; Kalimo, H.; Paetau, A.; Abramzon, Y.; Remes, A. M.; Kaganovich, A.; Scholz, S. W.; Duckworth, J.; Ding, J.; Harmer, D. W.; Hernandez, D. G.; Johnson, J. O.; Mok, K.; Ryten, M.; Trabzuni, D.; Guerreiro, R. J.; Orrell, R. W.; Neal, J.; Murray, A.; Pearson, J.; Jansen, I. E.; Sondervan, D.; Seelaar, H.; Blake, D.; Young, K.; Halliwell, N.; Callister, J. B.; Toulson, G.; Richardson, A.; Gerhard, A.; Snowden, J.; Mann, D.; Neary, D.; Nalls, M. A.; Peuralinna, T.; Jansson, L.; Isoviita, V.-M.; Kaivorinne, A.-L.; Hölttä-Vuori, M.; Ikonen, E.; Sulkava, R.; Benatar, M.; Wuu, J.; Chiò, A.; Restagno, G.; Borghero, G.; Sabatelli, M.; Heckerman, D.; Rogaeva, E.; Zinman, L.; Rothstein, J. D.; Sendtner, M.; Drepper, C.; Eichler, E. E.; Alkan, C.; Abdullaev, Z.; Pack, S. D.; Dutra, A.; Pak, E.; Hardy, J.; Singleton, A.; Williams, N. M.; Heutink, P.; Pickering-Brown, S.; Morris, H. R.; Tienari, P. J.; Traynor, B. J. A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron* **2011**, *72* (2), 257–268. https://doi.org/10.1016/j.neuron.2011.09.010.

(61) DeJesus-Hernandez, M.; Mackenzie, I. R.; Boeve, B. F.; Boxer, A. L.; Baker, M.; Rutherford, N. J.; Nicholson, A. M.; Finch, N. A.; Gilmer, H. F.; Adamson, J.; Kouri, N.; Wojtas, A.; Sengdy, P.; Hsiung, G.-Y. R.; Karydas, A.; Seeley, W. W.; Josephs, K. A.; Coppola, G.; Geschwind, D. H.; Wszolek, Z. K.; Feldman, H.; Knopman, D.; Petersen, R.; Miller, B. L.; Dickson, D.; Boylan, K.; Graff-Radford, N.; Rademakers, R. Expanded GGGGCC Hexanucleotide Repeat in Non-Coding Region of C9ORF72 Causes Chromosome 9p-Linked Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron* **2011**, *72* (2), 245–256. https://doi.org/10.1016/j.neuron.2011.09.011.

(62) Andresen, J. M.; Gayán, J.; Djoussé, L.; Roberts, S.; Brocklebank, D.; Cherny, S. S.; Cardon, L. R.; Gusella, J. F.; MacDonald, M. E.; Myers, R. H.; Housman, D. E.; Wexler, N. S. The Relationship Between CAG Repeat Length and Age of Onset Differs for Huntington's Disease Patients with Juvenile Onset or Adult Onset. *Ann. Hum. Gen.* **2007**, *71* (3), 295–301. https://doi.org/10.1111/j.1469-1809.2006.00335.x.

(63) Vasconcellos, L. R. C.; Dutra, F. F.; Siqueira, M. S.; Paula-Neto, H. A.; Dahan, J.; Kiarely, E.; Carneiro, L. A. M.; Bozza, M. T.; Travassos, L. H. Protein Aggregation as a Cellular Response to Oxidative Stress Induced by Heme and Iron. *Proc. Natl. Acad. Sci. USA* **2016**. https://doi.org/10.1073/pnas.1608928113.

(64) Stefani, M.; Rigacci, S. Protein Folding and Aggregation into Amyloid: The Interference by Natural Phenolic Compounds. *Int. J. Mol. Sci.* **2013**, *14* (6), 12411–12457. https://doi.org/10.3390/ijms140612411.

(65) Simpson, D. S. A.; Oliver, P. L. ROS Generation in Microglia: Understanding Oxidative Stress and Inflammation in Neurodegenerative Disease. *Antioxidants* **2020**, *9* (8), 743. https://doi.org/10.3390/antiox9080743.

(66) Stoppini, M.; Andreola, A.; Foresti, G.; Bellotti, V. Neurodegenerative Diseases Caused by Protein Aggregation: A Phenomenon at the Borderline between Molecular Evolution and Ageing. *Pharmacol. Res.* **2004**, *50* (4), 419–431. https://doi.org/10.1016/j.phrs.2003.12.025.

(67) Rogers, I.; Kerr, F; Martinez, P.; Hardy, J; Lovestone, S.; Partridge, L; Ageing Increases Vulnerability to Aβ42 Toxicity in Drosophila. *PLoS One* **2012**, *7*, e40569.https://doi.org/ 10.1371/journal.pone.0040569.

(68) Fusco, G.; Chen, S. W.; Williamson, P. T. F.; Cascella, R.; Perni, M.; Jarvis, J. A.; Cecchi, C.; Vendruscolo, M.; Chiti, F.; Cremades, N.; Ying, L.; Dobson, C. M.; De Simone, A. Structural Basis of Membrane Disruption and Cellular Toxicity by α-Synuclein Oligomers. *Science* **2017**, *358* (6369), 1440–1443. https://doi.org/10.1126/science.aan6160.

(69) Strohäker, T.; Jung, B. C.; Liou, S.-H.; Fernandez, C. O.; Riedel, D.; Becker, S.; Halliday, G. M.; Bennati, M.; Kim, W. S.; Lee, S.-J.; Zweckstetter, M. Structural

Heterogeneity of α-Synuclein Fibrils Amplified from Patient Brain Extracts. *Nat. Commun.* **2019**, *10* (1), 5535. https://doi.org/10.1038/s41467-019-13564-w.

(70) Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S.; Eisenberg, D. S.; Zhou, Z. H.; Jiang, L. Cryo-EM of Full-Length α-Synuclein Reveals Fibril Polymorphs with a Common Structural Kernel. *Nat. Commun.* **2018**, *9* (1), 3609. https://doi.org/10.1038/s41467-018-05971-2.

(71) Watt, J. A.; Pike, C. J.; Walencewicz-Wasserman, A. J.; Cotman, C. W. Ultrastructural Analysis of β-Amyloid-Induced Apoptosis in Cultured Hippocampal Neurons. *Brain Res.* **1994**, *661* (1), 147–156. https://doi.org/10.1016/0006-8993(94)91191-6.

(72) Bucciantini, M.; Rigacci, S.; Berti, A.; Pieri, L.; Cecchi, C.; Nosi, D.; Formigli, L.; Chiti, F.; Stefani, M. Patterns of Cell Death Triggered in Two Different Cell Lines by HypF-N Prefibrillar Aggregates. *FASEB J.* **2005**, *19* (3), 1–23. https://doi.org/10.1096/fj.04-3086fje.

(73) Ross, C. A. Polyglutamine Pathogenesis: Emergence of Unifying Mechanisms for Huntington's Disease and Related Disorders. *Neuron* **2002**, *35* (5), 819–822. https://doi.org/10.1016/S0896-6273(02)00872-3.

(74) Lane, N. J.; Balbo, A.; Fukuyama, R.; Rapoport, S. I.; Galdzicki, Z. The Ultrastructural Effects of β-Amyloid Peptide on Cultured PC12 Cells: Changes in Cytoplasmic and Intramembranous Features. *J. Neurocytol.* **1998**, *27* (10), 707–718. https://doi.org/10.1023/A:1006976400227.

(75) Morishima, Y.; Gotoh, Y.; Zieg, J.; Barrett, T.; Takano, H.; Flavell, R.; Davis, R. J.; Shirasaki, Y.; Greenberg, M. E. β-Amyloid Induces Neuronal Apoptosis Via a Mechanism That Involves the c-Jun N-Terminal Kinase Pathway and the Induction of Fas Ligand. *J. Neurosci.* **2001**, *21* (19), 7551–7560. https://doi.org/10.1523/JNEUROSCI.21-19-07551.2001.

(76) Velez-Pardo, C.; Arroyave, S. T.; Lopera, F.; Castaño, A. D.; Jimenez Del Rio, M. Ultrastructure Evidence of Necrotic Neural Cell Death in Familial Alzheimer's Disease Brains Bearing Presenilin-1 E280A Mutation. *J. Alzheimers Dis.* **2001**, *3* (4), 409–415. https://doi.org/10.3233/jad-2001-3408.

(77) Hipp, M. S.; Kasturi, P.; Hartl, F. U. The Proteostasis Network and Its Decline in Ageing. *Nat. Rev. Mol. Cell. Biol.* **2019**, *20* (7), 421–435. https://doi.org/10.1038/s41580-019-0101-y.

(78) Rooijen, B. D. van; Claessens, M. M. A. E.; Subramaniam, V. Membrane Permeabilization by Oligomeric α-Synuclein: In Search of the Mechanism. *PLoS ONE* **2010**, *5* (12), e14292. https://doi.org/10.1371/journal.pone.0014292.

(79) Campioni, S.; Mannini, B.; Zampagni, M.; Pensalfini, A.; Parrini, C.; Evangelisti, E.; Relini, A.; Stefani, M.; Dobson, C. M.; Cecchi, C.; Chiti, F. A Causative Link between the Structure of Aberrant Protein Oligomers and Their Toxicity. *Nat. Chem. Biol.* **2010**, *6* (2), 140–147. https://doi.org/10.1038/nchembio.283.

(80) Lee, J.-E.; Sang, J. C.; Rodrigues, M.; Carr, A. R.; Horrocks, M. H.; De, S.; Bongiovanni, M. N.; Flagmeier, P.; Dobson, C. M.; Wales, D. J.; Lee, S. F.; Klenerman, D. Mapping Surface Hydrophobicity of α-Synuclein Oligomers at the Nanoscale. *Nano Lett.* **2018**, *18* (12), 7494–7501. https://doi.org/10.1021/acs.nanolett.8b02916.

(81) Mannini, B.; Mulvihill, E.; Sgromo, C.; Cascella, R.; Khodarahmi, R.; Ramazzotti, M.; Dobson, C. M.; Cecchi, C.; Chiti, F. Toxicity of Protein Oligomers Is Rationalized by a Function Combining Size and Surface Hydrophobicity. *ACS Chem. Biol.* **2014**, *9* (10), 2309–2317. https://doi.org/10.1021/cb500505m.

(82) Farrugia, M. Y.; Caruana, M.; Ghio, S.; Camilleri, A.; Farrugia, C.; Cauchi, R. J.; Cappelli, S.; Chiti, F.; Vassallo, N. Toxic Oligomers of the Amyloidogenic HypF-N Protein Form Pores in Mitochondrial Membranes. *Sci. Rep.* **2020**, *10* (1), 17733. https://doi.org/10.1038/s41598-020-74841-z.

(83) Olzscha, H.; Schermann, S. M.; Woerner, A. C.; Pinkert, S.; Hecht, M. H.; Tartaglia, G. G.; Vendruscolo, M.; Hayer-Hartl, M.; Hartl, F. U.; Vabulas, R. M. Amyloid-like Aggregates Sequester Numerous Metastable Proteins with Essential Cellular Functions. *Cell* **2011**, *144* (1), 67–78. https://doi.org/10.1016/j.cell.2010.11.050.

(84) Darios, F.; Ruipérez, V.; López, I.; Villanueva, J.; Gutierrez, L. M.; Davletov, B. α-Synuclein Sequesters Arachidonic Acid to Modulate SNARE-Mediated

Exocytosis. *EMBO reports* **2010**, *11* (7), 528–533. https://doi.org/10.1038/embor.2010.66.

(85) Yasumoto, T.; Takamura, Y.; Tsuji, M.; Watanabe-Nakayama, T.; Imamura, K.; Inoue, H.; Nakamura, S.; Inoue, T.; Kimura, A.; Yano, S.; Nishijo, H.; Kiuchi, Y.; Teplow, D. B.; Ono, K. High Molecular Weight Amyloid B1-42 Oligomers Induce Neurotoxicity via Plasma Membrane Damage. *FASEB J.* **2019**, *33* (8), 9220–9234. https://doi.org/10.1096/fj.201900604R.

(86) Pepys, M. B. Pathogenesis, Diagnosis and Treatment of Systemic Amyloidosis. *Phil. Trans. R. Soc. Lond. B* **2001**, *356* (1406), 203–211. https://doi.org/10.1098/rstb.2000.0766.

(87) Hawkins, P. N.; Lavender, J. P.; Pepys, M. B. Evaluation of Systemic Amyloidosis by Scintigraphy with 123I-Labeled Serum Amyloid P Component. *New Engl. J. Med.* **1990**, *323* (8), 508–513. https://doi.org/10.1056/NEJM199008233230803.

(88) Dubrey, S. W.; Bilazarian, S.; LaValley, M.; Reisinger, J.; Skinner, M.; Falk, R. H. Signal-Averaged Electrocardiography in Patients with AL (Primary) Amyloidosis. *Am. Heart J.* **1997**, *134* (6), 994–1001. https://doi.org/10.1016/S0002-8703(97)70017-6.

(89) Angelini, A.; Zanco, F.; Castellani, C.; Francesco, A. D.; Barbera, M. D.; Vescovo, G. M.; Berno, T.; Fedrigo, M. Cardiac Amyloidosis: A Review of the Literature and a Practical Approach for the Clinicians. *Ital. J. Med.* **2019**, *13* (2), 73–90. https://doi.org/10.4081/itjm.2019.1149.

(90) Janelidze, S.; Pannee,J.; Mikulskis, A.; Chiao, P.; Zetterberg, H; Blennow, K.; Hansson, O.; Concordance Between Different Amyloid Immunoassays and Visual Amyloid Positron Emission Tomographic Assessment. *JAMA Neurol.* **2017,** *74*(12), 1492-1501. https://doi.org/10.1001/jamaneurol.2017.2814.

(91) Dietemann, S.; Nkoulou, R. Amyloid PET Imaging in Cardiac Amyloidosis: A Pilot Study Using 18F-Flutemetamol Positron Emission Tomography. *Ann. Nucl. Med.* **2019**, *33* (8), 624–628. https://doi.org/10.1007/s12149-019-01372-7.

(92) Eisele, Y. S.; Monteiro, C.; Fearns, C.; Encalada, S. E.; Wiseman, R. L.; Powers, E. T.; Kelly, J. W. Targeting Protein Aggregation for the Treatment of Degenerative

Diseases. *Nat. Rev. Drug Disc.* **2015**, *14* (11), 759–780. https://doi.org/10.1038/nrd4593.

(93) Bieschke, J.; Russ, J.; Friedrich, R. P.; Ehrnhoefer, D. E.; Wobst, H.; Neugebauer, K.; Wanker, E. E. EGCG Remodels Mature α-Synuclein and Amyloid-β Fibrils and Reduces Cellular Toxicity. *Proc. Natl. Acad. Sci. USA* **2010**, *107* (17), 7710–7715. https://doi.org/10.1073/pnas.0910723107.

(94) Trichopoulou, A.; Costacou, T.; Bamia, C.; Trichopoulos, D. Adherence to a Mediterranean Diet and Survival in a Greek Population. *New Engl. J. Med.* **2003**, *348* (26), 2599–2608. https://doi.org/10.1056/NEJMoa025039.

(95) Reddy, V. P.; Aryal, P.; Robinson, S.; Rafiu, R.; Obrenovich, M.; Perry, G. Polyphenols in Alzheimer's Disease and in the Gut–Brain Axis. *Microorganisms* **2020**, *8* (2), 199. https://doi.org/10.3390/microorganisms8020199.

(96) Sathya, S.; Pandima Devi, K. Chapter 15 - The Use of Polyphenols for the Treatment of Alzheimer's Disease. In *Role of the Mediterranean Diet in the Brain and Neurodegenerative Diseases*; Academic Press, **2018**, 239–252. https://doi.org/10.1016/B978-0-12-811959-4.00015-8.

(97) Lamberto, G. R.; Binolfi, A.; Orcellet, M. L.; Bertoncini, C. W.; Zweckstetter, M.; Griesinger, C.; Fernández, C. O. Structural and Mechanistic Basis behind the Inhibitory Interaction of PcTS on α-Synuclein Amyloid Fibril Formation. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (50), 21057–21062. https://doi.org/10.1073/pnas.0902603106.

(98) Porat, Y.; Abramowitz, A.; Gazit, E. Inhibition of Amyloid Fibril Formation by Polyphenols: Structural Similarity and Aromatic Interactions as a Common Inhibition Mechanism. *Chem. Biol. Drug Design* **2006**, *67* (1), 27–37. https://doi.org/10.1111/j.1747-0285.2005.00318.x.

(99) Velander, P.; Wu, L.; Henderson, F.; Zhang, S.; Bevan, D. R.; Xu, B. Natural Product-Based Amyloid Inhibitors. *Biochem. Pharmacol.* **2017**, *139*, 40–55. https://doi.org/10.1016/j.bcp.2017.04.004.

(100) Hora, M.; Carballo-Pacheco, M.; Weber, B.; Morris, V. K.; Wittkopf, A.; Buchner, J.; Strodel, B.; Reif, B. Epigallocatechin-3-Gallate Preferentially Induces

Aggregation of Amyloidogenic Immunoglobulin Light Chains. *Sci. Rep.* **2017**, *7* (1), 41515. https://doi.org/10.1038/srep41515.

(101) Ueda, M.; Okada, M.; Mizuguchi, M.; Kluve-Beckerman, B.; Kanenawa, K.; Isoguchi, A.; Misumi, Y.; Tasaki, M.; Ueda, A.; Kanai, A.; Sasaki, R.; Masuda, T.; Inoue, Y.; Nomura, T.; Shinriki, S.; Shuto, T.; Kai, H.; Yamashita, T.; Matsui, H.; Benson, M. D.; Ando, Y. A Cell-Based High-Throughput Screening Method to Directly Examine Transthyretin Amyloid Fibril Formation at Neutral PH. *J. Biol. Chem.* **2019**, *294* (29), 11259–11275. https://doi.org/10.1074/jbc.RA119.007851.

(102) Zou, Y.; Qian, Z.; Chen, Y.; Qian, H.; Wei, G.; Zhang, Q. Norepinephrine Inhibits Alzheimer's Amyloid-β Peptide Aggregation and Destabilizes Amyloid-β Protofibrils: A Molecular Dynamics Simulation Study. *ACS Chem. Neurosci.* **2019**, *10* (3), 1585–1594. https://doi.org/10.1021/acschemneuro.8b00537.

(103) Doig, A. J.; Derreumaux, P. Inhibition of Protein Aggregation and Amyloid Formation by Small Molecules. *Curr. Opin. Struct. Biol.* **2015**, *30*, 50–56. https://doi.org/10.1016/j.sbi.2014.12.004.

(104) Karthivashan, G.; Ganesan, P.; Park, S.-Y.; Kim, J.-S.; Choi, D.-K. Therapeutic Strategies and Nano-Drug Delivery Applications in Management of Ageing Alzheimer's Disease. *Drug Deliv.* **2018**, *25* (1), 307–320. https://doi.org/10.1080/10717544.2018.1428243.

(105) Baghel, S.; Cathcart, H.; O'Reilly, N. J. Polymeric Amorphous Solid Dispersions: A Review of Amorphization, Crystallization, Stabilization, Solid-State Characterization, and Aqueous Solubilization of Biopharmaceutical Classification System Class II Drugs. *J. Pharm. Sci.* **2016**, *105* (9), 2527–2544. https://doi.org/10.1016/j.xphs.2015.10.008.

(106) Ota, M.; Koike, R.; Amemiya, T.; Tenno, T.; Romero, P. R.; Hiroaki, H.; Dunker, A. K.; Fukuchi, S. An Assignment of Intrinsically Disordered Regions of Proteins Based on NMR Structures. *J. Struct. Biol.* **2013**, *181* (1), 29–36. https://doi.org/10.1016/j.jsb.2012.10.017.

(107) Sikorski, P.; Atkins, E. D. T.; Serpell, L. C. Structure and Texture of Fibrous Crystals Formed by Alzheimer's Aβ(11–25) Peptide Fragment. *Structure* **2003**, *11* (8), 915–926. https://doi.org/10.1016/S0969-2126(03)00149-7.

(108) Kreutzer, A. G.; Nowick, J. S. Elucidating the Structures of Amyloid Oligomers with Macrocyclic β-Hairpin Peptides: Insights into Alzheimer's Disease and Other Amyloid Diseases. *Acc. Chem. Res.* **2018**, *51* (3), 706–718. https://doi.org/10.1021/acs.accounts.7b00554.

(109) Madine, J.; Jack, E.; Stockley, P. G.; Radford, S. E.; Serpell, L. C.; Middleton, D. A. Structural Insights into the Polymorphism of Amyloid-Like Fibrils Formed by Region 20−29 of Amylin Revealed by Solid-State NMR and X-Ray Fiber Diffraction. *J. Am. Chem. Soc.* **2008**, *130* (45), 14990–15001. https://doi.org/10.1021/ja802483d.

(110) Loquet, A.; El Mammeri, N.; Stanek, J.; Berbon, M.; Bardiaux, B.; Pintacuda, G.; Habenstein, B. 3D Structure Determination of Amyloid Fibrils Using Solid-State NMR Spectroscopy. *Methods* **2018**, *138–139*, 26–38. https://doi.org/10.1016/j.ymeth.2018.03.014.

(111) Tycko, R. Solid-State NMR Studies of Amyloid Fibril Structure. *Annu. Rev. Phys. Chem.* **2011**, *62* (1), 279–299. https://doi.org/10.1146/annurev-physchem-032210-103539.

(112) Iwata, K.; Fujiwara, T.; Matsuki, Y.; Akutsu, H.; Takahashi, S.; Naiki, H.; Goto, Y. 3D Structure of Amyloid Protofilaments of B2-Microglobulin Fragment Probed by Solid-State NMR. *Proc. Natl. Acad. Sci. USA* **2006**, *103* (48), 18119–18124. https://doi.org/10.1073/pnas.0607180103.

(113) Liu, Y.; Huynh, D. T.; Yeates, T. O. A 3.8 Å Resolution Cryo-EM Structure of a Small Protein Bound to an Imaging Scaffold. *Nat. Commun.* **2019**, *10* (1), 1864. https://doi.org/10.1038/s41467-019-09836-0.

(114) Renaud, J.-P.; Chari, A.; Ciferri, C.; Liu, W.; Rémigy, H.-W.; Stark, H.; Wiesmann, C. Cryo-EM in Drug Discovery: Achievements, Limitations and Prospects. *Nat. Rev. Drug Disc.* **2018**, *17* (7), 471–492. https://doi.org/10.1038/nrd.2018.77.

(115) Greenfield, N. J. Analysis of Circular Dichroism Data. In *Methods in Enzymology*; Elsevier, **2004**, *383*, 282–317. https://doi.org/10.1016/S0076-6879(04)83012-X.

(116) Ami, D.; Lavatelli, F.; Rognoni, P.; Palladini, G.; Raimondi, S.; Giorgetti, S.; Monti, L.; Doglia, S. M.; Natalello, A.; Merlini, G. In Situ Characterization of Protein Aggregates in Human Tissues Affected by Light Chain Amyloidosis: A

FTIR Microspectroscopy Study. *Sci. Rep.* **2016**, *6* (1), 29096. https://doi.org/10.1038/srep29096.

(117) Barth, A. Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta* **2007**, *1767*, 1073–1101. https://doi.org/10.1016/j.bbabio.2007.06.004.

(118) Oladepo, S. A.; Xiong, K.; Hong, Z.; Asher, S. A.; Handen, J.; Lednev, I. K. UV Resonance Raman Investigations of Peptide and Protein Structure and Dynamics. *Chem. Rev.* **2012**, *112* (5), 2604–2628. https://doi.org/10.1021/cr200198a.

(119) Streets, A. M.; Sourigues, Y.; Kopito, R. R.; Melki, R.; Quake, S. R. Simultaneous Measurement of Amyloid Fibril Formation by Dynamic Light Scattering and Fluorescence Reveals Complex Aggregation Kinetics. *PLoS ONE* **2013**, *8* (1), e54541. https://doi.org/10.1371/journal.pone.0054541.

(120) Georgalis, Y.; Starikov, E. B.; Hollenbach, B.; Lurz, R.; Scherzinger, E.; Saenger, W.; Lehrach, H.; Wanker, E. E. Huntingtin Aggregation Monitored by Dynamic Light Scattering. *Proc. Natl. Acad. Sci. USA* **1998**, *95* (11), 6118–6121. https://doi.org/10.1073/pnas.95.11.6118.

(121) Chiang, Y.-L.; Chang, Y.-C.; Chiang, I.-C.; Mak, H.-M.; Hwang, I.-S.; Shih, Y.-L. Atomic Force Microscopy Characterization of Protein Fibrils Formed by the Amyloidogenic Region of the Bacterial Protein MinE on Mica and a Supported Lipid Bilayer. *PLoS ONE* **2015**, *10* (11), e0142506. https://doi.org/10.1371/journal.pone.0142506.

(122) Zhang, L.; Li, N.; Gao, F.; Hou, L.; Xu, Z. Insulin Amyloid Fibrils: An Excellent Platform for Controlled Synthesis of Ultrathin Superlong Platinum Nanowires with High Electrocatalytic Activity. *J. Am. Chem. Soc.* **2012**, *134* (28), 11326–11329. https://doi.org/10.1021/ja302959e.

(123) Espargaró, A.; Llabrés, S.; Saupe, S. J.; Curutchet, C.; Luque, F. J.; Sabaté, R. On the Binding of Congo Red to Amyloid Fibrils. *Angew. Chem. Int. Ed.* **2020**, *59* (21), 8104–8107. https://doi.org/10.1002/anie.201916630.

(124) Biancalana, M.; Koide, S. Molecular Mechanism of Thioflavin-T Binding to Amyloid Fibrils. *Biochim. Biophys. Acta* **2010**, *1804* (7), 1405–1412. https://doi.org/10.1016/j.bbapap.2010.04.001.

(125) Mishra, R.; Sjölander, D.; Hammarström, P. Spectroscopic Characterization of Diverse Amyloid Fibrilsin Vitro by the Fluorescent Dye Nile Red. *Mol. BioSyst.* **2011**, *7* (4), 1232–1240. https://doi.org/10.1039/C0MB00236D.

(126) Long, D. A. Raman Spectroscopy. *Great Britain: McGraw-Hill* **1977**.

(127) Long, D. A. The Raman Effect: A United Treatment of the Theory For Raman Scatteing by Molecules. *John Wiley & Sons Ltd* **2002**.

(128) Asher, S. A.; Ludwig, Michael.; Johnson, C. R. UV Resonance Raman Excitation Profiles of the Aromatic Amino Acids. *J. Am. Chem. Soc.* **1986**, *108* (12), 3186–3197. https://doi.org/10.1021/ja00272a005.

(129) Maiti, N. C.; Apetri, M. M.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E. Raman Spectroscopic Characterization of Secondary Structure in Natively Unfolded Proteins: α-Synuclein. *J. Am. Chem. Soc.* **2004**, *126* (8), 2399–2408. https://doi.org/10.1021/ja0356176.

(130) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. Obtaining Information about Protein Secondary Structures in Aqueous Solution Using Fourier Transform IR Spectroscopy. *Nat. Protoc.* **2015**, *10* (3), 382–396. https://doi.org/10.1038/nprot.2015.024.

(131) Joseph, M. M.; Narayanan, N.; Nair, J. B.; Karunakaran, V.; Ramya, A. N.; Sujai, P. T.; Saranya, G.; Arya, J. S.; Vijayan, V. M.; Maiti, K. K. Exploring the Margins of SERS in Practical Domain: An Emerging Diagnostic Modality for Modern Biomedical Applications. *Biomaterials* **2018**, *181*, 140–181. https://doi.org/10.1016/j.biomaterials.2018.07.045.

(132) Hong, Z.; Asher, S. A. Dependence of Raman and Resonance Raman Intensities on Sample Self-Absorption. *Appl. Spectrosc.* **2015**, *69* (1), 75–83. https://doi.org/10.1366/14-07531.

(133) Teresa, M.; Petersen, S.; Prakash, G. UV Light Effects on Proteins: From Photochemistry to Nanomedicine. In *Molecular Photochemistry - Various Aspects*; InTech, **2012**. https://doi.org/10.5772/37947.

(134) D'Amico, F.; Saito, M.; Bencivenga, F.; Marsi, M.; Gessini, A.; Camisasca, G.; Principi, E.; Cucini, R.; Di Fonzo, S.; Battistoni, A.; Giangrisostomi, E.; Masciovecchio, C. UV Resonant Raman Scattering Facility at Elettra. *Nucl.*

Instrum. Methods Phys. Res. **2013**, *703*, 33–37. https://doi.org/10.1016/j.nima.2012.11.037.

(135) Chi, Z.; Asher, S. A. UV Raman Determination of the Environment and Solvent Exposure of Tyr and Trp Residues. *J. Phys. Chem. B* **1998**, *102* (47), 9595–9602. https://doi.org/10.1021/jp9828336.

(136) Xu, M.; Ermolenkov, V. V.; Uversky, V. N.; Lednev, I. K. Hen Egg White Lysozyme Fibrillation: A Deep-UV Resonance Raman Spectroscopic Study. *J. Biophoton.* **2008**, *1* (3), 215–229. https://doi.org/10.1002/jbio.200710013.

(137) Lakowicz, J. R.; Masters, B. R. Principles of Fluorescence Spectroscopy, Third Edition. *J. Biomed. Opt.* **2008**, *13* (2), 029901. https://doi.org/10.1117/1.2904580.

(138) Barth, A.; Zscherp, C. What Vibrations Tell about Proteins. *Q. Rev. Biophys.* **2002**, *35* (4), 369–430. https://doi.org/10.1017/S0033583502003815.

(139) Colthup , N.; Daly, L.; Wiberley, S. Introduction to Infrared and Raman Spectroscopy; *Elsevier*, **1975**. https://doi.org/10.1016/B978-0-12-182552-2.X5001-3..

(140) Lupi, S.; Nucara, A.; Perucchi, A.; Calvani, P.; Ortolani, M.; Quaroni, L.; Kiskinova, M. Performance of SISSI, the Infrared Beamline of the ELETTRA Storage Ring. *J. Opt. Soc. Am. B* **2007**, *24* (4), 959. https://doi.org/10.1364/JOSAB.24.000959.

(141) Amenabar, I.; Poly, S.; Nuansing, W.; Hubrich, E. H.; Govyadinov, A. A.; Huth, F.; Krutokhvostov, R.; Zhang, L.; Knez, M.; Heberle, J.; Bittner, A. M.; Hillenbrand, R. Structural Analysis and Mapping of Individual Protein Complexes by Infrared Nanospectroscopy. *Nat. Commun.* **2013**, *4*. https://doi.org/10.1038/ncomms3890.

(142) Misra, G. Chapter 3 - Fluorescence Spectroscopy. In *Data Processing Handbook for Complex Biological Data Sources*; Academic Press, **2019**, 31–37. https://doi.org/10.1016/B978-0-12-816548-5.00003-4.

(143) Fleming, K. G. Fluorescence Theory. In *Encyclopedia of Spectroscopy and Spectrometry*; Elsevier, **2017**, 647–653. https://doi.org/10.1016/B978-0-12-803224-4.00357-5.

(144) Lehrer, S. Solute Perturbation of Protein Fluorescence. Quenching of the Tryptophyl Fluorescence of Model Compounds and of Lysozyme by Iodide Ion. *Biochemistry* **1971**, *10* (17), 3254–3263. https://doi.org/10.1021/bi00793a015.

(145) Chen, Y.; Barckley, M.D. Toward Understanding Tryptophan Fluorescence in Proteins. *Biochemistry* **1998**, *37*, 9976–9982. https://doi.org/10.1021/bi980274n.

(146) Vaughn, M. B.; Biren, C.; Li, Q.; Ragupathi, A.; Dyer, R. B. Site-Specific Tryptophan Labels Reveal Local Microsecond–Millisecond Motions of Dihydrofolate Reductase. *Molecules* **2020**, *25* (17), 3819. https://doi.org/10.3390/molecules25173819.

(147) Dusa, A.; Kaylor, J.; Edridge, S.; Bodner, N.; Hong, D.-P.; Fink, A. L. Characterization of Oligomers during α-Synuclein Aggregation Using Intrinsic Tryptophan Fluorescence. *Biochemistry* **2006**, *45* (8), 2752–2760. https://doi.org/10.1021/bi051426z.

(148) Dufrêne, Y. F.; Ando, T.; Garcia, R.; Alsteens, D.; Martinez-Martin, D.; Engel, A.; Gerber, C.; Müller, D. J. Imaging Modes of Atomic Force Microscopy for Application in Molecular and Cell Biology. *Nat. Nanotechnol.* **2017**, *12* (4), 295–307. https://doi.org/10.1038/nnano.2017.45.

(149) Alessandrini, A.; Facci, P. AFM: A Versatile Tool in Biophysics. *Meas. Sci. Technol.* **2005**, *16* (6), R65–R92. https://doi.org/10.1088/0957-0233/16/6/R01.

(150) Haugstad, G. Overview of AFM. In *Atomic Force Microscopy*; John Wiley & Sons, Ltd, **2012**; pp 1–32. https://doi.org/10.1002/9781118360668.ch1.

(151) Agarwal, D. H.; Bhatt, P. M.; Pathan, A. M.; Patel, H.; Joshi, U. S. Development of Portable Experimental Set-up for AFM to Work at Cryogenic Temperature. *AIP Conf. Proc.* **2012**, *1447* (1), 531–532. https://doi.org/10.1063/1.4710113.

(152) Vittorio, O.; Curcio, M.; Cojoc, M.; Goya, G. F.; Hampel, S.; Iemma, F.; Dubrovska, A.; Cirillo, G. Polyphenols Delivery by Polymeric Materials: Challenges in Cancer Treatment. *Drug Delivery* **2017**, *24* (1), 162–180. https://doi.org/10.1080/10717544.2016.1236846.

(153) Pachetti, M; D'Amico, F; Gessini, A; Parisse, P; Pucciarelli, S; Vaccari, L; Masciovecchio, C. UV Resonance Raman Explores Protein Structural Modification upon Fibrillation and Ligand-Interaction, *under peer-review.*

(154) Buxbaum, J. N. Diseases of Protein Conformation: What Do in Vitro Experiments Tell Us about in Vivo Diseases? *Trends Biochem. Sci.* **2003**, *28* (11), 585–592. https://doi.org/10.1016/j.tibs.2003.09.009.

(155) Young, A. C. M.; Tilton, R. F.; Dewan, J. C. Thermal Expansion of Hen Egg-White Lysozyme: Comparison of the 1·9 Å Resolution Structures of the Tetragonal Form of the Enzyme at 100 K and 298 K. *J. Mol. Biol.* **1994**, *235* (1), 302–317. https://doi.org/10.1016/S0022-2836(05)80034-8.

(156) Lemkul, J. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package. *LiveCoMS* **2018**, *1* (1), 5068. https://doi.org/10.33011/livecoms.1.1.5068.

(157) Poulsen, F. M.; Hoch, J. C.; Dobson, C. M. Structural Study of the Hydrophobic Box Region of Lysozyme in Solution Using Nuclear Overhauser Effects. *Biochemistry* **1980**, *19* (12), 2597–2607. https://doi.org/10.1021/bi00553a011.

(158) Goda, S.; Takano, K.; Yutani, K.; Yamagata, Y.; Nagata, R.; Akutsu, H.; Maki, S.; Namba, K. Amyloid Protofilament Formation of Hen Egg Lysozyme in Highly Concentrated Ethanol Solution. *Protein Sci.* **2008**, *9* (2), 369–375. https://doi.org/10.1110/ps.9.2.369.

(159) Cao, A. Formation of Amyloid Fibrils from Fully Reduced Hen Egg White Lysozyme. *Protein Sci.* **2004**, *13* (2), 319–324. https://doi.org/10.1110/ps.03183404.

(160) Booth, D. R.; Sunde, M.; Bellotti, V.; Robinson, C. V.; Hutchinson, W. L.; Fraser, P. E.; Hawkins, P. N.; Dobson, C. M.; Radford, S. E.; Blake, C. C. F.; Pepys, M. B. Instability, Unfolding and Aggregation of Human Lysozyme Variants Underlying Amyloid Fibrillogenesis. *Nature* **1997**, *385* (6619), 787–793. https://doi.org/10.1038/385787a0.

(161) Morozova-Roche, L. A.; Zurdo, J.; Spencer, A.; Noppe, W.; Receveur, V.; Archer, D. B.; Joniau, M.; Dobson, C. M. Amyloid Fibril Formation and Seeding by Wild-Type Human Lysozyme and Its Disease-Related Mutational Variants. *J. Struct. Biol.* **2000**, *130* (2–3), 339–351. https://doi.org/10.1006/jsbi.2000.4264.

(162) Blake, C. C.; Grace, D. E.; Johnson, L. N.; Perkins, S. J.; Phillips, D. C.; Cassels, R.; Dobson, C. M.; Poulsen, F. M.; Williams, R. J. Physical and Chemical Properties

of Lysozyme. *Ciba Found Symp.* **1977**, No. 60, 137–185. https://doi.org/10.1002/9780470720424.ch10.

(163) Herning, T.; Yutani, K.; Taniyama, Y.; Kikuchi, M. Effects of Proline Mutations on the Unfolding and Refolding of Human Lysozyme: The Slow Refolding Kinetic Phase Does Not Result from Proline Cis-Trans Isomerization. *Biochemistry* **1991**, *30*, 9882–9891. https://doi.org/10.1021/bi00105a011.

(164) Redfield, C.; Dobson, C. M. Proton NMR studies of human lysozyme: spectral assignment and comparison with hen lysozyme. *Biochemistry* **1990**, *29*, 7201–7214. https://doi.org/10.1021/bi00483a007.

(165) Swaminathan, R.; Ravi, V. K.; Kumar, S.; Kumar, M. V. S.; Chandra, N. Lysozyme: A Model Protein for Amyloid Research. In *Advances in Protein Chemistry and Structural Biology*; Academic Press, **2011**; Vol. 84, pp 63–111. https://doi.org/10.1016/B978-0-12-386483-3.00003-3.

(166) Miranker, A.; Radford, S.; Karplus, M.; Dobson, C.M. Demonstration by NMR of folding domains in lysozyme. *Nature,* **1991,** *349,* 633–636. https://doi.org/10.1038/349633a0..

(167) Hill, S. E.; Robinson, J.; Matthews, G.; Muschol, M. Amyloid Protofibrils of Lysozyme Nucleate and Grow Via Oligomer Fusion. *Biophys. J.* **2009**, *96* (9), 3781–3790. https://doi.org/10.1016/j.bpj.2009.01.044.

(168) Mayer, J. P.; Zhang, F.; DiMarchi, R. D. Insulin Structure and Function. *Peptide Sci.* **2007**, *88* (5), 687–713. https://doi.org/10.1002/bip.20734.

(169) van Lierop, B.; Ong, S. C.; Belgi, A.; Delaine, C.; Andrikopoulos, S.; Haworth, N. L.; Menting, J. G.; Lawrence, M. C.; Robinson, A. J.; Forbes, B. E. Insulin in Motion: The A6-A11 Disulfide Bond Allosterically Modulates Structural Transitions Required for Insulin Activity. *Sci. Rep.* **2017**, *7* (1), 17239. https://doi.org/10.1038/s41598-017-16876-3.

(170) Smith, M. I.; Foderà, V.; Sharp, J. S.; Roberts, C. J.; Donald, A. M. Factors Affecting the Formation of Insulin Amyloid Spherulites. *Colloids Surf. B* **2012**, *89*, 216–222. https://doi.org/10.1016/j.colsurfb.2011.09.018.

(171) Nielsen, L.; Frokjaer, S.; Carpenter, J. F.; Brange, J. Studies of the Structure of Insulin Fibrils by Fourier Transform Infrared (FTIR) Spectroscopy and Electron Microscopy. *J. Pharm. Sci.* **2001**, *90* (1), 9.

(172) Moussa, C.; Hebron, M.; Huang, X.; Ahn, J.; Rissman, R. A.; Aisen, P. S.; Turner, R. S. Resveratrol Regulates Neuro-Inflammation and Induces Adaptive Immunity in Alzheimer's Disease. *J. Neuroinflammation* **2017**, *14* (1), 1. https://doi.org/10.1186/s12974-016-0779-0.

(173) Turner, R. S.; Thomas, R. G.; Craft, S.; Dyck, C. H. van; Mintzer, J.; Reynolds, B. A.; Brewer, J. B.; Rissman, R. A.; Raman, R.; Aisen, P. S.; Study, F. the A. D. C. A Randomized, Double-Blind, Placebo-Controlled Trial of Resveratrol for Alzheimer Disease. *Neurology* **2015**, *85* (16), 1383–1391. https://doi.org/10.1212/WNL.0000000000002035.

(174) Berman, A. Y.; Motechin, R. A.; Wiesenfeld, M. Y.; Holz, M. K. The Therapeutic Potential of Resveratrol: A Review of Clinical Trials. *npj Precis. Oncol.* **2017**, *1* (1), 1–9. https://doi.org/10.1038/s41698-017-0038-6.

(175) Murakami, A. Dose-Dependent Functionality and Toxicity of Green Tea Polyphenols in Experimental Rodents. *Arch. Biochem. Biophys.* **2014**, *557*, 3–10. https://doi.org/10.1016/j.abb.2014.04.018.

(176) Arosio, P.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Chemical Kinetics for Drug Discovery to Combat Protein Aggregation Diseases. *Trends Pharmacol. Sci.* **2014**, *35* (3), 127–135. https://doi.org/10.1016/j.tips.2013.12.005.

(177) Gazova, Z.; Siposova, K.; Kurin, E.; Mučaji, P.; Nagy, M. Amyloid Aggregation of Lysozyme: The Synergy Study of Red Wine Polyphenols. *Proteins* **2013**, *81* (6), 994–1004. https://doi.org/10.1002/prot.24250.

(178) Mohammadi, F.; Mahmudian, A.; Moeeni, M.; Hassani, L. Inhibition of Amyloid Fibrillation of Hen Egg-White Lysozyme by the Natural and Synthetic Curcuminoids. *RSC Adv.* **2016**, *6* (28), 23148–23160. https://doi.org/10.1039/C5RA18992F.

(179) Pantusa, M.; Bartucci, R.; Rizzuti, B. Stability of *Trans* -Resveratrol Associated with Transport Proteins. *J. Agric. Food Chem.* **2014**, *62* (19), 4384–4391. https://doi.org/10.1021/jf405584a.

(180) Feng, Y.; Wang, X.; Yang, S.; Wang, Y.; Zhang, X.; Du, X.; Sun, X.; Zhao, M.; Huang, L.; Liu, R. Resveratrol Inhibits Beta-Amyloid Oligomeric Cytotoxicity but Does Not Prevent Oligomer Formation. *NeuroToxicology* **2009**, *30* (6), 986–995. https://doi.org/10.1016/j.neuro.2009.08.013.

(181) Vingtdeux, V.; Dreses-Werringloer, U.; Zhao, H.; Davies, P.; Marambaud, P. Therapeutic Potential of Resveratrol in Alzheimer's Disease. *BMC Neurosci* **2008**, *9* (2), S6. https://doi.org/10.1186/1471-2202-9-S2-S6.

(182) Mishra, R.; Sellin, D.; Radovan, D.; Gohlke, A.; Winter, R. Inhibiting Islet Amyloid Polypeptide Fibril Formation by the Red Wine Compound Resveratrol. *ChemBioChem* **2009**, *10* (3), 445–449. https://doi.org/10.1002/cbic.200800762.

(183) Sawda, C.; Moussa, C.; Turner, R. S. Resveratrol for Alzheimer's Disease. *Ann. N. Y. Acad. Sci.* **2017**, *1403* (1), 142–149. https://doi.org/10.1111/nyas.13431.

(184) Brasnyó, P.; Molnár, G. A.; Mohás, M.; Markó, L.; Laczy, B.; Cseh, J.; Mikolás, E.; Szijártó, I. A.; Mérei, Á.; Halmai, R.; Mészáros, L. G.; Sümegi, B.; Wittmann, I. Resveratrol Improves Insulin Sensitivity, Reduces Oxidative Stress and Activates the Akt Pathway in Type 2 Diabetic Patients. *Br. J. Nutr.* **2011**, *106* (3), 383–389. https://doi.org/10.1017/S0007114511000316.

(185) de Vries, K.; Strydom, M.; Steenkamp, V. Bioavailability of Resveratrol: Possibilities for Enhancement. *J. Herb. Med.* **2018**, *11*, 71–77. https://doi.org/10.1016/j.hermed.2017.09.002.

(186) Brown, V. A.; Patel, K. R.; Viskaduraki, M.; Crowell, J. A.; Perloff, M.; Booth, T. D.; Vasilinin, G.; Sen, A.; Schinas, A. M.; Piccirilli, G.; Brown, K.; Steward, W. P.; Gescher, A. J.; Brenner, D. E. Repeat Dose Study of the Cancer Chemopreventive Agent Resveratrol in Healthy Volunteers: Safety, Pharmacokinetics, and Effect on the Insulin-like Growth Factor Axis. *Cancer Res.* **2010**, *70* (22), 9003–9011. https://doi.org/10.1158/0008-5472.CAN-10-2364.

(187) Martins, L. A. M.; Coelho, B. P.; Behr, G.; Pettenuzzo, L. F.; Souza, I. C. C.; Moreira, J. C. F.; Borojevic, R.; Gottfried, C.; Guma, F. C. R. Resveratrol Induces Pro-Oxidant Effects and Time-Dependent Resistance to Cytotoxicity in Activated Hepatic Stellate Cells. *Cell Biochem. Biophys.* **2014**, *68* (2), 247–257. https://doi.org/10.1007/s12013-013-9703-8.

(188) Nečas, D.; Klapetek, P. Gwyddion: An Open-Source Software for SPM Data Analysis. *Open Phys.* **2012**, *10* (1). https://doi.org/10.2478/s11534-011-0096-2.

(189) Graham, J. S.; McCullough, B. R.; Kang, H.; Elam, W. A.; Cao, W.; De La Cruz, E. M. Multi-Platform Compatible Software for Analysis of Polymer Bending Mechanics. *PLoS ONE* **2014**, *9* (4), e94766. https://doi.org/10.1371/journal.pone.0094766.

(190) Byler, D. M.; Susi, H.; Regional, E. Examination of the Secondary Structure of Proteins by Deconvolved FTIR Spectra. *Biopolymers* **1986**, *25*, 469–487. https://doi.org/10.1002/bip.360250307

(191) Yang, H.; Yang, S.; Kong, J.; Dong, A.; Yu, S. Obtaining Information about Protein Secondary Structures in Aqueous Solution Using Fourier Transform IR Spectroscopy. *Nat Protoc.* **2015**. https://doi.org/10.1038/nprot.2015.024.

(192) Mossuto, M. F.; Dhulesia, A.; Devlin, G.; Frare, E.; Kumita, J. R.; de Laureto, P. P.; Dumoulin, M.; Fontana, A.; Dobson, C. M.; Salvatella, X. The Non-Core Regions of Human Lysozyme Amyloid Fibrils Influence Cytotoxicity. *J. Mol. Biol.* **2010**, *402* (5–2), 783–796. https://doi.org/10.1016/j.jmb.2010.07.005.

(193) Zako, T.; Sakono, M.; Hashimoto, N.; Ihara, M.; Maeda, M. Bovine Insulin Filaments Induced by Reducing Disulfide Bonds Show a Different Morphology, Secondary Structure, and Cell Toxicity from Intact Insulin Amyloid Fibrils. *Biophys. J.* **2009**, *96* (8), 3331–3340. https://doi.org/10.1016/j.bpj.2008.12.3957.

(194) Konar, M.; Mathew, A.; Dasgupta, S. Effect of Silica Nanoparticles on the Amyloid Fibrillation of Lysozyme. *ACS Omega* **2019**, *4* (1), 1015–1026. https://doi.org/10.1021/acsomega.8b03169.

(195) Mulaj, M.; Foley, J.; Muschol, M. Amyloid Oligomers and Protofibrils, but Not Filaments, Self-Replicate from Native Lysozyme. *J. Am. Chem. Soc.* **2014**, *136* (25), 8947–8956. https://doi.org/10.1021/ja502529m.

(196) Venkataramani, S.; Truntzer, J.; Coleman, D. Thermal Stability of High Concentration Lysozyme across Varying PH: A Fourier Transform Infrared Study. *J. Pharm. Bioall. Sci.* **2013**, *5* (2), 148. https://doi.org/10.4103/0975-7406.111821.

(197) Matheus, S.; Friess, W.; Mahler, H.-C. FTIR and NDSC as Analytical Tools for High-Concentration Protein Formulations. *Pharm. Res.* **2006**, *23* (6), 1350–1363. https://doi.org/10.1007/s11095-006-0142-8.

(198) Meersman, F.; Heremans, K. Temperature-Induced Dissociation of Protein Aggregates: Accessing the Denatured State. *Biochemistry* **2003**, *42* (48), 14234–14241. https://doi.org/10.1021/bi035623e.

(199) Piccirilli, F.; Plotegher, N.; Ortore, M. G.; Tessari, I.; Brucale, M.; Spinozzi, F.; Beltramini, M.; Mariani, P.; Militello, V.; Lupi, S.; Perucchi, A.; Bubacco, L. High-Pressure-Driven Reversible Dissociation of α-Synuclein Fibrils Reveals Structural Hierarchy. *Biophys. J.* **2017**, *113* (8), 1685–1696. https://doi.org/10.1016/j.bpj.2017.08.042.

(200) Villegas, V.; Zurdo, J.; Filimonov, V. V.; Avilés, F. X.; Dobson, C. M.; Serrano, L. Protein Engineering as a Strategy to Avoid Formation of Amyloid Fibrils. *Protein Sci.* **2000**, *9* (9), 1700–1708.

(201) Zandomeneghi, G.; Krebs, M. R. H.; McCammon, M. G.; Fändrich, M. FTIR Reveals Structural Differences between Native β-Sheet Proteins and Amyloid Fibrils. *Protein Sci.* **2004**, *13* (12), 3314–3321. https://doi.org/10.1110/ps.041024904.

(202) Fändrich, M.; Forge, V.; Buder, K.; Kittler, M.; Dobson, C. M.; Diekmann, S. Myoglobin Forms Amyloid Fibrils by Association of Unfolded Polypeptide Segments. *Proc. Natl. Acad. Sci. USA* **2003**, *100* (26), 15463–15468. https://doi.org/10.1073/pnas.0303758100.

(203) Fändrich, M.; Dobson, C. M. The Behaviour of Polyamino Acids Reveals an Inverse Side Chain Effect in Amyloid Structure Formation. *EMBO J.* **2002**, *21* (21), 5682–5690. https://doi.org/10.1093/emboj/cdf573.

(204) Zou, Y.; Li, Y.; Hao, W.; Hu, X.; Ma, G. Parallel β-Sheet Fibril and Antiparallel β-Sheet Oligomer: New Insights into Amyloid Formation of Hen Egg White Lysozyme under Heat and Acidic Condition from FTIR Spectroscopy. *J. Phys. Chem. B* **2013**, *117* (15), 4003–4013. https://doi.org/10.1021/jp4003559.

(205) Bouchard, M.; Zurdo, J.; Nettleton, E. J.; Dobson, C. M.; Robinson, C. V. Formation of Insulin Amyloid Fibrils Followed by FTIR Simultaneously with CD

and Electron Microscopy. *Protein Sci.* **2000**, *9* (10), 1960–1967. https://doi.org/10.1110/ps.9.10.1960.

(206) Lomont, J. P.; Ostrander, J. S.; Ho, J.-J.; Petti, M. K.; Zanni, M. T. Not All β-Sheets Are the Same: Amyloid Infrared Spectra, Transition Dipole Strengths, and Couplings Investigated by 2D IR Spectroscopy. *J. Phys. Chem. B* **2017**, *121* (38), 8935–8945. https://doi.org/10.1021/acs.jpcb.7b06826.

(207) Schmick, S. D.; Weliky, D. P. Major Antiparallel and Minor Parallel β Sheet Populations Detected in the Membrane-Associated Human Immunodeficiency Virus Fusion Peptide. *Biochemistry* **2010**, *49* (50), 10623–10635. https://doi.org/10.1021/bi101389r.

(208) Tremmel, S.; Beyermann, M.; Oschkinat, H.; Bienert, M.; Naumann, D.; Fabian, H. 13C-Labeled Tyrosine Residues as Local IR Probes for Monitoring Conformational Changes in Peptides and Proteins. *Angew. Chem. Int. Ed.* **2005**, *44* (29), 4631–4635. https://doi.org/10.1002/anie.200500547.

(209) Reinstädler, D.; Fabian, H.; Backmann, J.; Naumann, D. Refolding of Thermally and Urea-Denatured Ribonuclease A Monitored by Time-Resolved FTIR Spectroscopy. *Biochemistry* **1996**, *35* (49), 15822–15830. https://doi.org/10.1021/bi961810j.

(210) Berthomieu, C.; Hienerwadel, R. Vibrational Spectroscopy to Study the Properties of Redox-Active Tyrosines in Photosystem II and Other Proteins. *Biochim. Biophys. Acta* **2005**, *1707* (1), 51–66. https://doi.org/10.1016/j.bbabio.2004.03.011.

(211) Chirgadze, Y. N.; Fedorov, O. V.; Trushina, N. P. Estimation of Amino Acid Residue Side-Chain Absorption in the Infrared Spectra of Protein Solutions in Heavy Water. *Biopolymers* **1975**, *14* (4), 679–694. https://doi.org/10.1002/bip.1975.360140402.

(212) Zscherp, C.; Aygün, H.; Engels, J. W.; Mäntele, W. Effect of Proline to Alanine Mutation on the Thermal Stability of the All-β-Sheet Protein Tendamistat. *Biochim. Biophys. Acta* **2003**, *1651* (1), 139–145. https://doi.org/10.1016/S1570-9639(03)00263-2.

(213) Gabashvili, I. S.; Menikh, A.; Fragata, M. MOLECULAR Protein Structure of Photosystem II Studied by FT-IR Spectroscopy . Effect of Digalactosyldiacylglycerol on the Tyrosine Side Chain Residues. *J. Mol. Struct.* **1998**, *444*, 123–133.

(214) Fabian, H.; Schultz, C.; Naumann, D.; Landt, O.; Hahn, U.; Saenger, W. Secondary Structure and Temperature-Induced Unfolding and Refolding of Ribonuclease T1 in Aqueous Solution: A Fourier Transform Infrared Spectroscopic Study. *J. Mol. Biol.***1993**, *232* (3), 967–981. https://doi.org/10.1006/jmbi.1993.1442.

(215) Arrondo, J. L. R.; Muga, A.; Castresana, J.; Goñi, F. M. Quantitative Studies of the Structure of Proteins in Solution by Fourier-Transform Infrared Spectroscopy. *Prog. Biophys. Mol. Biol.* **1993**, *59* (1), 23–56. https://doi.org/10.1016/0079-6107(93)90006-6.

(216) Shashilov, V. A.; Sikirzhytski, V.; Popova, L. A.; Lednev, I. K. Quantitative Methods for Structural Characterization of Proteins Based on Deep UV Resonance Raman Spectroscopy. *Methods* **2010**, *52* (1), 23–37. https://doi.org/10.1016/j.ymeth.2010.05.004.

(217) Shashilov, V.; Xu, M.; Ermolenkov, V. V.; Fredriksen, L.; Lednev, I. K. Probing a Fibrillation Nucleus Directly by Deep Ultraviolet Raman Spectroscopy. *J. Am. Chem. Soc.* **2007**, *129* (22), 6972–6973. https://doi.org/10.1021/ja070038c.

(218) Kurouski, D.; Van Duyne, R. P.; Lednev, I. K. Exploring the Structure and Formation Mechanism of Amyloid Fibrils by Raman Spectroscopy: A Review. *Analyst* **2015**, *140* (15), 4967–4980. https://doi.org/10.1039/C5AN00342C.

(219) Chi, Z.; Asher, S. A. UV Resonance Raman Determination of Protein Acid Denaturation: Selective Unfolding of Helical Segments of Horse Myoglobin †. *Biochemistry* **1998**, *37* (9), 2865–2872. https://doi.org/10.1021/bi971161r.

(220) Schlamadinger, D. E.; Gable, J. E.; Kim, J. E. Hydrogen Bonding and Solvent Polarity Markers in the UV Resonance Raman Spectrum of Tryptophan: Application to Membrane Proteins. *J. Phys. Chem. B* **2009**, *113* (44), 14769–14778. https://doi.org/10.1021/jp905473y.

(221) JiJi, R. D.; Balakrishnan, G.; Hu, Y.; Spiro, T. G. Intermediacy of Poly(L -Proline) II and β-Strand Conformations in Poly(L -Lysine) β-Sheet Formation Probed by Temperature-Jump/UV Resonance Raman Spectroscopy . *Biochemistry* **2006**, *45* (1), 34–41. https://doi.org/10.1021/bi051507v.

(222) V. Simpson, J.; Balakrishnan, G.; D. JiJi, R. MCR-ALS Analysis of Two-Way UV Resonance Raman Spectra to Resolve Discrete Protein Secondary Structural Motifs. *Analyst* **2009**, *134* (1), 138–147. https://doi.org/10.1039/B814392G.

(223) Xu, M.; Shashilov, V.; Lednev, I. K. Probing the Cross-β Core Structure of Amyloid Fibrils by Hydrogen−Deuterium Exchange Deep Ultraviolet Resonance Raman Spectroscopy. *J. Am. Chem. Soc.* **2007**, *129* (36), 11002–11003. https://doi.org/10.1021/ja073798w.

(224) Sikirzhytski, V.; Topilina, N. I.; Takor, G. A.; Higashiya, S.; Welch, J. T.; Uversky, V. N.; Lednev, I. K. Fibrillation Mechanism of a Model Intrinsically Disordered Protein Revealed by 2D Correlation Deep UV Resonance Raman Spectroscopy. *Biomacromolecules* **2012**, *13* (5), 1503–1509. https://doi.org/10.1021/bm300193f.

(225) Chi, Z.; Asher, S. A. Ultraviolet Resonance Raman Examination of Horse Apomyoglobin Acid Unfolding Intermediates. *Biochemistry* **1999**, *38* (26), 8196–8203. https://doi.org/10.1021/bi982654e.

(226) Kurouski, D.; Deckert-Gaudig, T.; Deckert, V.; Lednev, I. K. Structure and Composition of Insulin Fibril Surfaces Probed by TERS. *J. Am. Chem. Soc.* **2012**, *134* (32), 13323–13329. https://doi.org/10.1021/ja303263y.

(227) Kurouski, D.; Lauro, W.; Lednev, I. K. Amyloid Fibrils Are "Alive": Spontaneous Refolding from One Polymorph to Another. *Chem. Commun.* **2010**, *46* (24), 4249–4251. https://doi.org/10.1039/B926758A.

(228) Kurouski, D.; Washington, J.; Ozbil, M.; Prabhakar, R.; Shekhtman, A.; Lednev, I. K. Disulfide Bridges Remain Intact While Native Insulin Converts into Amyloid Fibrils. *PLoS ONE* **2012**, *7* (6), e36989. https://doi.org/10.1371/journal.pone.0036989.

(229) Xu, M.; Ermolenkov, V. V.; He, W.; Uversky, V. N.; Fredriksen, L.; Lednev, I. K. Lysozyme Fibrillation: Deep UV Raman Spectroscopic Characterization of

Protein Structural Transformation. *Biopolymers* **2005**, *79* (1), 58–61. https://doi.org/10.1002/bip.20330.

(230) Miura, T.; Takeuchi, H.; Harada, I. Characterization of Individual Tryptophan Side Chains in Proteins Using Raman Spectroscopy and Hydrogen-Deuterium Exchange Kinetics. *Biochemistry* **1988**, *27* (1), 88–94. https://doi.org/10.1021/bi00401a015.

(231) Walsh, M. A.; Schneider, T. R.; Sieker, L. C.; Dauter, Z.; Lamzin, V. S.; Wilson, K. S. Refinement of Triclinic Hen Egg-White Lysozyme at Atomic Resolution. *Acta Cryst* **1998**, D*52* (4), 522–546. https://doi.org/10.1107/S0907444997013656.

(232) Takeuchi, H. Raman Structural Markers of Tryptophan and Histidine Side Chains in Proteins. *Biopolymers* **2003**, *72* (5), 305–317. https://doi.org/10.1002/bip.10440.

(233) Takeuchi, H.; Ohtsuka, Y.; Harada, I. Ultraviolet Resonance Raman Study on the Binding Mode of Enkephalin to Phospholipid Membranes. *J. Am. Chem. Soc.***1992**, *114* (13), 5321–5328. https://doi.org/10.1021/ja00039a049.

(234) Jakubek, R. S.; Handen, J.; White, S. E.; Asher, S. A.; Lednev, I. K. Ultraviolet Resonance Raman Spectroscopic Markers for Protein Structure and Dynamics. *Trends Anal Chem* **2018**, *103*, 223–229. https://doi.org/10.1016/j.trac.2017.12.002.

(235) Hildebrandt, P. G.; Copeland, R. A.; Spiro, T. G.; Otlewski, J.; Laskowski, M.; Prendergast, F. G. Tyrosine Hydrogen-Bonding and Environmental Effects in Proteins Probed by Ultraviolet Resonance Raman Spectroscopy. *Biochemistry* **1988**, *27* (15), 5426–5433. https://doi.org/10.1021/bi00415a007.

(236) Asher, S. A.; Larkin, P. J.; Teraoka, J. Ultraviolet Resonance Raman and Absorption Difference Spectroscopy of Myoglobins: Titration Behavior of Individual Tyrosine Residues. *Biochemistry* **1991**, *30* (24), 5944–5954. https://doi.org/10.1021/bi00238a019.

(237) Imoto, T.; Forster, L. S.; Rupley, J. A.; Tanaka, F. Fluorescence of Lysozyme: Emissions from Tryptophan Residues 62 and 108 and Energy Migration. *Proc. Natl. Acad. Sci. USA* **1972**, *69* (5), 1151–1155. https://doi.org/10.1073/pnas.69.5.1151.

(238) Nishimoto, E.; Yamashita, S.; Yamasaki, N.; Imoto, T. Resolution and Characterization of Tryptophyl Fluorescence of Hen Egg-White Lysozyme by Quenching- and Time-Resolved Spectroscopy. *Biosci. Biotechnol. Biochem.* **1999**, *63* (2), 329–336. https://doi.org/10.1271/bbb.63.329.

(239) Chaari, A.; Fahy, C.; Chevillot-Biraud, A.; Rholam, M. Insights into Kinetics of Agitation-Induced Aggregation of Hen Lysozyme under Heat and Acidic Conditions from Various Spectroscopic Methods. *PLoS ONE* **2015**, *10* (11), e0142095. https://doi.org/10.1371/journal.pone.0142095.

(240) Frare, E.; Laureto, P. P. De; Dobson, C. M.; Fontana, A. A Highly Amyloidogenic Region of Hen Lysozyme. *J. Mol. Biol.* **2004**, 1153–1165. https://doi.org/10.1016/j.jmb.2004.05.056.

(241) Bekard, I. B.; Dunstan, D. E. Tyrosine Autofluorescence as a Measure of Bovine Insulin Fibrillation. *Biophys. J.* **2009**, *97* (9), 2521–2531. https://doi.org/10.1016/j.bpj.2009.07.064.

(242) Hernández, B.; Coïc, Y. M.; Pflüger, F.; Kruglik, S. G.; Ghomi, M. All Characteristic Raman Markers of Tyrosine and Tyrosinate Originate from Phenol Ring Fundamental Vibrations. *J. Raman Spectrosc.* **2016**, *47* (2), 210–220. https://doi.org/10.1002/jrs.4776.

(243) Hildebrandt, P. G.; Copeland, R. A.; Spiro, T. G.; Otlewski, J.; Laskowski, M.; Prendergast, F. G. Tyrosine Hydrogen-Bonding and Environmental Effects in Proteins Probed by Ultraviolet Resonance Raman Spectroscopy. *Biochemistry* **1988**, *27* (15), 5426–5433. https://doi.org/10.1021/bi00415a007.

(244) Hernández, B.; Pflüger, F.; Kruglik, S. G.; Ghomi, M. Characteristic Raman Lines of Phenylalanine Analyzed by a Multiconformational Approach. *J. Raman Spectrosc.* **2013**, *44* (6), 827–833. https://doi.org/10.1002/jrs.4290.

(245) Rava, T.G; Spiro, T.G. Resonance Enhancement in the Ultraviolet Raman Spectra of Aromatic Amino Acids. *J. Phys. Chem.* **1985**, *89*, 1856–1861. https://doi.org/10.1021/j100256a007

(246) Tito, P.; Nettleton, E.J.; Robinsonet C.V.; Dissecting the Hydrogen Exchange Properties of Insulin under Amyloid Fibril Forming Conditions: A Site-specific

Investigation by Mass Spectrometry *J. Mol. Biol.* **2000**, *303* (2), 267–278.https://doi.org/ 10.1006/jmbi.2000.4142

(247) Ivanova, M. I.; Sievers, S. A.; Sawaya, M. R.; Wall, J. S.; Eisenberg, D. Molecular Basis for Insulin Fibril Assembly. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (45), 18990–18995. https://doi.org/10.1073/pnas.0910080106.

(248) Whittingham, J. L.; Scott, D. J.; Chance, K.; Wilson, A.; Finch, J.; Brange, J.; Dodson, G. G. Insulin at PH 2 : Structural Analysis of the Conditions Promoting Insulin Fibre. *J. Mol. Biol.*, **2002**, *318*, 479-490. https://doi.org/10.1016/S0022-2836(02)00021-9.

(249) Alijanvand, S. H.; Christensen, M. H.; Christiansen, G.; Harikandei, K. B.; Salehi, P.; Schiøtt, B.; Moosavi-Movahedi, A. A.; Otzen, D. E. Novel Noscapine Derivatives Stabilize the Native State of Insulin against Fibrillation. *Int. J. Biol. Macromol.* **2020**, *147*, 98–108. https://doi.org/10.1016/j.ijbiomac.2020.01.061.

(250) Memarpoor-Yazdi, M.; Mahaki, H. Probing the Interaction of Human Serum Albumin with Vitamin B2 (Riboflavin) and l-Arginine (l-Arg) Using Multi-Spectroscopic, Molecular Modeling and Zeta Potential Techniques. *J. Lumin.* **2013**, *136*, 150–159. https://doi.org/10.1016/j.jlumin.2012.11.016.

(251) Gasymov, O. K.; Abduragimov, A. R.; Glasgow, B. J. A Simple Model-Free Method for Direct Assessment of Fluorescent Ligand Binding by Linear Spectral Summation. *J. Fluoresc.* **2014**, *24* (1), 231–238. https://doi.org/10.1007/s10895-013-1290-y.

(252) van Rooijen, B. D.; van Leijenhorst-Groener, K. A.; Claessens, M. M. A. E.; Subramaniam, V. Tryptophan Fluorescence Reveals Structural Features of α-Synuclein Oligomers. *J. Mol. Biol.* **2009**, *394* (5), 826–833. https://doi.org/10.1016/j.jmb.2009.10.021.

(253) Poniková, S.; Antošová, A.; Demjén, E.; Sedláková, D.; Marek, J.; Varhač, R.; Gažová, Z.; Sedlák, E. Lysozyme Stability and Amyloid Fibrillization Dependence on Hofmeister Anions in Acidic pH. *J. Biol. Inorg. Chem.* **2015**, *20* (6), 921–933. https://doi.org/10.1007/s00775-015-1276-0.

(254) Ding, F.; Zhao, G.; Huang, J.; Sun, Y.; Zhang, L. Fluorescence Spectroscopic Investigation of the Interaction between Chloramphenicol and Lysozyme. *Eur. J. Med. Chem.* **2009**, *44* (10), 4083–4089. https://doi.org/10.1016/j.ejmech.2009.04.047.

(255) Roqanian, S.; Meratan, A. A.; Ahmadian, S.; Shafizadeh, M.; Ghasemi, A.; Karami, L. Polyphenols Protect Mitochondrial Membrane against Permeabilization Induced by HEWL Oligomers: Possible Mechanism of Action. *Int. J. Biol. Macromol.* **2017**, *103*, 709–720. https://doi.org/10.1016/j.ijbiomac.2017.05.130.

(256) He, J.; Wang, Y.; Chang, A. K.; Xu, L.; Wang, N.; Chong, X.; Li, H.; Zhang, B.; Jones, G. W.; Song, Y. Myricetin Prevents Fibrillogenesis of Hen Egg White Lysozyme. *J. Agric. Food. Chem.* **2014**, *62* (39), 9442–9449. https://doi.org/10.1021/jf5025449.

(257) Roqanian, S.; Meratan, A. A.; Ahmadian, S.; Shafizadeh, M.; Ghasemi, A.; Karami, L. Polyphenols Protect Mitochondrial Membrane against Permeabilization Induced by HEWL Oligomers: Possible Mechanism of Action. *Int. J. Biol. Macromol.* **2017**, *103*, 709–720. https://doi.org/10.1016/j.ijbiomac.2017.05.130.

(258) Mishra, N. K.; Deepak, R. N. V.K.; Sankararamakrishnan, R; Verma, S.; Controlling in Vitro Insulin Amyloidosis with Stable Peptide Conjugates: A Combined Experimental and Computational Study. *J. Phys.Chem.B* **2015**, *119 (50),* 15395–15406..

(259) Gong, H.; He, Z.; Peng, A.; Zhang, X.; Cheng, B.; Sun, Y.; Zheng, L.; Huang, K. Effects of Several Quinones on Insulin Aggregation. *Sci. Rep.* **2014**, *4*, 1–8. https://doi.org/10.1038/srep05648.

(260) Carbonaro, M.; Ripanti, F.; Filabozzi, A.; Minicozzi, V.; Stellato, F.; Placidi, E.; Morante, S.; Di Venere, A.; Nicolai, E.; Postorino, P.; Nucara, A. Human Insulin Fibrillogenesis in the Presence of Epigallocatechin Gallate and Melatonin: Structural Insights from a Biophysical Approach. *Int. J. Biol. Macromol.* **2018**, *115*, 1157–1164. https://doi.org/10.1016/j.ijbiomac.2018.04.134.

(261) Porzoor, A.; Alford, B.; Hügel, H. M.; Grando, D.; Caine, J.; Macreadie, I. Anti-Amyloidogenic Properties of Some Phenolic Compounds. *Biomolecules* **2015**, *5* (2), 505–527. https://doi.org/10.3390/biom5020505.

(262) Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E. EGCG Redirects Amyloidogenic Polypeptides into Unstructured, off-Pathway Oligomers. *Nat. Struct. Mol. Biol.* **2008**, *15* (6), 558–566. https://doi.org/10.1038/nsmb.1437.

(263) Woods, L. A.; Platt, G. W.; Hellewell, A. L.; Hewitt, E. W.; Homans, S. W.; Ashcroft, A. E.; Radford, S. E. Ligand Binding to Distinct States Diverts Aggregation of an Amyloid-Forming Protein. *Nat. Chem. Biol.* **2011**, *7* (10), 730–739. https://doi.org/10.1038/nchembio.635.

(264) Ono, K.; Yamada, M. Antioxidant Compounds Have Potent Anti-Fibrillogenic and Fibril-Destabilizing Effects for Alpha-Synuclein Fibrils in Vitro. *J. Neurochem.* **2006**, *97* (1), 105–115. https://doi.org/10.1111/j.1471-4159.2006.03707.x.

(265) He, J.; Xing, Y.-F.; Huang, B.; Zhang, Y.-Z.; Zeng, C.-M. Tea Catechins Induce the Conversion of Preformed Lysozyme Amyloid Fibrils to Amorphous Aggregates. *J. Agric. Food Chem.* **2009**, *57* (23), 11391–11396. https://doi.org/10.1021/jf902664f.

(266) Zscherp, C.; Aygün, H.; Engels, J. W.; Mäntele, W. Effect of Proline to Alanine Mutation on the Thermal Stability of the All-β-Sheet Protein Tendamistat. *Biochim. Biophys. Acta* **2003**, *1651* (1–2), 139–145. https://doi.org/10.1016/S1570-9639(03)00263-2.

(267) Fabian, H.; Schultz, C-; Backmann, J.; Hahn, U.; Saenger, W.; Henry, I; Mantsch, H.; Naumann, D. Impact of Point Mutations on the Structure and Thermal Stability of Ribonuclease T1 in Aqueous Solution Probed by Fourier Transform Infrared Spectroscopy. *J. Mol Biol* **1993**, 232 (3), 967–981.

(268) Tremmel, S.; Beyermann, M.; Oschkinat, H.; Bienert, M.; Naumann, D.; Fabian, H. Probes for Monitoring Conformational Changes. *Angew. Chem. Int. Ed*. **2005**, 4631–4635. https://doi.org/10.1002/anie.200500547.

(269) Fabian, H.; Schultz, C.; Backmann, J.; Hahn, U.; Saenger, W.; Mantsch, I. H. H.; Naumann, D. Impact of Point Mutations on the Structure and Thermal Stability

of Ribonuclease. *Biochemistry*, **1994**, *33*, 10725–10730. https://doi.org/10.1021/bi00201a021.

(270) Tokunaga, Y.; Sakakibara, Y.; Kamada, Y.; Watanabe, K.; Sugimoto, Y. Analysis of Core Region from Egg White Lysozyme Forming Amyloid Fibrils. *Int. J. Biol. Sci.* **2013**, *9* (2), 219–227. https://doi.org/10.7150/ijbs.5380.

(271) Pantusa, M.; Bartucci, R.; Rizzuti, B. Stability of Trans -Resveratrol Associated with Transport Proteins. *J. Agric. Food Chem* **2014**, *62* (19), 4384–4391. https://doi.org/10.1021/jf405584a.

(272) Chen, Y. Y.; Xiao, L.; Cui, J. H.; Chen, G. F.; Zhang, J.; Wang, P. Biological Relevance of the Interaction Between Resveratrol and Insulin. *Food Biophys.* **2013**, *8* (4), 282–289. https://doi.org/10.1007/s11483-013-9301-8.

(273) Allan, K. E.; Lenehan, A. C. E.; A, A. V. E. UV Light Stability of α -Cyclodextrin / Resveratrol Host – Guest Complexes and Isomer Stability at Varying pH. *Aust. J. Chem*, **2009**, *62*, 921–926. https://doi.org/10.1071/CH08506.

(274) Miura, T.; Takeuchi, H.; Harada, I. Characterization of Individual Tryptophan Side Chains in Proteins Using Raman Spectroscopy and Hydrogen-Deuterium Exchange Kinetics. *Biochemistry* **1988**, *27* (1), 88–94. https://doi.org/10.1021/bi00401a015.

(275) Wang, L.; Berne, B. J.; Friesner, R. A. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein – Ligand Binding Affinities. *Proc. Natl. Acad. Sci. USA* **2012**, *109* (6). https://doi.org/10.1073/pnas.1114017109.

(276) Wang, M.; JiJi, R. D. Resolution of Localized Small Molecule–Aβ Interactions by Deep-Ultraviolet Resonance Raman Spectroscopy. *Biophys. Chem.* **2011**, *158* (2), 96–103. https://doi.org/10.1016/j.bpc.2011.05.017.

(277) Wang, M.; JiJi, R. D. Spectroscopic Detection of β -Sheet Structure in Nascent Aβ Oligomers. *J. Biophoton.* **2011**, n/a-n/a. https://doi.org/10.1002/jbio.201100023.

(278) Chi, Z.; Chen, X. G.; Holtz, J. S. W.; Asher, S. A. Uv Resonance Raman-Selective Amide Vibrational Enhancement: Quantitative Methodology for Determining Protein Secondary Structure. *Biochemistry* **1998**, *37* (9), 2854–2864. https://doi.org/10.1021/bi971160z.

(279) Das, S.; Ghosh, P.; Koley, S.; Singha Roy, A. Binding of Naringin and Naringenin with Hen Egg White Lysozyme: A Spectroscopic Investigation and Molecular Docking Study. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2018**, *192*, 211–221. https://doi.org/10.1016/j.saa.2017.11.015.

(280) Das, S.; Hazarika, Z.; Sarmah, S.; Baruah, K.; Rohman, M. A.; Paul, D.; Jha, A. N.; Singha Roy, A. Exploring the Interaction of Bioactive Kaempferol with Serum Albumin, Lysozyme and Hemoglobin: A Biophysical Investigation Using Multi-Spectroscopic, Docking and Molecular Dynamics Simulation Studies. *J. Photochem. Photobiol. B* **2020**, *205*, 111825. https://doi.org/10.1016/j.jphotobiol.2020.111825.

(281) Frare, E.; Polverino de Laureto, P.; Zurdo, J.; Dobson, C. M.; Fontana, A. A Highly Amyloidogenic Region of Hen Lysozyme. *J. Mol. Biol.* **2004**, *340* (5), 1153–1165. https://doi.org/10.1016/j.jmb.2004.05.056.

(282) Flieger, J.; Tatarczak-Michalewska, M.; Blicharska, E. Characterization of the cis/trans Isomerization of Resveratrol by High-Performance Liquid Chromatography. *Anal. Lett.*, **2016**, *50*, 294-303. https://doi.org/10.1080/00032719.2016.1178756.

(283) Wang, L.; Berne, B. J.; Friesner, R. A. On Achieving High Accuracy and Reliability in the Calculation of Relative Protein-Ligand Binding Affinities. *Proc. Nat.l Acad. Sci. USA* **2012**, *109* (6), 1937–1942. https://doi.org/10.1073/pnas.1114017109.

(284) Jones, B. E.; Matthews, C. R. Early Intermediates in the Folding of Dihydrofolate Reductase from Escherichia Coli Detected by Hydrogen Exchange and NMR. *Protein Sci.* **1995**, *4* (2), 167–177. https://doi.org/10.1002/pro.5560040204.

(285) Boehr, D. D.; McElheny, D.; Dyson, H. J.; Wright, P. E. The Dynamic Energy Landscape of Dihydrofolate Reductase Catalysis. *Science* **2006**, *313* (5793), 1638–1642. https://doi.org/10.1126/science.1130258.

(286) Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E. A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science* **2011**, *332* (6026), 234–238. https://doi.org/10.1126/science.1198542.

(287) Wróbel, A.; Arciszewska, K.; Maliszewski, D.; Drozdowska, D. Trimethoprim and Other Nonclassical Antifolates an Excellent Template for Searching Modifications of Dihydrofolate Reductase Enzyme Inhibitors. *J. Antibiot.* **2020**, *73* (1), 5–27. https://doi.org/10.1038/s41429-019-0240-6.

(288) Huang, F.; Yang, Q.-X.; Huang, T.; Gelbaum, L.; Kuyper, L. F. The Conformations of Trimethoprim/E. Coli Dihydrofolate Reductase Complexes A 15N and 31P NMR Study. *FEBS Letters* **1991**, *283* (1), 44–46. https://doi.org/10.1016/0014-5793(91)80549-I.

(289) Yang, Q. X.; Huang, F. Y.; Huang, T. H.; Gelbaum, L. The Effect of Hydration on the Dynamics of Trimethoprim Bound to Dihydrofolate Reductase. A Deuterium NMR Study. *Biophys. J.***1993**, *64* (4), 1361–1365. https://doi.org/10.1016/S0006-3495(93)81486-3.

(290) Cammarata, M.; Thyer, R.; Lombardo, M.; Anderson, A.; Wright, D.; Ellington, A.; Brodbelt, J. S. Characterization of Trimethoprim Resistant E. Coli Dihydrofolate Reductase Mutants by Mass Spectrometry and Inhibition by Propargyl-Linked Antifolates. *Chem. Sci.* **2017**, *8* (5), 4062–4072. https://doi.org/10.1039/C6SC05235E.

(291) Pegg, A. E.; McCann, P. P. Polyamine Metabolism and Function. *Am. J. Physiol.-Cell Physiol.* **1982**, *243* (5), C212–C221. https://doi.org/10.1152/ajpcell.1982.243.5.C212.

(292) Wallace, H. M.; Fraser, A. V.; Hughes, A. A Perspective of Polyamine Metabolism. *Biochem. J*, **2003**, *376*, 1–14. https://doi.org/10.1042/BJ20031327.

(293) Couling, V. W.; Fischer, P.; Klenerman, D.; Huber, W. Ultraviolet Resonance Raman Study of Drug Binding in Dihydrofolate Reductase, Gyrase, and Catechol O-Methyltransferase. *Biophys. J.* **1998**, *75* (2), 1097–1106. https://doi.org/10.1016/S0006-3495(98)77599-X.

(294) Matthews, D. A.; Bolin, J. T.; Burridge, J. M.; Filman, D. J.; Volz, K. W.; Kaufman, B. T.; Beddell, C. R.; Champness, J. N.; Stammers, D. K.; Kraut, J. Refined Crystal Structures of Escherichia Coli and Chicken Liver Dihydrofolate Reductase Containing Bound Trimethoprim. *J. Biol. Chem.* **1985**, *260* (1), 381–391. https://doi.org/10.1016/S0021-9258(18)89743-5.

(295) Bajorath, J.; Kitson, D. H.; Hagler, A. T.; Kraut, J. The Electrostatic Potential of Escherichia Coli Dihydrofolate Reductase. *Proteins* **1991**, *11* (1), 1–12. https://doi.org/10.1002/prot.340110102.

(296) Fernández, C. O., Hoyer, W., Zweckstetter, M., Jares-Erijman, E. A., Subramaniam, V., Griesinger, C., & Jovin, T. M. NMR of alpha-synuclein-polyamine complexes elucidates the mechanism and kinetics of induced aggregation. *EMBO J.*, **2004,** *23*, 2039–2046. https://doi.org/10.1038/sj.emboj.7600211.

(297) Ohmae, E., Sasaki, Y., & Gekko, K. Effects of five-tryptophan mutations on structure, stability and function of Escherichia coli dihydrofolate reductase. *J. Biochem.*, **2001**, *130*(3), 439–447. https://doi.org/10.1093/oxfordjournals.jbchem.a003004.

(298) Grishina, I. B.; Woody, R. W. Contributions of Tryptophan Side Chains to the Circular Dichroism of Globular Proteins: Exciton Couplets and Coupled Oscillators. *Faraday Discuss.* **1994**, *99* (0), 245–262. https://doi.org/10.1039/FD9949900245.

(299) Kasper, J. R.; Liu, P.-F.; Park, C. Structure of a Partially Unfolded Form of Escherichia Coli Dihydrofolate Reductase Provides Insight into Its Folding Pathway. *Protein Sci.* **2014**, *23* (12), 1728–1737. https://doi.org/10.1002/pro.2555.

(300) Burré, J. The Synaptic Function of α-Synuclein. *JPD* **2015**, *5* (4), 699–713. https://doi.org/10.3233/JPD-150642.

(301) Ramachandran, G.; Milán-Garcés, E. A.; Udgaonkar, J. B.; Puranik, M. Resonance Raman Spectroscopic Measurements Delineate the Structural Changes That Occur during Tau Fibril Formation. *Biochemistry* **2014**, *53* (41), 6550–6565. https://doi.org/10.1021/bi500528x.

(302) Song, S.; Asher, S. A. Internal Intensity Standards for Heme Protein UV Resonance Raman Studies: Excitation Profiles of Cacodylic Acid and Sodium Selenate. *Biochemistry* **1991**, *30* (5), 1199–1205. https://doi.org/10.1021/bi00219a006.

(303) Munishkina, L. A.; Henriques, J.; Uversky, V. N.; Fink, A. L. Role of Protein−Water Interactions and Electrostatics in α-Synuclein Fibril Formation. *Biochemistry* **2004**, *43* (11), 3289–3300. https://doi.org/10.1021/bi034938r.

(304) Mathis, C. A.; Lopresti, B. J.; Ikonomovic, M. D.; Klunk, W. E. Small-Molecule PET Tracers for Imaging Proteinopathies. *Semin. Nucl. Med.* **2017**, *47* (5), 553–575. https://doi.org/10.1053/j.semnuclmed.2017.06.003.

(305) Uversky, V. N. Neuropathology, Biochemistry, and Biophysics of α-Synuclein Aggregation. *J. Neurochem.* **2007**, *103* (1), 17–37. https://doi.org/10.1111/j.1471-4159.2007.04764.x.

(306) Post, M. R.; Lieberman, O. J.; Mosharov, E. V. Can Interactions Between α-Synuclein, Dopamine and Calcium Explain Selective Neurodegeneration in Parkinson's Disease? *Front. Neurosci.* **2018**, *12*. https://doi.org/10.3389/fnins.2018.00161.

(307) Fakhree, M. A. A.; Blum, C.; Claessens, M. M. A. E. Shaping Membranes with Disordered Proteins. *Arch. Biochem. Biophys.* **2019**, *677*, 108163. https://doi.org/10.1016/j.abb.2019.108163.

(308) Drescher, M.; Rooijen, B. D. van; Veldhuis, G.; Subramaniam, V.; Huber, M. A Stable Lipid-Induced Aggregate of α-Synuclein. *J. Am. Chem. Soc.* **2010**, *132* (12), 4080–4082. https://doi.org/10.1021/ja909247j.

(309) Woods, W. S.; Boettcher, J. M.; Zhou, D. H.; Kloepper, K. D.; Hartman, K. L.; Ladror, D. T.; Qi, Z.; Rienstra, C. M.; George, J. M. Conformation-Specific Binding of α-Synuclein to Novel Protein Partners Detected by Phage Display and NMR Spectroscopy. *J. Biol. Chem.* **2007**, *282* (47), 34555–34567. https://doi.org/10.1074/jbc.M705283200.

(310) Theillet, F.-X.; Binolfi, A.; Bekei, B.; Martorana, A.; Rose, H. M.; Stuiver, M.; Verzini, S.; Lorenz, D.; van Rossum, M.; Goldfarb, D.; Selenko, P. Structural Disorder of Monomeric α-Synuclein Persists in Mammalian Cells. *Nature* **2016**, *530* (7588), 45–50. https://doi.org/10.1038/nature16531.

(311) Cattani, J.; Subramaniam, V.; Drescher, M. Room-Temperature in-Cell EPR Spectroscopy: Alpha-Synuclein Disease Variants Remain Intrinsically

Disordered in the Cell. *Phys. Chem. Chem. Phys.* **2017**, *19* (28), 18147–18151. https://doi.org/10.1039/C7CP03432F.

(312) Bartels, T.; Ahlstrom, L. S.; Leftin, A.; Kamp, F.; Haass, C.; Brown, M. F.; Beyer, K. The N-Terminus of the Intrinsically Disordered Protein α-Synuclein Triggers Membrane Binding and Helix Folding. *Biophys. J.* **2010**, *99* (7), 2116–2124. https://doi.org/10.1016/j.bpj.2010.06.035.

(313) Karube, H.; Sakamoto, M.; Arawaka, S.; Hara, S.; Sato, H.; Ren, C.-H.; Goto, S.; Koyama, S.; Wada, M.; Kawanami, T.; Kurita, K.; Kato, T. N-Terminal Region of α-Synuclein Is Essential for the Fatty Acid-Induced Oligomerization of the Molecules. *FEBS Letters* **2008**, *582* (25), 3693–3700. https://doi.org/10.1016/j.febslet.2008.10.001.

(314) Fakhree, M. A. A.; Nolten, I. S.; Blum, C.; Claessens, M. M. A. E. Different Conformational Subensembles of the Intrinsically Disordered Protein α-Synuclein in Cells. *J. Phys. Chem. Lett.* **2018**, *9* (6), 1249–1253. https://doi.org/10.1021/acs.jpclett.8b00092.

(315) Stephens, A. D.; Zacharopoulou, M.; Moons, R.; Fusco, G.; Seetaloo, N.; Chiki, A.; Woodhams, P. J.; Mela, I.; Lashuel, H. A.; Phillips, J. J.; De Simone, A.; Sobott, F.; Schierle, G. S. K. Extent of N-Terminus Exposure of Monomeric Alpha-Synuclein Determines Its Aggregation Propensity. *Nat. Commun.* **2020**, *11* (1), 2820. https://doi.org/10.1038/s41467-020-16564-3.

(316) Ranjan, P.; Kumar, A. Perturbation in Long-Range Contacts Modulates the Kinetics of Amyloid Formation in α-Synuclein Familial Mutants. *ACS Chem. Neurosci.* **2017**, *8* (10), 2235–2246. https://doi.org/10.1021/acschemneuro.7b00149.

(317) McClendon, S.; Rospigliosi, C. C.; Eliezer, D. Charge Neutralization and Collapse of the C-Terminal Tail of Alpha-Synuclein at Low PH. *Protein Sci.* **2009**, *18* (7), 1531–1540. https://doi.org/10.1002/pro.149.

(318) Li, J.; Uversky, V. N.; Fink, A. L. Effect of Familial Parkinson's Disease Point Mutations A30P and A53T on the Structural Properties, Aggregation, and Fibrillation of Human α-Synuclein. *Biochemistry* **2001**, *40* (38), 11604–11613. https://doi.org/10.1021/bi010616g.

(319) Boyer, D. R.; Li, B.; Sun, C.; Fan, W.; Zhou, K.; Hughes, M. P.; Sawaya, M. R.; Jiang, L.; Eisenberg, D. S. The α-Synuclein Hereditary Mutation E46K Unlocks a More Stable, Pathogenic Fibril Structure. *Proc. Natl. Acad. Sci. USA* **2020**, *117* (7), 3592–3602. https://doi.org/10.1073/pnas.1917914117.

(320) Boyer, D. R.; Li, B.; Sun, C.; Fan, W.; Sawaya, M. R.; Jiang, L.; Eisenberg, D. S. Structures of Fibrils Formed by α-Synuclein Hereditary Disease Mutant H50Q Reveal New Polymorphs. *Nat. Struct. Mol. Biol.*, **2019**, *26*, 1044–1052. https://doi.org/10.1038/s41594-019-0322-y

(321) Lázaro, D. F.; Dias, M. C.; Carija, A.; Navarro, S.; Madaleno, C. S.; Tenreiro, S.; Ventura, S.; Outeiro, T. F. The Effects of the Novel A53E Alpha-Synuclein Mutation on Its Oligomerization and Aggregation. *Acta Neuropathol. Commun.* **2016**, *4* (1), 128. https://doi.org/10.1186/s40478-016-0402-8.

(322) Fares, M.-B.; Ait-Bouziad, N.; Dikiy, I.; Mbefo, M. K.; Jovičić, A.; Kiely, A.; Holton, J. L.; Lee, S.-J.; Gitler, A. D.; Eliezer, D.; Lashuel, H. A. The Novel Parkinson's Disease Linked Mutation G51D Attenuates in Vitro Aggregation and Membrane Binding of α-Synuclein, and Enhances Its Secretion and Nuclear Localization in Cells. *Hum. Mol. Genet..* **2014**, *23* (17), 4491–4509. https://doi.org/10.1093/hmg/ddu165.

(323) Villar-Piqué, A.; Fonseca, T. L. da; Sant'Anna, R.; Szegö, É. M.; Fonseca-Ornelas, L.; Pinho, R.; Carija, A.; Gerhardt, E.; Masaracchia, C.; Gonzalez, E. A.; Rossetti, G.; Carloni, P.; Fernández, C. O.; Foguel, D.; Milosevic, I.; Zweckstetter, M.; Ventura, S.; Outeiro, T. F. Environmental and Genetic Factors Support the Dissociation between α-Synuclein Aggregation and Toxicity. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (42), E6506–E6515. https://doi.org/10.1073/pnas.1606791113.

(324) Stephens, A. D.; Zacharopoulou, M.; Kaminski Schierle, G. S. The Cellular Environment Affects Monomeric α-Synuclein Structure. *Trends Biochem. Sci.* **2019**, *44* (5), 453–466. https://doi.org/10.1016/j.tibs.2018.11.005.

(325) Yu, S.; Uéda, K.; Chan, P. α-Synuclein and Dopamine Metabolism. *Mol. Neurobiol.* **2005**, *31* (1), 243–254. https://doi.org/10.1385/MN:31:1-3:243.

(326) Masato, A.; Plotegher, N.; Boassa, D.; Bubacco, L. Impaired Dopamine Metabolism in Parkinson's Disease Pathogenesis. *Mol. Neurodegener.* **2019**, *14*. https://doi.org/10.1186/s13024-019-0332-6.

(327) Shirasaka, M.; Kuwata, K.; Honda, R. α-Synuclein Chaperone Suppresses Nucleation and Amyloidogenesis of Prion Protein. *Biochem. Biophys. Res. Commun.* **2020**, *521* (1), 259–264. https://doi.org/10.1016/j.bbrc.2019.10.120.

(328) Peelaerts, W.; Bousset, L.; Van der Perren, A.; Moskalyuk, A.; Pulizzi, R.; Giugliano, M.; Van den Haute, C.; Melki, R.; Baekelandt, V. α-Synuclein Strains Cause Distinct Synucleinopathies after Local and Systemic Administration. *Nature* **2015**, *522* (7556), 340–344. https://doi.org/10.1038/nature14547.

(329) Pacheco, C.; Aguayo, L. G.; Opazo, C. An Extracellular Mechanism That Can Explain the Neurotoxic Effects of α-Synuclein Aggregates in the Brain. *Front. Physiol.* **2012**, *3*. https://doi.org/10.3389/fphys.2012.00297.

(330) Heise, H.; Hoyer, W.; Becker, S.; Andronesi, O. C.; Riedel, D.; Baldus, M. Molecular-Level Secondary Structure, Polymorphism, and Dynamics of Full-Length -Synuclein Fibrils Studied by Solid-State NMR. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (44), 15871–15876. https://doi.org/10.1073/pnas.0506109102.

(331) Guerrero-Ferreira, R.; Taylor, N. M.; Mona, D.; Ringler, P.; Lauer, M. E.; Riek, R.; Britschgi, M.; Stahlberg, H. Cryo-EM Structure of Alpha-Synuclein Fibrils. *eLife* **2018**, *7*, e36402. https://doi.org/10.7554/eLife.36402.

(332) Ni, X.; McGlinchey, R. P.; Jiang, J.; Lee, J. C. Structural Insights into α-Synuclein Fibril Polymorphism: Effects of Parkinson's Disease-Related C-Terminal Truncations. *J. Mol. Biol.* **2019**, *431* (19), 3913–3919. https://doi.org/10.1016/j.jmb.2019.07.001.

(333) Waxman, E. A.; Giasson, B. I. Specificity and Regulation of Casein Kinase/Mediated Phosphorylation of α-Synuclein. *J. Neuropathol. Exp. Neurol.* **2008**, *67* (5), 402–416. https://doi.org/10.1097/NEN.0b013e3186fc995.

(334) Sato, H.; Arawaka, S.; Hara, S.; Fukushima, S.; Koga, K.; Koyama, S.; Kato, T. Authentically Phosphorylated α-Synuclein at Ser129 Accelerates Neurodegeneration in a Rat Model of Familial Parkinson's Disease. *J. Neurosci.* **2011**, *31* (46), 16884–16894. https://doi.org/10.1523/JNEUROSCI.3967-11.2011.

(335) Oueslati, A.; Paleologou, K. E.; Schneider, B. L.; Aebischer, P.; Lashuel, H. A. Mimicking Phosphorylation at Serine 87 Inhibits the Aggregation of Human α-Synuclein and Protects against Its Toxicity in a Rat Model of Parkinson's Disease. *J. Neurosci.* **2012**, *32* (5), 1536–1544. https://doi.org/10.1523/JNEUROSCI.3784-11.2012.

(336) Xu, Y.; Deng, Y.; Qing, H. The Phosphorylation of α-Synuclein: Development and Implication for the Mechanism and Therapy of the Parkinson's Disease. *J. Neurochem.* **2015**, *135* (1), 4–18. https://doi.org/10.1111/jnc.13234.

(337) Qiao, H.-H.; Zhu, L.-N.; Wang, Y.; Hui, J.-L.; Xie, W.-B.; Liu, C.; Chen, L.; Qiu, P.-M. Implications of Alpha-Synuclein Nitration at Tyrosine 39 in Methamphetamine-Induced Neurotoxicity in Vitro and in Vivo. *Neural Regen. Res.* **2019**, *14* (2), 319. https://doi.org/10.4103/1673-5374.244795.

(338) Danielson, S. R.; Held, J. M.; Schilling, B.; Oo, M.; Gibson, B. W.; Andersen, J. K. Preferentially Increased Nitration of α-Synuclein at Tyrosine-39 in a Cellular Oxidative Model of Parkinson's Disease. *Anal. Chem.* **2009**, *81* (18), 7823–7828. https://doi.org/10.1021/ac901176t.

(339) Rott, R.; Szargel, R.; Shani, V.; Hamza, H.; Savyon, M.; Elghani, F. A.; Bandopadhyay, R.; Engelender, S. SUMOylation and Ubiquitination Reciprocally Regulate α-Synuclein Degradation and Pathological Aggregation. *Proc. Natl. Acad. Sci. USA* **2017**, *114* (50), 13176–13181. https://doi.org/10.1073/pnas.1704351114.

(340) Moon, S. P.; Balana, A. T.; Galesic, A.; Rakshit, A.; Pratt, M. R. Ubiquitination Can Change the Structure of the α-Synuclein Amyloid Fiber in a Site Selective Fashion. *J. Org. Chem.* **2020**, *85* (3), 1548–1555. https://doi.org/10.1021/acs.joc.9b02641.

(341) Z.A. Sorrentino; B.I. Giasson. The Emerging Role of α-Synuclein Truncation in Aggregation and Disease. *J. Biol. Chem.* **2020**, *295*, 10224–10244.

(342) Iyer, A.; Roeters, S. J.; Kogan, V.; Woutersen, S.; Claessens, M. M. A. E.; Subramaniam, V. C-Terminal Truncated α-Synuclein Fibrils Contain Strongly Twisted β-Sheets. *J. Am. Chem. Soc.* **2017**, *139* (43), 15392–15400. https://doi.org/10.1021/jacs.7b07403.

(343) Palomino-Hernandez, O.; Buratti, F. A.; Sacco, P. S.; Rossetti, G.; Carloni, P.; Fernandez, C. O. Role of Tyr-39 for the Structural Features of α-Synuclein and for the Interaction with a Strong Modulator of Its Amyloid Assembly. *Int. J. Mol. Sci.*, **2020**, *21* (14), 5061. https://doi.org/10.3390/ijms21145061.

(344) Zhao, K.; Lim, Y.-J.; Liu, Z.; Long, H.; Sun, Y.; Hu, J.-J.; Zhao, C.; Tao, Y.; Zhang, X.; Li, D.; Li, Y.-M.; Liu, C. Parkinson's Disease-Related Phosphorylation at Tyr39 Rearranges α-Synuclein Amyloid Fibril Structure Revealed by Cryo-EM. *Proc. Natl. Acad. Sci. USA* **2020**, *117* (33), 20305–20315. https://doi.org/10.1073/pnas.1922741117.

(345) Kotzbauer, P. T.; Tu, Z.; Mach, R. H. Current Status of the Development of PET Radiotracers for Imaging Alpha Synuclein Aggregates in Lewy Bodies and Lewy Neurites. *Clin. Transl. Imaging* **2017**, *5* (1), 3–14. https://doi.org/10.1007/s40336-016-0217-4.

(346) Arad, E.; Green, H.; Jelinek, R.; Rapaport, H. Revisiting Thioflavin T (ThT) Fluorescence as a Marker of Protein Fibrillation – The Prominent Role of Electrostatic Interactions. *J. Colloid Interface Sci.* **2020**, *573*, 87–95. https://doi.org/10.1016/j.jcis.2020.03.075.

(347) Plotegher, N.; Kumar, D.; Tessari, I.; Brucale, M.; Munari, F.; Tosatto, L.; Belluzzi, E.; Greggio, E.; Bisaglia, M.; Capaldi, S.; Aioanei, D.; Mammi, S.; Monaco, H. L.; Samo, B.; Bubacco, L. The Chaperone-like Protein 14-3-3η Interacts with Human α-Synuclein Aggregation Intermediates Rerouting the Amyloidogenic Pathway and Reducing α-Synuclein Cellular Toxicity. *Hum. Mol. Gen.* **2014**, *23* (21), 5615–5629. https://doi.org/10.1093/hmg/ddu275.

(348) Plotegher, N.; Berti, G.; Ferrari, E.; Tessari, I.; Zanetti, M.; Lunelli, L.; Greggio, E.; Bisaglia, M.; Veronesi, M.; Girotto, S.; Dalla Serra, M.; Perego, C.; Casella, L.; Bubacco, L. DOPAL Derived Alpha-Synuclein Oligomers Impair Synaptic Vesicles Physiological Function. *Sci. Rep.* **2017**, *7* (1), 40699. https://doi.org/10.1038/srep40699.

(349) Ye, S.; Li, H.; Yang, W.; Luo, Y. Accurate Determination of Interfacial Protein Secondary Structure by Combining Interfacial-Sensitive Amide I and Amide III

Spectral Signals. *J. Am. Chem. Soc.* **2014**, *136* (4), 1206–1209. https://doi.org/10.1021/ja411081t.

(350) Copeland, R. A.; Spiro, T. G. Ultraviolet Resonance Raman Spectra of Cytochrome c Conformational States. *Biochemistry* **1985**, *24* (18), 4960–4968. https://doi.org/10.1021/bi00339a035.

(351) Takeuchi, H.; Watanabe, N.; Satoh, Y.; Harada, I. Effects of Hydrogen Bonding on the Tyrosine Raman Bands in the 1300–1150 Cm−1 Region. *J. Raman Spectrosc.* **1989**, *20* (4), 233–237. https://doi.org/10.1002/jrs.1250200407.

(352) Su, C.; Wang, Y.; Spiro, T. G. Saturation Effects on Ultraviolet Resonance Raman Intensities: Excimer/YAG Laser Comparison and Aromatic Amino Acid Cross-Sections. *J. Raman Spectrosc.* **1990**, *21* (7), 435–440. https://doi.org/10.1002/jrs.1250210709.

(353) Sereda, V.; Lednev, I. K. Polarized Raman Spectroscopy of Aligned Insulin Fibrils: Polarized Raman Spectroscopy of Insulin Fibrils. *J. Raman Spectrosc.* **2014**, *45* (8), 665–671. https://doi.org/10.1002/jrs.4523.

(354) Wang, Y.; Purrello, R.; Jordan, T.; Spiro, T. G. UVRR Spectroscopy of the Peptide Bond. 1. Amide S, a Nonhelical Structure Marker, Is a $C_{\alpha}H$ Bending Mode. *J. Am. Chem. Soc.* **1991**, *113* (17), 6359–6368. https://doi.org/10.1021/ja00017a002.

(355) Vilar, M.; Chou, H.-T.; Luhrs, T.; Maji, S. K.; Riek-Loher, D.; Verel, R.; Manning, G.; Stahlberg, H.; Riek, R. The Fold of -Synuclein Fibrils. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (25), 8637–8642. https://doi.org/10.1073/pnas.0712179105.

(356) Gade Malmos, K.; Blancas-Mejia, L. M.; Weber, B.; Buchner, J.; Ramirez-Alvarado, M.; Naiki, H.; Otzen, D. ThT 101: A Primer on the Use of Thioflavin T to Investigate Amyloid Formation. *Amyloid* **2017**, *24* (1), 1–16. https://doi.org/10.1080/13506129.2017.1304905.

(357) Hsieh, C.-J.; Ferrie, J. J.; Xu, K.; Lee, I.; Graham, T. J. A.; Tu, Z.; Yu, J.; Dhavale, D.; Kotzbauer, P.; Petersson, E. J.; Mach, R. H. Alpha Synuclein Fibrils Contain Multiple Binding Sites for Small Molecules. *ACS Chem. Neurosci.* **2018**, *9* (11), 2521–2527. https://doi.org/10.1021/acschemneuro.8b00177.

(358) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. Structural Reorganization of α-Synuclein at Low PH Observed by NMR and REMD

Simulations. *J. Mol. Biol.* **2009**, *391* (4), 784–796. https://doi.org/10.1016/j.jmb.2009.06.063.

(359) Lewis, K. A.; Su, Y.; Jou, O.; Ritchie, C.; Foong, C.; Hynan, L. S.; White, C. L.; Thomas, P. J.; Hatanpaa, K. J. Abnormal Neurites Containing C-Terminally Truncated α-Synuclein Are Present in Alzheimer's Disease without Conventional Lewy Body Pathology. *Am. J. Pathol.* **2010**, *177* (6), 3037–3050. https://doi.org/10.2353/ajpath.2010.100552.

(360) Mar, C. D.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L. Structure and Properties of α-Synuclein and Other Amyloids Determined at the Amino Acid Level. *Proc. Natl. Acad. Sci. USA* **2005**, *102* (43), 15477–15482. https://doi.org/10.1073/pnas.0507405102.

(361) Stephens, A. D.; Nespovitaya, N.; Zacharopoulou, M.; Kaminski, C. F.; Phillips, J. J.; Kaminski Schierle, G. S. Different Structural Conformers of Monomeric α-Synuclein Identified after Lyophilizing and Freezing. *Anal. Chem.* **2018**, *90* (11), 6975–6983. https://doi.org/10.1021/acs.analchem.8b01264.

(362) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping Long-Range Interactions in α-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127*, 476-477 .

(363) Der-Sarkissian A., Jao, C.C.; Chen, J.; Langen, R. Structural Organization of Alpha-Synuclein Fibrils Studied by Site-Directed Spin Labeling. *J. Biol. Chem.* **2003**, *278*, 37530–37535.

(364) Apetri, M. M.; Maiti, N. C.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E. Secondary Structure of α-Synuclein Oligomers: Characterization by Raman and Atomic Force Microscopy. *J. Mol. Biol.* **2006**, *355* (1), 63–71. https://doi.org/10.1016/j.jmb.2005.10.071.

(365) Kang, L; Moriarty, G.M.; Woods, L.A.; Ashcroft, A.E.; Radford, S.E.; Baum, J; N-terminal acetylation of α-synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Sci*. **2012,** *21*, 911-917.https://doi.org/ 10.1002/pro.2088.

(366) Watson, M. D.; Lee, J. C. N-Terminal Acetylation Affects α-Synuclein Fibril Polymorphism. *Biochemistry* **2019**, *58* (35), 3630–3633. https://doi.org/10.1021/acs.biochem.9b00629.

(367) Pachetti, M.; Marini, B.; Benedetti, F.; Giudici, F.; Mauro, E.; Storici, P.; Masciovecchio, C.; Angeletti, S.; Ciccozzi, M.; Gallo, R. C.; Zella, D.; Ippodrino, R. Emerging SARS-CoV-2 Mutation Hot Spots Include a Novel RNA-Dependent-RNA Polymerase Variant. *J. Transl. Med.* **2020**, *18* (1), 179. https://doi.org/10.1186/s12967-020-02344-6.

(368) Pachetti, M.; Marini, B.; Giudici, F.; Benedetti, F.; Angeletti, S.; Ciccozzi, M.; Masciovecchio, C.; Ippodrino, R.; Zella, D. Impact of Lockdown on Covid-19 Case Fatality Rate and Viral Mutations Spread in 7 Countries in Europe and North America. *J. Transl. Med.* **2020**, *18* (1), 338. https://doi.org/10.1186/s12967-020-02501-x.

(369) Benedetti, F.; Pachetti, M.; Marini, B.; Ippodrino, R.; Gallo, R. C.; Ciccozzi, M.; Zella, D. Inverse Correlation between Average Monthly High Temperatures and COVID-19-Related Death Rates in Different Geographical Areas. *J. Transl. Med.* **2020**, *18* (1), 251. https://doi.org/10.1186/s12967-020-02418-5.

(370) Benedetti, F.; Pachetti, M.; Marini, B.; Ippodrino, R.; Ciccozzi, M.; Zella, D. SARS-CoV-2: March toward Adaptation. *J. Med. Virol.* **2020**, *92* (11), 2274–2276. https://doi.org/10.1002/jmv.26233.

(371) Hu, B.; Guo, H.; Zhou, P.; Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* **2020**, 1–14. https://doi.org/10.1038/s41579-020-00459-7.

(372) Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J. W.; Kim, V. N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181* (4), 914-921.e10. https://doi.org/10.1016/j.cell.2020.04.011.

(373) Baud, D.; Qi, X.; Nielsen-Saines, K.; Musso, D.; Pomar, L.; Favre, G. Real Estimates of Mortality Following COVID-19 Infection. *Lancet Infect. Dis.* **2020**, *20* (7), 773. https://doi.org/10.1016/S1473-3099(20)30195-X.

(374) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan,

W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New Engl. J. Med.* **2020**, *382* (8), 727–733. https://doi.org/10.1056/NEJMoa2001017.

(375) Wang, C.; Liu, Z.; Chen, Z.; Huang, X.; Xu, M.; He, T.; Zhang, Z. The Establishment of Reference Sequence for SARS-CoV-2 and Variation Analysis. *J. Med. Virol.* **2020**, *92* (6), 667–674. https://doi.org/10.1002/jmv.25762.

(376) Domingo, E.; Holland, J. J. Rna Virus Mutations and Fitness for Survival. *Annu. Rev. Microbiol.* **1997**, *51* (1), 151–178. https://doi.org/10.1146/annurev.micro.51.1.151.

(377) Domingo, E. Viruses at the Edge of Adaptation. *Virology* **2000**, *270* (2), 251–253. https://doi.org/10.1006/viro.2000.0320.

(378) Domingo, E. Quasispecies Theory in Virology. *J. Virol.* **2002**, *76* (1), 463–465. https://doi.org/10.1128/JVI.76.1.463-465.2002.

(379) Bruenn, J. A. A Structural and Primary Sequence Comparison of the Viral RNA-dependent RNA Polymerases. *Nucleic Acids Res.* **2003**, *31* (7), 1821–1829. https://doi.org/10.1093/nar/gkg277.

(380) Venkataraman, S.; Prasad, B. V. L. S.; Selvarajan, R. RNA Dependent RNA Polymerases: Insights from Structure, Function and Evolution. *Viruses* **2018**, *10* (2), 76. https://doi.org/10.3390/v10020076.

(381) Jia, H.; Gong, P. A Structure-Function Diversity Survey of the RNA-Dependent RNA Polymerases From the Positive-Strand RNA Viruses. *Front. Microbiol.* **2019**, *10*. https://doi.org/10.3389/fmicb.2019.01945.

(382) Ziebuhr, J. The Coronavirus Replicase. In *Coronavirus Replication and Reverse Genetics*; Current Topics in Microbiology and Immunology; Springer: Berlin, Heidelberg, 2005; pp 57–94. https://doi.org/10.1007/3-540-26765-4_3.

(383) Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M.-L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.-Z. A New Coronavirus Associated with Human Respiratory Disease in China. *Nature* **2020**, *579* (7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3.

(384) Gao, Y.; Yan, L.; Huang, Y.; Liu, F.; Zhao, Y.; Cao, L.; Wang, T.; Sun, Q.; Ming, Z.; Zhang, L.; Ge, J.; Zheng, L.; Zhang, Y.; Wang, H.; Zhu, Y.; Zhu, C.; Hu, T.; Hua,

T.; Zhang, B.; Yang, X.; Li, J.; Yang, H.; Liu, Z.; Xu, W.; Guddat, L. W.; Wang, Q.; Lou, Z.; Rao, Z. Structure of the RNA-Dependent RNA Polymerase from COVID-19 Virus. *Science* **2020**, *368* (6492), 779–782. https://doi.org/10.1126/science.abb7498.

(385) Ma, Y.; Wu, L.; Shaw, N.; Gao, Y.; Wang, J.; Sun, Y.; Lou, Z.; Yan, L.; Zhang, R.; Rao, Z. Structural Basis and Functional Analysis of the SARS Coronavirus Nsp14–Nsp10 Complex. *Proc. Natl. Acad. Sci.. USA* **2015**, *112* (30), 9436–9441. https://doi.org/10.1073/pnas.1508686112.

(386) Ogando, N. S.; Ferron, F.; Decroly, E.; Canard, B.; Posthuma, C. C.; Snijder, E. J. The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity. *Front. Microbiol.* **2019**, *10*. https://doi.org/10.3389/fmicb.2019.01813.

(387) Eckerle, L. D.; Becker, M. M.; Halpin, R. A.; Li, K.; Venter, E.; Lu, X.; Scherbakova, S.; Graham, R. L.; Baric, R. S.; Stockwell, T. B.; Spiro, D. J.; Denison, M. R. Infidelity of SARS-CoV Nsp14-Exonuclease Mutant Virus Replication Is Revealed by Complete Genome Sequencing. *PLoS Pathogens* **2010**, *6* (5), e1000896. https://doi.org/10.1371/journal.ppat.1000896.

(388) Kirchdoerfer, R. N.; Ward, A. B. Structure of the SARS-CoV Nsp12 Polymerase Bound to Nsp7 and Nsp8 Co-Factors. *Nat. Commun.* **2019**, *10* (1), 2342. https://doi.org/10.1038/s41467-019-10280-3.

(389) Furuta, Y.; Gowen, B. B.; Takahashi, K.; Shiraki, K.; Smee, D. F.; Barnard, D. L. Favipiravir (T-705), a Novel Viral RNA Polymerase Inhibitor. *Antivir. Res.* **2013**, *100* (2), 446–454. https://doi.org/10.1016/j.antiviral.2013.09.015.

(390) Lim, S.-Y.; Osuna, C.; Lakritz, J.; Chen, E.; Yoon, G.; Taylor, R.; MacLennan, S.; Leonard, M.; Giuliano, E.; Mathis, A.; Berger, E.; Babu, Y.; Sheridan, W.; Whitney, J. Galidesivir, a Direct-Acting Antiviral Drug, Abrogates Viremia in Rhesus Macaques Challenged with Zika Virus. *Open Forum Infect. Dis.* **2017**, *4* (suppl_1), S55–S55. https://doi.org/10.1093/ofid/ofx162.129.

(391) Agostini, M. L.; Andres, E. L.; Sims, A. C.; Graham, R. L.; Sheahan, T. P.; Lu, X.; Smith, E. C.; Case, J. B.; Feng, J. Y.; Jordan, R.; Ray, A. S.; Cihlar, T.; Siegel, D.; Mackman, R. L.; Clarke, M. O.; Baric, R. S.; Denison, M. R. Coronavirus

Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *mBio* **2018**, *9* (2). https://doi.org/10.1128/mBio.00221-18.

(392) Morgenstern, B.; Michaelis, M.; Baer, P. C.; Doerr, H. W.; Cinatl, J. Ribavirin and Interferon-β Synergistically Inhibit SARS-Associated Coronavirus Replication in Animal and Human Cell Lines. *Biochem. Biophys. Res. Commun* **2005**, *326* (4), 905–908. https://doi.org/10.1016/j.bbrc.2004.11.128.

(393) Chang, Y.-C.; Tung, Y.-A.; Lee, K.-H.; Chen, T.-F.; Hsiao, Y.-C.; Chang, H.-C.; Hsieh, T.-T.; Su, C.-H.; Wang, S.-S.; Yu, J.-Y.; Shih, S.; Lin, Y.-H.; Lin, Y.-H.; Tu, Y.-C. E.; Tung, C.-W.; Chen, C.-Y. Potential Therapeutic Agents for COVID-19 Based on the Analysis of Protease and RNA Polymerase Docking. *Preprints* **2020**. https://doi.org/10.20944/preprints202002.0242.v1.

(394) Ruan, Z.; Liu, C.; Guo, Y.; He, Z.; Huang, X.; Jia, X.; Yang, T. Potential Inhibitors Targeting RNA-Dependent RNA Polymerase Activity (NSP12) of SARS-CoV-2. *Preprints* **2020**. https://doi.org/10.20944/preprints202003.0024.v1.

(395) Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; Cui, J.; Lu, J. On the Origin and Continuing Evolution of SARS-CoV-2. *Natl. Sci. Rev.* **2020**, *7* (6), 1012–1023. https://doi.org/10.1093/nsr/nwaa036.

(396) Shen, Z.; Xiao, Y.; Kang, L.; Ma, W.; Shi, L.; Zhang, L.; Zhou, Z.; Yang, J.; Zhong, J.; Yang, D.; Guo, L.; Zhang, G.; Li, H.; Xu, Y.; Chen, M.; Gao, Z.; Wang, J.; Ren, L.; Li, M. Genomic Diversity of Severe Acute Respiratory Syndrome–Coronavirus 2 in Patients With Coronavirus Disease 2019. *Clin. Infect. Dis* **2020**, *71* (15), 713–720. https://doi.org/10.1093/cid/ciaa203.

(397) Phan, T. Genetic Diversity and Evolution of SARS-CoV-2. *Infect. Genet. Evol.* **2020**, *81*, 104260. https://doi.org/10.1016/j.meegid.2020.104260.

(398) Pfeiffer, J. K.; Kirkegaard, K. A Single Mutation in Poliovirus RNA-Dependent RNA Polymerase Confers Resistance to Mutagenic Nucleotide Analogs via Increased Fidelity. *Proc. Natl. Acad. Sci. USA* **2003**, *100* (12), 7289–7294. https://doi.org/10.1073/pnas.1232294100.

(399) Zhai, Y.; Sun, F.; Li, X.; Pang, H.; Xu, X.; Bartlam, M.; Rao, Z. Insights into SARS-CoV Transcription and Replication from the Structure of the Nsp7–Nsp8

Hexadecamer. *Nat. Struct. Mol. Biol.* **2005**, *12* (11), 980–986. https://doi.org/10.1038/nsmb999.

(400) Li, G.; Clercq, E. D. Therapeutic Options for the 2019 Novel Coronavirus (2019-NCoV). *Nat. Rev. Drug. Discov.* **2020**, *19* (3), 149–150. https://doi.org/10.1038/d41573-020-00016-0.

(401) Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; Tummino, T. A.; Hüttenhain, R.; Kaake, R. M.; Richards, A. L.; Tutuncuoglu, B.; Foussard, H.; Batra, J.; Haas, K.; Modak, M.; Kim, M.; Haas, P.; Polacco, B. J.; Braberg, H.; Fabius, J. M.; Eckhardt, M.; Soucheray, M.; Bennett, M. J.; Cakir, M.; McGregor, M. J.; Li, Q.; Meyer, B.; Roesch, F.; Vallet, T.; Mac Kain, A.; Miorin, L.; Moreno, E.; Naing, Z. Z. C.; Zhou, Y.; Peng, S.; Shi, Y.; Zhang, Z.; Shen, W.; Kirby, I. T.; Melnyk, J. E.; Chorba, J. S.; Lou, K.; Dai, S. A.; Barrio-Hernandez, I.; Memon, D.; Hernandez-Armenta, C.; Lyu, J.; Mathy, C. J. P.; Perica, T.; Pilla, K. B.; Ganesan, S. J.; Saltzberg, D. J.; Rakesh, R.; Liu, X.; Rosenthal, S. B.; Calviello, L.; Venkataramanan, S.; Liboy-Lugo, J.; Lin, Y.; Huang, X.-P.; Liu, Y.; Wankowicz, S. A.; Bohn, M.; Safari, M.; Ugur, F. S.; Koh, C.; Savar, N. S.; Tran, Q. D.; Shengjuler, D.; Fletcher, S. J.; O'Neal, M. C.; Cai, Y.; Chang, J. C. J.; Broadhurst, D. J.; Klippsten, S.; Sharp, P. P.; Wenzell, N. A.; Kuzuoglu-Ozturk, D.; Wang, H.-Y.; Trenker, R.; Young, J. M.; Cavero, D. A.; Hiatt, J.; Roth, T. L.; Rathore, U.; Subramanian, A.; Noack, J.; Hubert, M.; Stroud, R. M.; Frankel, A. D.; Rosenberg, O. S.; Verba, K. A.; Agard, D. A.; Ott, M.; Emerman, M.; Jura, N.; von Zastrow, M.; Verdin, E.; Ashworth, A.; Schwartz, O.; d'Enfert, C.; Mukherjee, S.; Jacobson, M.; Malik, H. S.; Fujimori, D. G.; Ideker, T.; Craik, C. S.; Floor, S. N.; Fraser, J. S.; Gross, J. D.; Sali, A.; Roth, B. L.; Ruggero, D.; Taunton, J.; Kortemme, T.; Beltrao, P.; Vignuzzi, M.; García-Sastre, A.; Shokat, K. M.; Shoichet, B. K.; Krogan, N. J. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* **2020**, *583* (7816), 459–468. https://doi.org/10.1038/s41586-020-2286-9.

(402) Wang, M.; Cao, R.; Zhang, L.; Yang, X.; Liu, J.; Xu, M.; Shi, Z.; Hu, Z.; Zhong, W.; Xiao, G. Remdesivir and Chloroquine Effectively Inhibit the Recently Emerged

Novel Coronavirus (2019-NCoV) in Vitro. *Cell Res.* **2020**, *30* (3), 269–271. https://doi.org/10.1038/s41422-020-0282-0.

(403) Najjar, M.; Suebsuwong, C.; Ray, S. S.; Thapa, R. J.; Maki, J. L.; Nogusa, S.; Shah, S.; Saleh, D.; Gough, P. J.; Bertin, J.; Yuan, J.; Balachandran, S.; Cuny, G. D.; Degterev, A. Structure Guided Design of Potent and Selective Ponatinib-Based Hybrid Inhibitors for RIPK1. *Cell Rep.* **2015**, *10* (11), 1850–1860. https://doi.org/10.1016/j.celrep.2015.02.052.

(404) Young, K.-C.; Lindsay, K. L.; Lee, K.-J.; Liu, W.-C.; He, J.-W.; Milstein, S. L.; Lai, M. M. C. Identification of a Ribavirin-Resistant NS5B Mutation of Hepatitis C Virus during Ribavirin Monotherapy. *Hepatology* **2003**, *38* (4), 869–878. https://doi.org/10.1002/hep.1840380413.

(405) Goldhill, D. H.; Velthuis, A. J. W. te; Fletcher, R. A.; Langat, P.; Zambon, M.; Lackenby, A.; Barclay, W. S. The Mechanism of Resistance to Favipiravir in Influenza. *Proc. Natl. Acad. Sci. USA* **2018**, *115* (45), 11613–11618. https://doi.org/10.1073/pnas.1811345115.

(406) Ksiazek, T. G.; Erdman, D.; Goldsmith, C. S.; Zaki, S. R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J. A.; Lim, W.; Rollin, P. E.; Dowell, S. F.; Ling, A.-E.; Humphrey, C. D.; Shieh, W.-J.; Guarner, J.; Paddock, C. D.; Rota, P.; Fields, B.; DeRisi, J.; Yang, J.-Y.; Cox, N.; Hughes, J. M.; LeDuc, J. W.; Bellini, W. J.; Anderson, L. J.; SARS Working Group. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *New Engl. J. Med.* **2003**, *348* (20), 1953–1966. https://doi.org/10.1056/NEJMoa030781.

(407) Kuiken, T.; Fouchier, R. A.; Schutten, M.; Rimmelzwaan, G. F.; Amerongen, G. van; Riel, D. van; Laman, J. D.; Jong, T. de; Doornum, G. van; Lim, W.; Ling, A. E.; Chan, P. K.; Tam, J. S.; Zambon, M. C.; Gopal, R.; Drosten, C.; Werf, S. van der; Escriou, N.; Manuguerra, J.-C.; Stöhr, K.; Peiris, J. S. M.; Osterhaus, A. D. Newly Discovered Coronavirus as the Primary Cause of Severe Acute Respiratory Syndrome. *Lancet* **2003**, *362* (9380), 263–270. https://doi.org/10.1016/S0140-6736(03)13967-0.

(408) Drosten, C.; Günther, S.; Preiser, W.; van der Werf, S.; Brodt, H.-R.; Becker, S.; Rabenau, H.; Panning, M.; Kolesnikova, L.; Fouchier, R. A. M.; Berger, A.;

Burguière, A.-M.; Cinatl, J.; Eickmann, M.; Escriou, N.; Grywna, K.; Kramme, S.; Manuguerra, J.-C.; Müller, S.; Rickerts, V.; Stürmer, M.; Vieth, S.; Klenk, H.-D.; Osterhaus, A. D. M. E.; Schmitz, H.; Doerr, H. W. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *New Engl. J. Med.* **2003**, *348* (20), 1967–1976. https://doi.org/10.1056/NEJMoa030747.

(409) Groot, R. J. de; Baker, S. C.; Baric, R. S.; Brown, C. S.; Drosten, C.; Enjuanes, L.; Fouchier, R. A. M.; Galiano, M.; Gorbalenya, A. E.; Memish, Z. A.; Perlman, S.; Poon, L. L. M.; Snijder, E. J.; Stephens, G. M.; Woo, P. C. Y.; Zaki, A. M.; Zambon, M.; Ziebuhr, J. Commentary: Middle East Respiratory Syndrome Coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. *J. Virol.* **2013**, *87* (14), 7790–7792. https://doi.org/10.1128/JVI.01244-13.

(410) Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; Yu, T.; Xia, J.; Wei, Y.; Wu, W.; Xie, X.; Yin, W.; Li, H.; Liu, M.; Xiao, Y.; Gao, H.; Guo, L.; Xie, J.; Wang, G.; Jiang, R.; Gao, Z.; Jin, Q.; Wang, J.; Cao, B. Clinical Features of Patients Infected with 2019 Novel Coronavirus in Wuhan, China. *Lancet* **2020**, *395* (10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.

(411) Battegay, M.; Kuehl, R.; Tschudin-Sutter, S.; Hirsch, H. H.; Widmer, A. F.; Neher, R. A. 2019-Novel Coronavirus (2019-NCoV): Estimating the Case Fatality Rate – a Word of Caution. *Swiss Med. Wkly* **2020**, *150* (0506). https://doi.org/10.4414/smw.2020.20203.

(412) Li, Q.; Guan, X.; Wu, P.; Wang, X.; Zhou, L.; Tong, Y.; Ren, R.; Leung, K. S. M.; Lau, E. H. Y.; Wong, J. Y.; Xing, X.; Xiang, N.; Wu, Y.; Li, C.; Chen, Q.; Li, D.; Liu, T.; Zhao, J.; Liu, M.; Tu, W.; Chen, C.; Jin, L.; Yang, R.; Wang, Q.; Zhou, S.; Wang, R.; Liu, H.; Luo, Y.; Liu, Y.; Shao, G.; Li, H.; Tao, Z.; Yang, Y.; Deng, Z.; Liu, B.; Ma, Z.; Zhang, Y.; Shi, G.; Lam, T. T. Y.; Wu, J. T.; Gao, G. F.; Cowling, B. J.; Yang, B.; Leung, G. M.; Feng, Z. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New Engl. J. Med.* **2020**. https://doi.org/10.1056/NEJMoa2001316.

(413) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; Hastie, K. M.; Parker,

M. D.; Partridge, D. G.; Evans, C. M.; Freeman, T. M.; Silva, T. I. de; Angyal, A.; Brown, R. L.; Carrilero, L.; Green, L. R.; Groves, D. C.; Johnson, K. J.; Keeley, A. J.; Lindsey, B. B.; Parsons, P. J.; Raza, M.; Rowland-Jones, S.; Smith, N.; Tucker, R. M.; Wang, D.; Wyles, M. D.; McDanal, C.; Perez, L. G.; Tang, H.; Moon-Walker, A.; Whelan, S. P.; LaBranche, C. C.; Saphire, E. O.; Montefiori, D. C. Tracking Changes in SARS-CoV-2 Spike: Evidence That D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182* (4), 812-827.e19. https://doi.org/10.1016/j.cell.2020.06.043.

(414) Chan, W.-E.; Chuang, C.-K.; Yeh, S.-H.; Chang, M.-S.; Chen, S. S.-L. Functional Characterization of Heptad Repeat 1 and 2 Mutants of the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus. *J. Virol.* **2006**, *80* (7), 3225–3237. https://doi.org/10.1128/JVI.80.7.3225-3237.2006.

(415) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581* (7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5.

(416) Bosch, B. J.; Martina, B. E. E.; Zee, R. van der; Lepault, J.; Haijema, B. J.; Versluis, C.; Heck, A. J. R.; Groot, R. de; Osterhaus, A. D. M. E.; Rottier, P. J. M. Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Infection Inhibition Using Spike Protein Heptad Repeat-Derived Peptides. *Proc. Natl. Acad. Sci. USA* **2004**, *101* (22), 8455–8460. https://doi.org/10.1073/pnas.0400576101.

(417) Xia, S.; Liu, M.; Wang, C.; Xu, W.; Lan, Q.; Feng, S.; Qi, F.; Bao, L.; Du, L.; Liu, S.; Qin, C.; Sun, F.; Shi, Z.; Zhu, Y.; Jiang, S.; Lu, L. Inhibition of SARS-CoV-2 (Previously 2019-NCoV) Infection by a Highly Potent Pan-Coronavirus Fusion Inhibitor Targeting Its Spike Protein That Harbors a High Capacity to Mediate Membrane Fusion. *Cell Res.* **2020**, *30* (4), 343–355. https://doi.org/10.1038/s41422-020-0305-x.

List of Papers

(1) Katava, M.; Stirnemann, G.; Zanatta, M.; Capaccioli, S.; Pachetti, M.; Ngai, K. L.; Sterpone, F.; Paciaroni, A. Critical structural fluctuations of proteins upon thermal unfolding challenge the Lindemann criterion, *Proc. Natl. Acad. Sci. U.S.A* **2017**, *114*, 9361-9366. https://doi.org/10.1073/pnas.1707357114

2) Garoli, D., Schirato, A., Giovannini, G., Cattarin, S., Ponzellini, P., Calandrini, E., Proietti Zaccaria, R., D'Amico, F., Pachetti, M., Yang, W., Jin, H. J., Krahne, R., & Alabastri, A. Galvanic Replacement Reaction as a Route to Prepare Nanoporous Aluminum for UV Plasmonics. *Nanomaterials* **2020**, 10, 102. https://doi.org/10.3390/nano10010102

3) Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., Masciovecchio, C., Angeletti, S., Ciccozzi, M., Gallo, R. C., Zella, D., & Ippodrino, R. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.*, **2020,** *18*, 179. https://doi.org/10.1186/s12967-020-02344-6 (2020).

4) Pachetti, M., Marini, B., Giudici, F., Benedetti, F., Angeletti, S., Ciccozzi, M., Masciovecchio, C., Ippodrino, R., & Zella, D. Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. *J. Transl. Med.*, **2020**, *18*, 338. https://doi.org/10.1186/s12967-020-02501-x

5) Benedetti, F., Pachetti, M., Marini, B., Ippodrino, R., Ciccozzi, M., & Zella, D. SARS-CoV-2: March toward adaptation. *J. Med. Virol.,* **2020**, *92*, 2274–2276. https://doi.org/10.1002/jmv.26233.

6) Benedetti, F., Pachetti, M., Marini, B., Ippodrino, R., Gallo, R. C., Ciccozzi, M., & Zella, D. Inverse correlation between average monthly high temperatures and COVID-19-related death rates in different geographical areas. *J. Transl. Med.*, **2020**, *18*, 251. https://doi.org/10.1186/s12967-020-02418-5.

7) Perissinotto, F.; Senigagliesi, B.; Vaccari, L.; Pachetti, M.; D'Amico, F.; Amenitsch, H.; Sartori, B.; Pachler, K.; Mayr, M.; Gimona, M.; Rohde, E.; Caponnetto, F.; Cesselli, D.; Casalis, L.; Parisse, P. Multi-technique analysis of extracellular vesicles: not only size matters, Chapter in *Advances in Biomembranes and Lipid Self-Assembly*, **2020**, vol. 32, 157-177. https://doi.org/10.1016/bs.abl.2020.05.004

8) D'Amico, F.; Zucchiatti, P.; Latella, K.; Pachetti, M.; Gessini, A.; Masciovecchio C., Vaccari, L.; Pascolo, L. Investigation of genomic DNA methylation by UV Resonant Raman spectroscopy *J. Biophot*. **2020**, 13. https://doi.org/10.1002/jbio.202000150

9) Pachetti, M., Zupin, L., Venturin, I., Mitri, E., Boscolo, R., D'Amico, F., Vaccari, L., Crovella, S., Ricci, G., & Pascolo, L. FTIR Spectroscopy to Reveal Lipid and Protein Changes Induced on Sperm by Capacitation: Bases for an Improvement of Sample Selection in ART. *Int. J. Mol. Sci.*, **2020**, *21*, 8659. https://doi.org/10.3390/ijms21228659.

10) Musto, M.; Parisse, P.; Pachetti, M.; Memo, C.; Di Mauro, G.; Kostarelos, K.; Casalis, L.; & Ballerini, L. Shedding of Physiological Membrane Vesicles Induced by Graphene Oxide Nanoflakes in Brain Cultured Astrocytes, *Carbon*, **2021**, *176*, 458-469. https://doi.org/10.1016/j.carbon.2021.01.142

11) Pachetti, M.; D'Amico, F.; Pucciarelli, S.; Gessini, A.; Parisse, P.; Vaccari, L.; Masciovecchio, C. UV Resonance Raman as an alternative tool to shed light into protein structural modification upon fibrillation and ligand-interaction, *in peer-review*.

12) Katava, M.; Stirnemann, G.; Pachetti, M.; Capaccioli, S.; Paciaroni, A.; Sterpone, F. Specific interactions and environment floppiness tune protein stability under extreme crowding, *in peer-review*.

Acknowledgments

Alla fine di ogni viaggio, sia esso fisico, professionale o emotivo, ci troviamo inevitabilmente a ricordare, a rivivere momenti salienti del percorso, a ricordare le persone che ti hanno fatto crescere e a ripercorrere nuovamente scelte già prese.

Il mio bagaglio professionale e umano si è accresciuto notevolmente in questi anni, coltivando la resilienza, l'accettazione ma anche sentendo una forte spinta interiore di superare le iniziali difficoltà.

Durante questi tre anni, ho imparato a capire di più il mondo scientifico e le persone che vi gravitano intorno, ho affinato quello che per me significa "fare scienza" e costruito una mia integrità scientifica. Ho sempre cercato di fare la differenza, rimanendo però conscia che le grandi sfide si combattono in gruppo.

La nascita di questa consapevolezza è conseguente alla grande indipendenza di cui ho goduto e alla possibilità di esprimermi liberamente in campo scientifico. Di questo ringrazio di cuore il mio relatore, il prof. Claudio Masciovecchio che, con lungimiranza, mi ha lasciato libera espressione e movimento sia nella ricerca che nella costruzione di un mio piccolo network di collaborazioni scientifiche.

Un altro sincero ringraziamento lo devo al dr. Francesco D'Amico che è stata la prima persona ad accogliermi nella linea dove ho lavorato, insegnandomi come essere indipendente sull'uso dei macchinari e come rapportarmi nei confronti dei quesiti scientifici. Lo ringrazio per la sua disponibilità e bontà d'animo, nonché la sua competenza scientifica.

Ringrazio i miei genitori e mia sorella per il supporto emotivo che mi hanno sempre dato, per aver creduto in me e per l'energia che mi hanno sempre trasmesso nell'andare avanti e farmi conoscere, nel cogliere le opportunità. Ringrazio Rudy, la mia nuova famiglia, per avermi supportato, per aver creduto immensamente in me e per essere un grande esempio di umanità e profondo genio.

Il mio cuore va anche ai miei amici toscani e triestini che hanno saputo scuotermi e accarezzarmi allo stesso tempo, facendomi sempre sentire il loro affetto.

Ringrazio anche Ulisse BioMed S.p.A e la dott.ssa Bruna Marini per avermi permesso di collaborare con loro e per avermi insegnato così tanto, professionalmente ed umanamente, facendomi conoscere il brivido delle sfide aziendali e di cosa significa traslare la ricerca scientifica nella creazione di prodotti fruibili.
Non posso che ricordare a tutti quelli che avranno per le mani queste pagine, di inseguire sempre i propri sogni, di andare oltre le cadute, di investire in sé stessi e nelle persone che la vita vi porrà davanti.

Maria Pachetti