Tesi di Dottorato di ANDREA PIZZI
Matricola 821294

POLITECNICO
MILANO 1863

**DIPARTIMENTO
DI
CHIMICA, MATERIALI
E INGEGNERIA CHIMICA**

"Giulio Natta"

HALOGEN BONDING AS A NEW SUPRAMOLECULAR TOOL TO CONTROL PROTEIN AND PEPTIDE SELF-ASSEMBLY

**Dottorato di Ricerca in Chimica
Industriale e Ingegneria Chimica
(CII)**

**PhD in Industrial Chemistry and
Chemical Engineering**

XXIX ciclo

2013 - 2016

Coordinatore: prof. Alessio Frassoldati

Tutore: prof. Enrico Tronconi

Relatore: prof. Pierangelo Metrangolo

Table of Contents

Chapter 3: Halogenation as a powerful strategy to tune the self-assembly of an amyloid peptide fragment 63

Abstract

More than 5000 naturally occurring halogenated compounds have been discovered up to date. Halogenation usually occurs at the periphery of molecules and is a minimal structural modification. However, it can induce deep changes in the molecular properties because of the peculiar chemical features of halogen atoms. Among these features, their tendency to function as electrophilic species has recently attracted wide interest. In particular, halogen atoms may behave as donors of halogen bonding (XB), which is a strong, specific, and directional noncovalent interaction. This thesis highlights the potential role of XB in the fine tuning of self-assembly involving biomolecules containing halogenated amino acids.

The first part of the thesis describes how halogenation improves the self-assembly of a well-known phenylalanine-based organic gelator: *N*-Fmoc-Phe. The hydrogel formation ability and properties of monohalogenated derivatives of *N*-Fmoc-4-X-Phe (X = F, Cl, Br, I) has been related for the first time to halogen atom polarizability, *i.e.*, the propensity to function as XB-donor. In fact, iodinated and brominated derivatives were found to form the strongest gels, and the XB occurrence was confirmed by single crystal structure analysis.

The second part of the thesis shows the impact of halogenation on the self-assembly behavior of a pentapeptide. The segment KLVFF, core sequence of the amyloid beta (Aβ) protein, has been modified at the *para* position of the phenylalanine residues obtaining seven different halogenated derivatives. It has been found that the nature of the halogen atoms, their number and position in the sequence affect the peptide nanostructures obtained in solution. By single point mutation, *i.e.*, halogen for hydrogen replacement, we could control the selective obtainment of four various nanostructures. Single crystal X-ray analysis demonstrated the potential role of XB in controlling the self-assembly of these kind of halogenated peptides. Our findings demonstrate for the first time XB as a new tool to carefully design peptide nanostructures. In a more biological context, these results may also shed new light on the impact of halogenation *in vivo* and diseases thereof, as a consequence of oxidative stress mechanisms.

These studies were corroborated by extensive crystal structure determinations, which is rather unusual for amyloidogenic peptides. This may be related to the presence of halogen atoms in the sequences, which promotes peptide crystallization along with phase determination. These crystal structures confirm the overall features of the amyloid cross-β spine, showing in detail the key noncovalent interactions driving the self-assembly of the peptides. The obtained crystal structures may contribute to further enlighten the amyloidogenic self-assembly of the widely studied KLVFF segment and related sequences.

Chapter 1

Introduction

1.1 Naturally occurring halogenated compounds

The term “halogen”, proposed in 1842 by the Swedish chemist Jöns Jakob Berzelius, intrinsically encompasses the peculiar property of the five elements belonging to the group 17 of the periodic table, which is high reactivity. The etymology of the word combines the Ancient Greek term *ἅλς* (háls, “salt”) and the suffix *-gen* (“produce”),[1] since a reaction involving an element of this group (fluorine, chlorine, bromine, iodine, astatine) with a metal “produces a salt”. Because of their remarkable electronegativity, halogens spread in nature mainly as halides. Chloride, fluoride and bromide salts are abundant both in the oceans and in the Earth's crust (Table 1). Although their abundance is not comparable to other elements, halogen atoms enrich a wide array of naturally produced scaffolds. Modern isolation and analytical techniques led to identify more than 5000 organohalogens,[2] which number is constantly increasing at a rate of 100-200 new compounds per year. The increasing discovery of new halogen-containing natural products is mainly due to a renewed interest in new sources for potential medicinal drugs.

Halide	Oceans	Sedimentary rocks	Fungi	Wood pulp	Plants
Cl^-	19000	10 – 320	-	70 – 2100	200 – 10000
Br^-	65	1.6 – 3	100	-	-
I^-	0.05	0.3	-	-	-
F^-	1.4	270 – 340	-	-	-

Table 1: Distribution of halides (ppm) in the environment.[3]

Up to the recent past organohalogens were perceived as chemical accidents of nature, while nowadays this general perception is known to be wrong. Most of these molecules possess a huge variety of biological activity (anticancer, antiviral, antibacterial etc.), and halogen atoms usually play a key role in these systems. The large abundance of chloride and bromide ions in the environment explains the predominance of chlorometabolites and bromometabolites, as reported in a recent survey.[4] Fluorinated natural products are the least common, while organoiodines are not so rare (182 in a review of 2014)[5] considering the poor amount of iodine on Earth (a thousand fold less than bromine). Organohalogens origin can be biogenic (synthesized by living organisms) or abiotic (formed during abiogenic processes like forest fires, volcanoes or other geothermal processes).[6] The marine environment is the largest source of biogenic organohalogens, which are plentiful in all known marine organisms. Because of their easy accessibility compared to other sea organisms, sponges have been widely examined and they revealed to be the most prolific species in halo-metabolites synthesis. Other marine species like seaweeds, mollusc, mussels and soft corals produce a huge variety of halo-metabolites. Although most of the known biogenic organohalogens are marine-derived, a wide collection of halogen-containing molecules is biosynthesized by bacteria, fungi, terrestrial plants, insects and higher animals including humans. Most of the known halo-metabolites are peculiar of a specific organism, but it is not unusual to find the same compound in different species. This is the case of the first reported halo-metabolite, 3,5-diiodotyrosine (Fig. 1), isolated in 1896 from the soft coral *Gorgonia cavolinii*[7] and also produced in the thyroid gland of mammals.

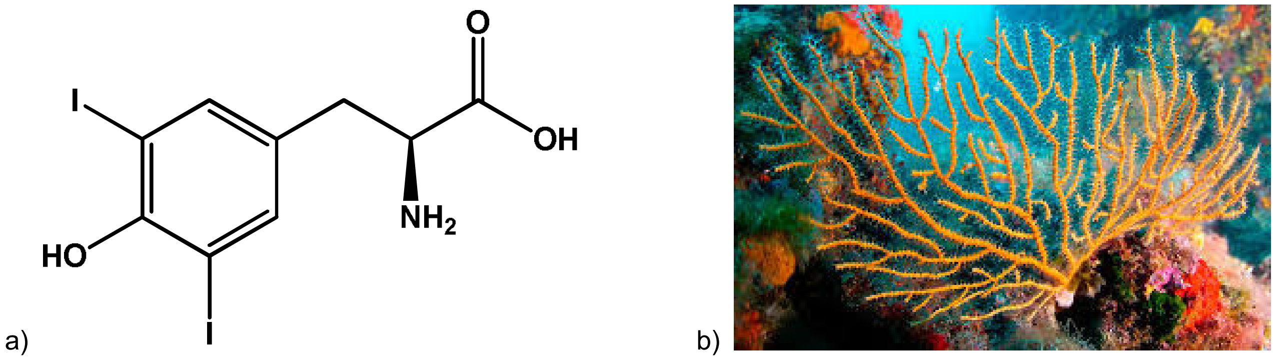


Figure 1: 3,5-diiodotyrosine a) Gorgonia cavolinii b).

1.2 Biohalogenation

Although the existence of halo-metabolites is known since 1896, the first example of enzymatic halogenating activity was reported long after (1959).[8] Since biological halogenation occurs at very different molecular targets (aromatic and heterocyclic rings, aliphatic carbons, olefin centers), the variety of halogenating enzymes is impressive, spreading in a broad range of organisms. Halogenating enzymes can be classified in two macro-categories,[9] according to substrate-specificity:

(I) Highly specific halogenases requiring molecular oxygen for enzymatic activity

(II) Less specific haloperoxidases (HPO) utilizing hydrogen peroxide.

Two different kinds of halogenases have been discovered so far, based on the co-substrate they use to perform their enzymatic activity: $FADH_2$-dependent and Fe^{II}/α-ketoglutarate/O_2-dependent halogenases. Specific $FADH_2$-dependent halogenases exist both for the halogenation of aromatic[10,11] and aliphatic moieties,[12,13] but their mechanism of activity is almost the same (Fig. 2).

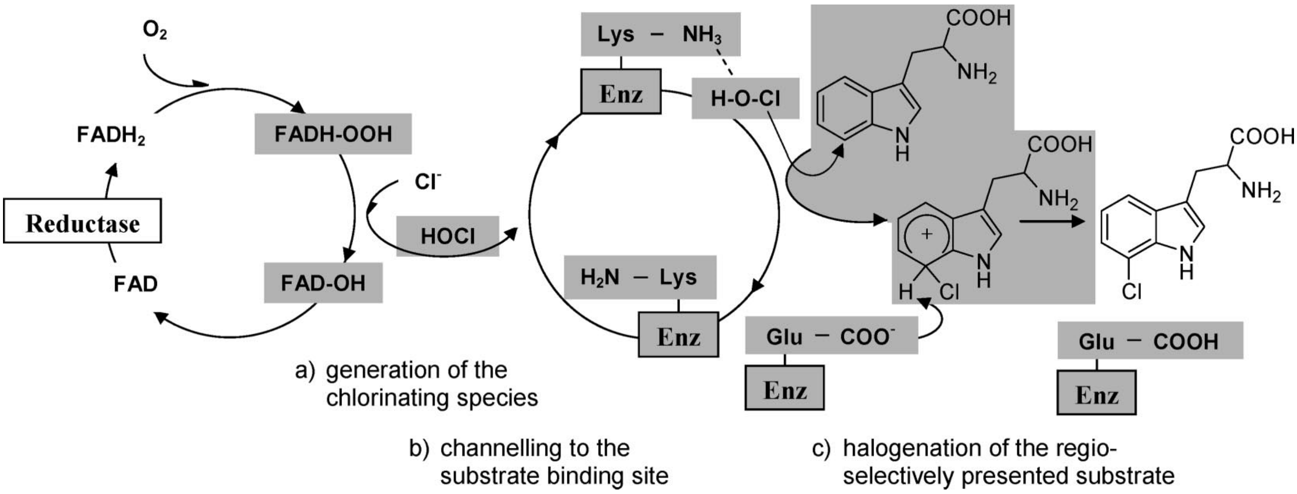


Figure 2: Mechanism of $FADH_2$-dependent halogenation, in this specific example the enzyme is tryptophan-7-halogenase.[9]

$FADH_2$ reacts with O_2 and forms a peroxide-linked co-substrate. A chloride anion, bound at the same region of the co-substrate, becomes HOCl after oxidation. Then the resulting chlorinating agent passes through an intra-enzymatic tunnel to the substrate binding-pocket. HOCl is not allowed to freely diffuse in solution, but remains bound to a Lysine residue – located at the end of the intra-enzymatic tunnel – *via* hydrogen bond. This interaction increases chlorine electrophilicity, allowing it to react with the substrate. A Glutamate residue stabilizes and then deprotonates the substrate intermediate. In general, the strictly controlled release of the halogenating species in the substrate binding-site is a fundamental requirement for regioselective halogenation.

The discovery of Fe^{II}/α-ketoglutarate/O_2-dependent halogenases[14,15] is more recent. The halogenation target of these enzymes is the terminal methyl group of amino acids linked to peptidyl carrier proteins. Fe(II), stabilized in the active site by Histidine residues,[16] is coordinated by water, α-ketoglutarate (α-KG) and a halide ligand (Cl^- or Br^-). Addition of molecular oxygen results in decarboxylation of α-KG and generation of a highly reactive Fe(IV) species. This species reacts with the substrate forming a radical centered on an aliphatic carbon. Then the radical-substrate picks the halide from the coordination sphere of iron, giving rise to the final halogenated molecule (Fig. 3).

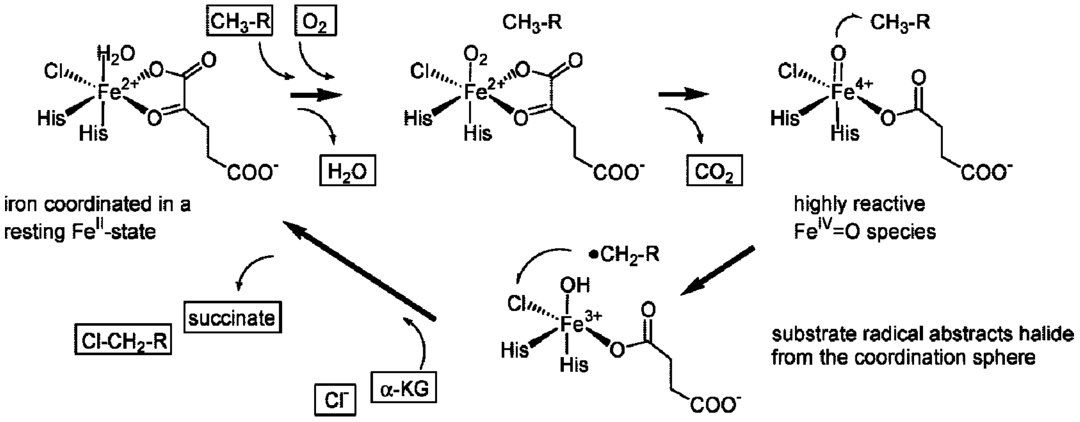


Figure 3: Mechanism of Fe^{II}/α-ketoglutarate/O_2-dependent halogenation.[9] α-KG is α-ketoglutarate, CH_3-R is a generic amino acid with a terminal methyl group.

Haloperoxidases (HPO) are the first class of halogenating enzymes isolated from nature. Working in combination with hydrogen peroxide, these enzymes generate hypohalous acids (HOX) as halogenating species. Once released from the enzyme, HOX acts on organic substrates that are suitable to react with electrophilic species. As already shown for halogenases, almost all known HPOs perform their activity by using a specific co-factor, which is heme or vanadate[17,18] (Fig. 4). Heme-HPOs form hypohalous acids according to a redox process, with the porphyrin iron changing its oxidation state. In Vanadate-HPOs, the cofactor does not change its oxidation state but simply acts as Lewis acid.

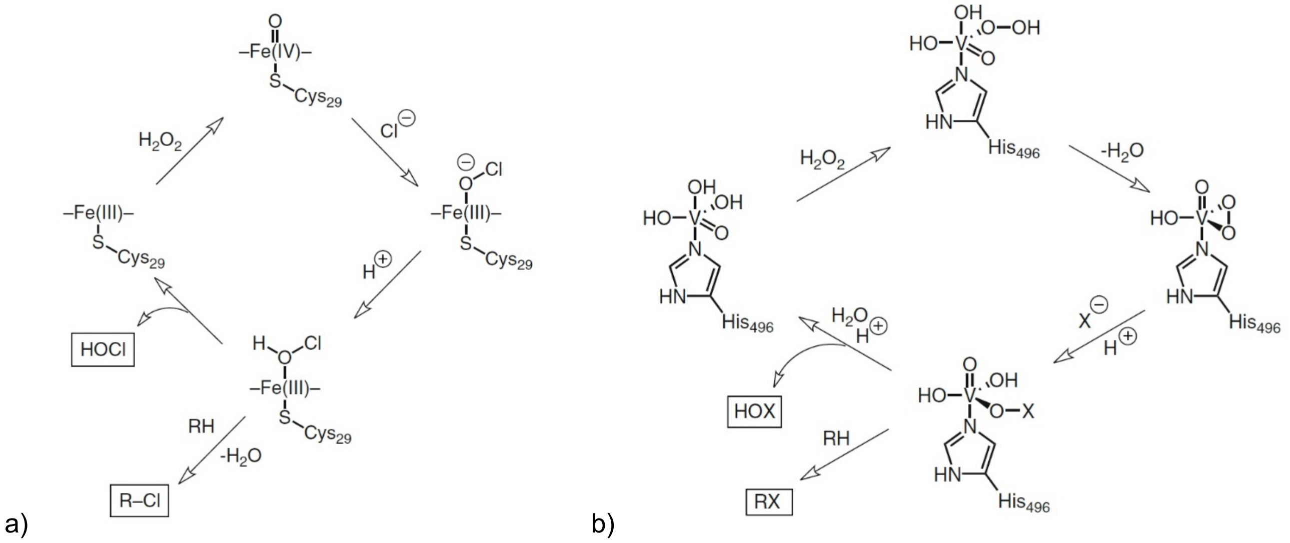


Figure 4: Mechanism of heme-HPO halogenation a) and vanadate-HPO halogenation b).[3]

Haloperoxidases are less halogen-specific than halogenases, as most of them are able to catalyze the oxidation of different halides. Chloroperoxidases can oxidize either chloride, bromide or iodide.[19] Bromoperoxydases can oxidize both bromide and iodide,[20] while iodoperoxidase activity is strictly specific for iodination.[21] Compared to halogenases, HPOs lack in substrate-specificity. A recent study on a *Caldariomyces fumago* Chloroperoxidase[22] confirms that chlorination occurs outside the active site, where the halogenating species freely diffuse reacting with the substrate without any special mode of recognition.

1.2.1 Peroxidases and oxidative stress

Phagocytes are cells that protect a living organism from infections by ingesting invading pathogens. Neutrophils, monocytes, macrophages and eosinophils release different kinds of peroxidases, the role of which is producing reactive oxygen species (ROS) that are lethal for pathogens.[6] However, the occurrence of acute and chronic inflammations creates an imbalance between the production ROS and antioxidant mechanisms, resulting in a potential damage of the inflamed tissues. The exposure of tissues to high concentrations of ROS consists in oxidative stress.[23,24] The first response to an infection is the activation of NADPH oxidase,[25] which reduces molecular oxygen to superoxide ($O_2^{\bullet -}$) (**1**).

$$(\mathbf{1})\ \mathrm{NADPH} + 2O_2 \longrightarrow 2O_2^{\bullet -} + \mathrm{NADP}^+ + H^+$$

The dismutation product of superoxide is hydrogen peroxide (H_2O_2), which is the lethal substance acting on pathogens. The family of mammalian peroxidases includes myeloperoxidase and eosinophil peroxidase. These enzymes use H_2O_2 as substrate to produce hypohalous acids (HOX), which are much more effective in killing pathogens compared to hydrogen peroxide. For instance, HOCl is at least a thousand times more effective than H_2O_2 in killing *Escherichia coli*.[26] Overproduction of hydrogen peroxide during chronic inflammations directly implies an overproduction of hypohalous acids by myeloperoxidase[27,28] and eosinophil peroxidase.[29–31] The excess of HOX and deriving reactive halogenated species in tissues is a kind of oxidative stress named halogenation stress.[32] Reactive halogenated species produced by peroxidases rapidly act on different biochemical targets such as proteins,[33] lipids and nucleotide precursors,[34,35] forming halogen-modified biomolecules. Hypohalous acids can react with primary amines to form halogenated amines, which are even more reactive than HOX. Moreover, hypohalous acids and halide ions can react at acidic pH to form molecular halides or interhalogen compounds[36] (**2**), which are combinations of different halogens (Fig. 5).

$$(\mathbf{2})\ HOX + H^+ + X'^- \leftrightarrow XX' + H_2O$$

Both molecular halides and interhalogens are very strong halogenating agents compared to HOX. In vitro studies[37] demonstrate that HOCl generated by myeloperoxidase is not able to chlorinate a tyrosine substrate in the absence of Cl^-, as a confirmation that chloride is required to allow the formation of Cl_2 (**3**).

$$(\mathbf{3})\ HOCl + H^+ + Cl^- \leftrightarrow Cl_2 + H_2O$$

Further evidence that Cl_2 is the effective chlorinating compound comes from the fact that molecular chlorine is able to halogenate the substrate without the aid of any other agent.

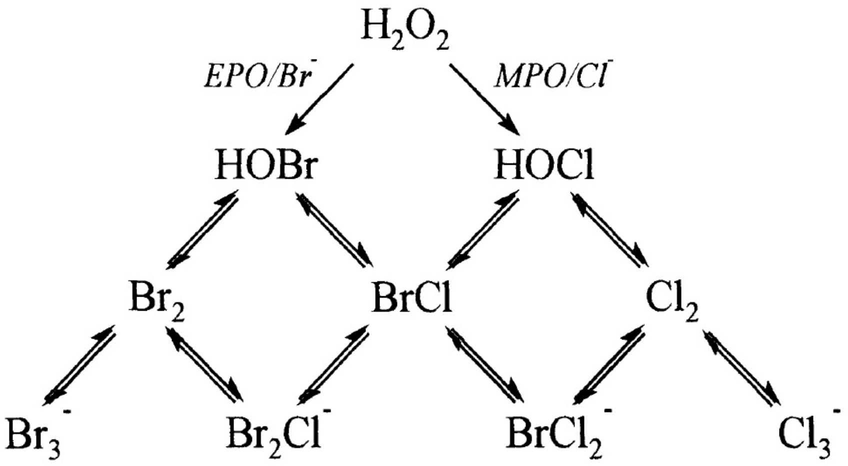


Figure 5: Formation of molecular halides and interhalogen compounds from eosinophil peroxidase and myeloperoxidase.[6]

Halogenated tyrosines, in particular 3-chlorotyrosine and 3-bromotyrosine, are biomarkers of tissues suffering from halogenation stress. *In vitro* studies suggest that only myeloperoxidase and eosinophil peroxidase can halogenate tyrosine substrates,[37] thus their detection in biological material indicates the occurring of halogenation stress. Mouse models of acute infection and inflammation revealed high concentrations of chlorotyrosines, while genetically altered mice lacking in functional myeloperoxidase did not contain any trace.[38,39] This finding confirms that myeloperoxidase is primarily responsible for chlorotyrosine detection *in vivo*. Striking levels of 3-chlorotyrosine have been found in sputum proteins of cystic fibrosis patients[40] and in lung washings of patients suffering from acute respiratory distress syndrome.[41] Significant levels of 3-bromotyrosine

– the main halogenated product resulting from the enzymatic activity of eosinophil peroxidase – have been detected in asthmatic people upon exposure to allergens.[42] Halogenated tyrosines are supposed to be involved also in some neurodegenerative disorders like Parkinson's (PD) and Alzheimer's disease (AD).[43] Pathologies affecting the nervous system are generally associated with dysfunctions of the dopamine system, although its role in AD is still an open debate.[44] Parkinson's disease, which causes are better known, occurs upon a loss of dopaminergic neurons in the *substantia nigra* (Fig. 6), a small area of midbrain having an important role in movement. Since halogenated tyrosines – mainly chlorotyrosines and iodotyrosines – are considered neurotoxic for dopamine-secreting neurons,[45,46] they are deeply studied for understanding PD's triggering factors. Elevated level of halogenated tyrosines have been detected in brain sections of patients suffering from AD and PD, further suggesting a possible pathogenic role of halogenation stress in these disorders.[47,48]

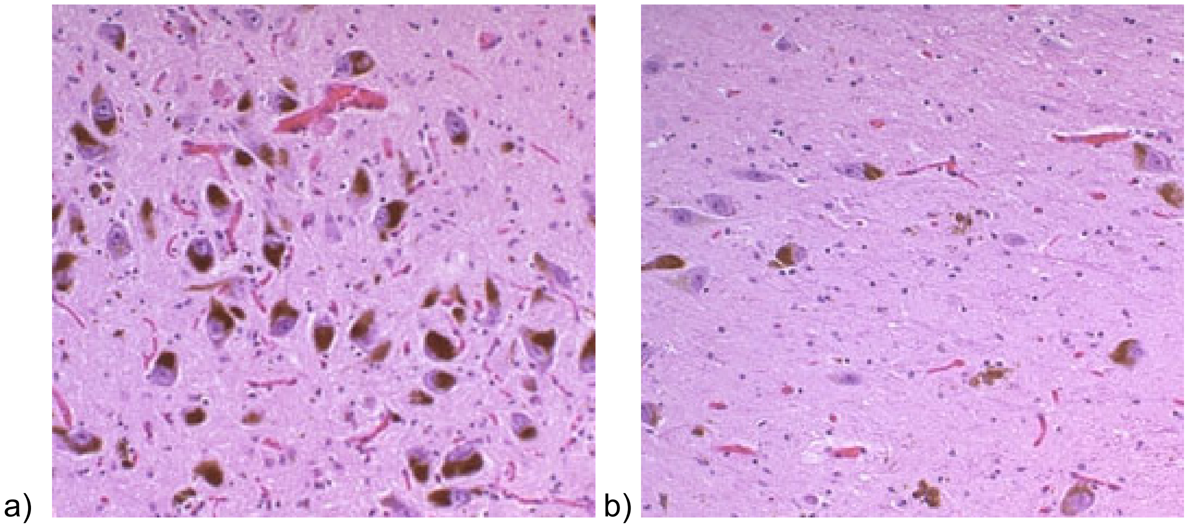


Figure 6: Healthy substantia nigra a) substantia nigra affected by Parkinson's disease b).

Besides amino acids and proteins, also nucleic acids are possible halogenation targets.[49,50] The presence of halogen-modified nucleotides in halogenation stress-exposed tissues can lead to their incorporation into DNA and consequent transition mutations, whose extreme consequence could be cancer development.[51,52]

1.3 Halogen bonding: general features

After a brief description of naturally occurring halogenation processes, a molecular insight about the role of the halogen atoms in halogen-modified compounds is necessary. Halogenation is a minimal structural modification; however, it can induce deep changes in the molecular properties because of the wide range of noncovalent interactions involving halogen atoms.[53] Among these interactions, one of the most recently studied due to its strength, specificity and directionality is halogen bonding[54] (XB). Almost two hundred years after the first example of halogen-bonded complex[55–57] ($H_3N\cdots I_2$), in 2013 IUPAC proposed the following definition[58] of XB: “A halogen bond occurs when there is evidence of a net attractive interaction between an electrophilic region associated with a halogen atom in a molecular entity and a nucleophilic region in another, or the same, molecular entity”. The following scheme (Fig. 7) describes the halogen bond as Y—X•••D: X represents the electrophilic halogen atom (Lewis acid, defined as XB donor) covalently bonded to Y moiety. D is the electron-rich species (Lewis base, defined as XB acceptor) interacting with the halogen atom.

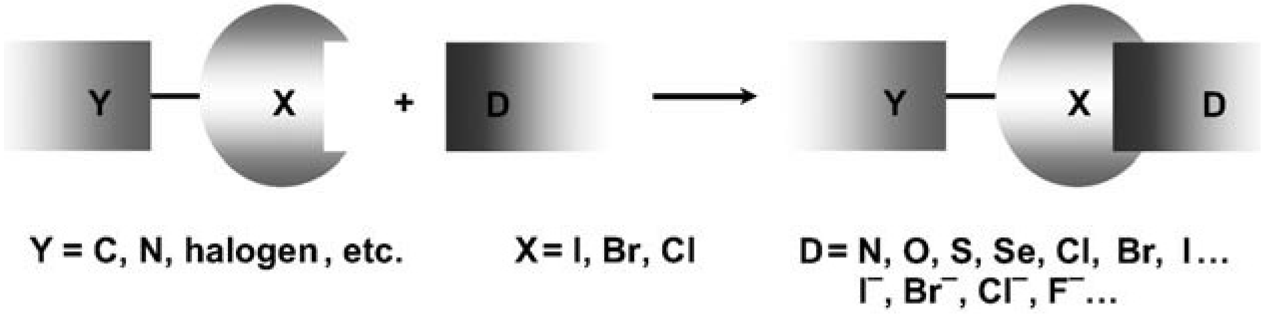


Figure 7: Schematic representation of the halogen bond.[59] X represents the electrophilic halogen atom (defined as XB donor) covalently bonded to Y moiety. D is the electron-rich species (defined as XB acceptor) interacting with the halogen atom.

Although halogens are among the most electronegative elements, they can act as electrophilic species because of the anisotropic distribution of the electron density around their nucleus.[60,61] When halogen atoms form covalent bonds, the new chemical environment affects the halogen electron density,[62,63] which is no more equally distributed in all directions. The electrostatic potential of the halogen atom becomes anisotropic, showing distinct regions around the nucleus with different chemical behavior. The external region located along the extension of the covalent bond results in lower electron density. Differently, the region perpendicular to the covalent bond shows increased electron density. The resulting shape of the halogen atom is an ellipsoid having the shorter radius (named polar flattening[64]) along the covalent bond direction (Fig. 8).

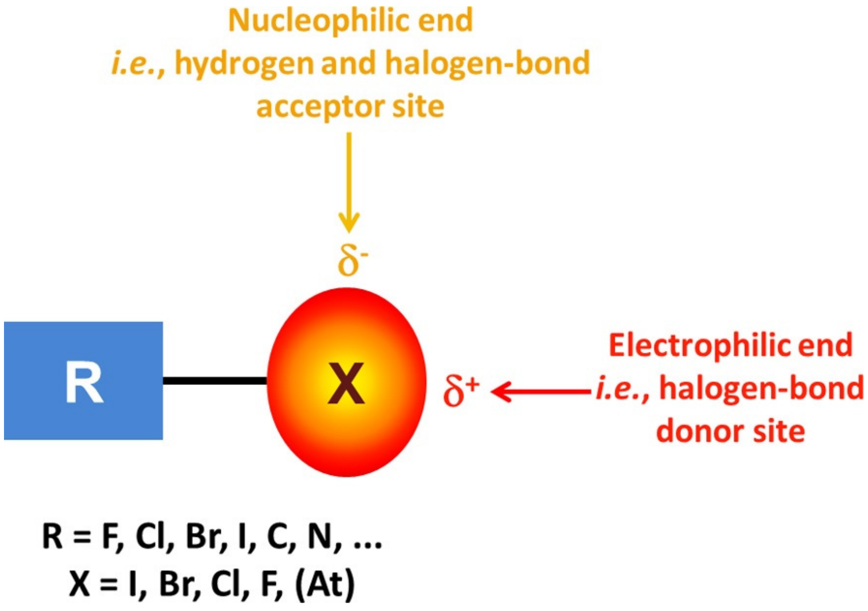


Figure 8: Representation of the anisotropic distribution of the electron density around a covalently bound halogen atom indicating the peculiar interaction pattern of the resulting electronic regions.

Politzer defined the region of positive electrostatic potential along the C−X covalent bond as "σ-hole" because it represents a local deficit of electron charge opposite to σ-bond.[65] A belt having a negative electrostatic potential – orthogonal to the C−X covalent bond – surrounds the σ-hole. According to this electrostatic configuration, nucleophiles form linear interactions (160°-180°) respect to the C−X bond, while electrophiles give rise to lateral contacts (90°-120°) against the bond axis.[66] Computational studies on short contacts of the R−X•••H type show that hydrogen atoms (positively charged) tend to approach the halogen orthogonally to the covalent bond, nicely confirming the existence of the negative belt on the halogen surface. The amphoteric properties of halogens, hence the peculiar localization of the σ-hole – in the same direction of the covalent bond involving the halogen atom – is the key feature conferring striking directionality to XB. The strength of halogen bond, depending on both size and magnitude of the σ-hole,[67] can be tuned upon changing the nature of the halogen atom and its covalently bound residues (Fig. 9). Increasing halogen atom polarizability (decreasing electronegativity) and raising electron-withdrawing propensity of its covalently bound moiety result in a more positive and larger halogen σ-hole,[68–70] which corresponds to a stronger XB. Consequently, halogen bond strength increases following the trend F << Cl < Br < I. In general, strong – hence short –

interactions are more directional than long ones, thus even XB directionality follows the same trend depending on halogen atom polarizability, with iodine the more prone to be involved in halogen bonding. The interaction energies related to halogen bonding spread in a wide interval depending on the interacting partner of the XB donor. The energy values range from 5 kJ/mol for weak interactions (N•••Cl contacts)[71] to 150 kJ/mol (very strong interaction observed in the $I_2 \cdots I^-$ adduct).[72] Halogen bond becomes the favorite noncovalent interaction – prevailing over the more common π–π stacking, dipole−dipole and hydrophobic forces – when the involved energies are significant.

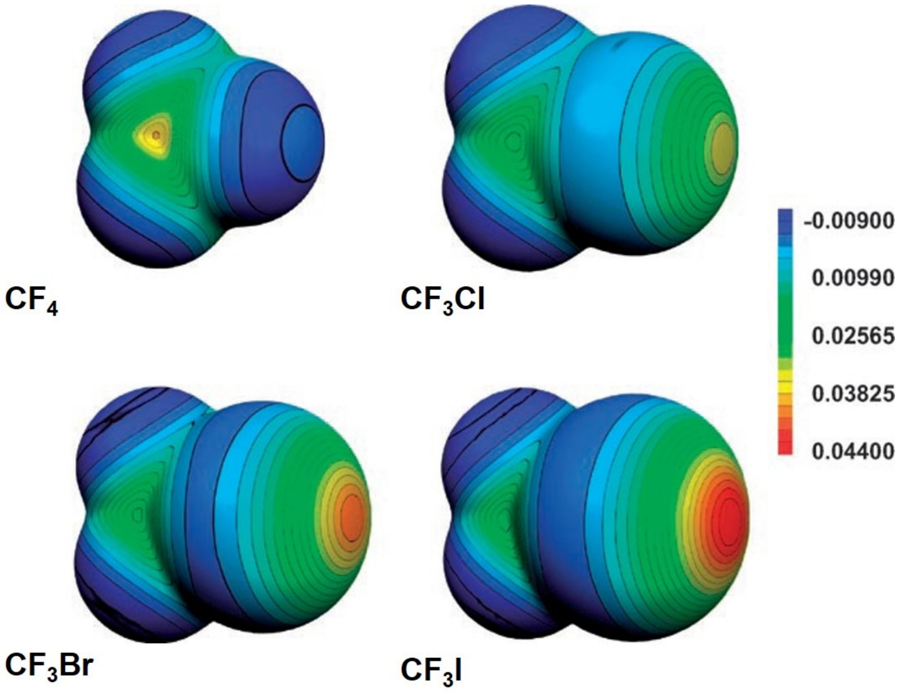


Figure 9: Molecular electrostatic potential (Hartrees) at the isodensity surface with 0.001 au for CF_3X.

1.3.1 Halogen bonding and hydrogen bonding: competition and cooperation

The term halogen bond (XB) immediately recalls its close relative hydrogen bond (HB). The first feature shared by XB and HB is an evident structural similarity.[73] Both hydrogen bonding (R−H•••Y) and halogen bonding (R−X•••Y) involve a covalently bound electropositive atom (Lewis acid) interacting with a negative site (Lewis base). Halogen and hydrogen bonds share comparable strengths. Keeping fixed R and Y moieties, hydrogen bonding is generally stronger when X is fluorine or chlorine,[74,75] while they are comparable when X is bromine.[76] Differently, halogen bonding is often stronger when X is iodine.[77] Both halogen and hydrogen bonding are directional interactions, although this feature is less marked for HB.[78,79] The lower directionality depends on the electronic configuration of the hydrogen atom, having only one valence electron participating in the R−H bond. The absence of any other valence electrons results in a more distributed electronic depletion on the atom's surface. Consequently, even the lateral sides of hydrogen's surface have relatively low electronic densities.[80] It means that the positive σ-hole is more hemispherically shaped and not strictly localized in a specific region like in halogen atoms (Fig 10).

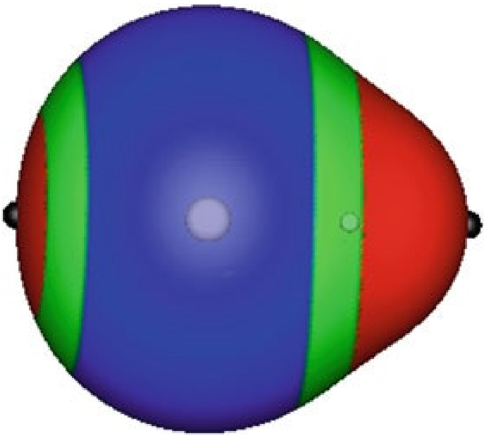


Figure 10: Calculated electrostatic potential on the 0.001 au molecular surface of HI, showing the more focused σ-hole of iodine (left side) and the hemispherical-shaped σ-hole of hydrogen (right side). Color ranges (volts): red, > 0.43 V; green, 0.43 < V < 0; blue, < 0. The most positive values of the electrostatic potential are shown by black hemispheres; hydrogen, 1.26 V; iodine, 0.91 V.

Due to their similarities, halogen bonding and hydrogen bonding can compete or cooperate. Extensive computational and experimental studies investigated the competition between XB and HB. In 2007, Aakeroy studied this competition by a series of co-crystallizations involving a ditopic probe molecule containing two sites – pyridyl and benzimidazole – that can act as either HB or XB acceptors.[81] He found as key factor of this competition is the propensity to form XB, since C−Br – a much weaker Lewis acid than the C−I moiety – does not form a halogen bond that can compete with the C−H•••N hydrogen bond. Recently, competitive studies between XB and HB focused their interest even in the context of biomolecular self-assembly processes. A recent survey on the Protein Data Bank[82] (PDB) demonstrates that XB and HB often coexist. This study shows that halogen bonds and hydrogen bonds are orthogonal molecular interactions,[83] both geometrically perpendicular and energetically independent. Nearly all the protein-ligand complexes belonging to the dataset show XB and HB sharing a common carbonyl oxygen as Lewis base. Generally, the carbonyl oxygen belongs to a peptide bond, while the orthogonal HB comes from other peptide bonds, amino acid side chains or water molecules. This is the best example of cooperation of the two interactions conferring increasing structural stability to biomolecular systems. The X•••O•••H angle formed by halogen and hydrogen bond (Fig. 11) has an average value of 88.28°. Angles close to the theoretical value of 90° are more frequent when iodine and bromine are involved in XB.

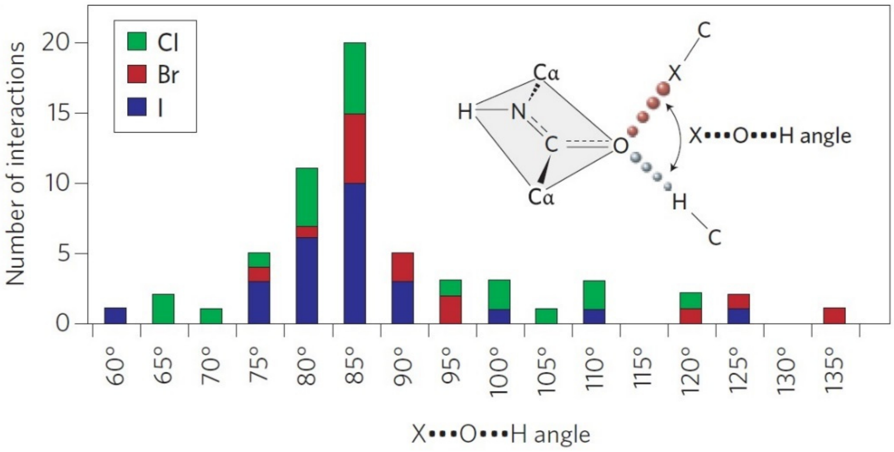


Figure 11: Histogram of the X•••O•••H angle between HBs and XBs sharing a carbonyl oxygen as Lewis base. Different colors highlight each type of XB (involving Cl, Br or I).

As already mentioned in the previous paragraph, halogens can act as HB acceptors (Lewis bases) sharing their electron density with electropositive hydrogen atoms (HB donor). Because of the anisotropic electrostatic potential on the halogen's surface, halogen bonding and hydrogen bonding can coexist on the same moiety. The positive σ-hole acts as halogen bond donor, while the negative belt surrounding the σ-hole and perpendicular to the R−X axis forms hydrogen bonds. The strength of HB involving halogen atoms increases with the electronegativity of the halogen (F > Cl >Br > I). The amphoteric properties of halogen atoms make XB and HB complementary – and not competitive – interactions.

1.3.2 Halogen bonding: a new perspective for biological systems

Because of its peculiar properties of strength and directionality, halogen bonding has become one of the most exploited interactions in supramolecular chemistry and crystal engineering. However, extensive studies focusing on XB relevance in biomolecular systems started quite recently.[84] A wide collection of naturally occurring halogenated products exists and continuously expands every year with new discoveries. Almost all natural organohalogens possess striking biological activity, which is the main reason making these products an inspiration for new drugs. Since it is known that halogen atoms can improve biochemical properties like membrane permeability and half-life, their introduction in potential drug candidates is a common practic in pharmaceutical optimization.[85–87] Although the importance of halogens in biomolecular systems is well established, a molecular interpretation correlating the interaction pattern of halogen atoms with the chemical properties of biomolecules is rare.[88,89] There are different examples of the occurrence of halogen bond in biological systems. Among the naturally occurring halo-metabolites, the iodinated thyroid hormones T_3 (triiodothyronine) and T_4 (thyroxine) work as XB donors – through I•••O short contacts – in the molecular recognition process regulating their activity.[90–92]

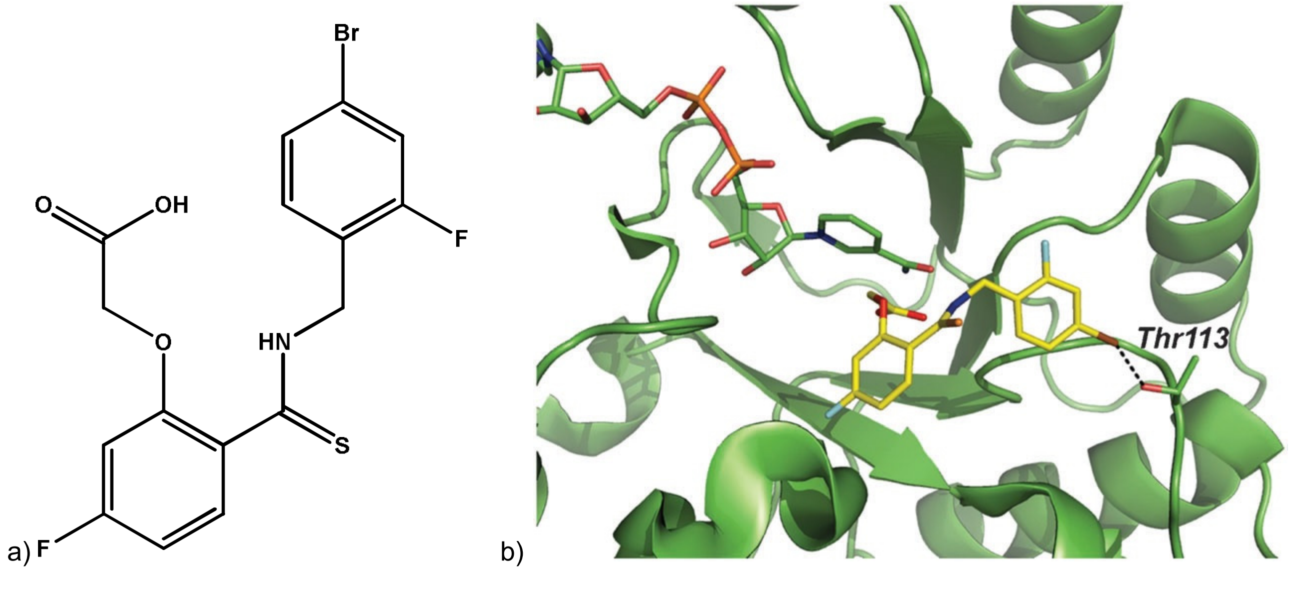


Figure 12: Chemical structure of IDD594 a) Inhibitor-protein complex showing Br•••O contact between IDD594 and Thr113 of human aldose reductase b).

High-resolution X-ray diffraction data (0.66 Å) of a protein-inhibitor complex involving human aldose reductase (AR) provide the first clear evidence of halogen bonding in a large bio-molecular system.[93,94] The crystal structure of AR-inhibitor IDD594 (Fig. 12) shows XB between the bromine atom of the inhibitor and the hydroxyl oxygen of Thr113 (Br•••O distance 2.97 A˚, C−Br•••O angle 152.81°). The brominated phenyl ring of the inhibitor has also a fluorine atom, electron withdrawing ability of which enhances the bromine propensity to act as halogen bond donor. This interaction makes IDD594 highly specific for AR; in fact, the inhibitor does not bind the enzyme aldehyde reductase, which binding pocket contains Tyr (a much bulkier amino acid) instead of Thr113. The presence of this Br•••O contact deals not only with specificity, but also with the potency of the inhibitor. Indeed, replacing bromine with chlorine (weaker halogen bond donor) the IC_{50} value increases from 500 to 1300 nM, suggesting a key role of XB in recognition and binding processes. A recent survey on the Protein Data Bank (PDB) nicely rationalizes XB in biological systems,[95] showing the more frequently occurring halogen bond acceptors.

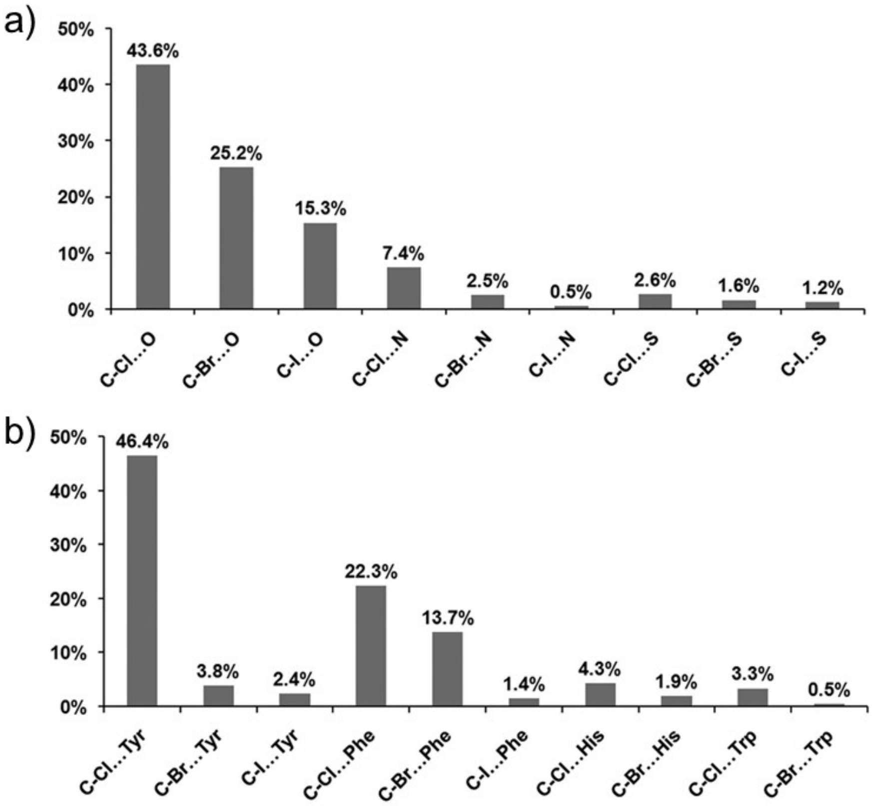


Figure 13: C−X•••Y halogen bonds in Protein Data Bank (PDB) a) C−X•••π halogen bonds involving aromatic amino acid side chains b).

778 short contacts involves halogen atoms: this number includes 567 C−X•••Y contacts and 211 C−X•••π contacts. The latter includes also interactions between the negative belt of the halogen atoms and π systems, thus the number of real C−X•••π halogen bonds is lower. Among C−X•••Y halogen bonds, C−X•••O contacts are prevalent (84.1%), with chlorine being the most frequent halogen bond donor (C−Cl•••O accounts for 43.6%) (Fig. 13a). Halogen bonds involving other XB acceptors are much less common (7.4% C−Cl•••N, 2.5% C−Br•••N, 0.5% C−I•••N, 2.6% C−Cl•••S, 1.6% C−Br•••S, and 1.2% C−I•••S). Considering the 567 C−X•••Y halogen bonds, 430 involve halogen atoms and protein residues. All of the twenty amino acids form C−X•••Y halogen bonds (Fig. 14), with leucine as favorite residue. XBs involving amino acid main chains are the majority (64.6%) respect to the ones formed with side chains (35.4%). According to the nature of the amino acid residue, side chains could prevail over main chains in forming halogen bonds (this is the case of Tyr, Met, Cys, Thr, Asn, Gln, Asp and Arg). Considering C−X··· π halogen bonds involving aromatic amino acids (Fig. 13b), nearly half (46.4%) are C−Cl•••Tyr contacts while 22.3% are C−Cl•••Phe interactions.

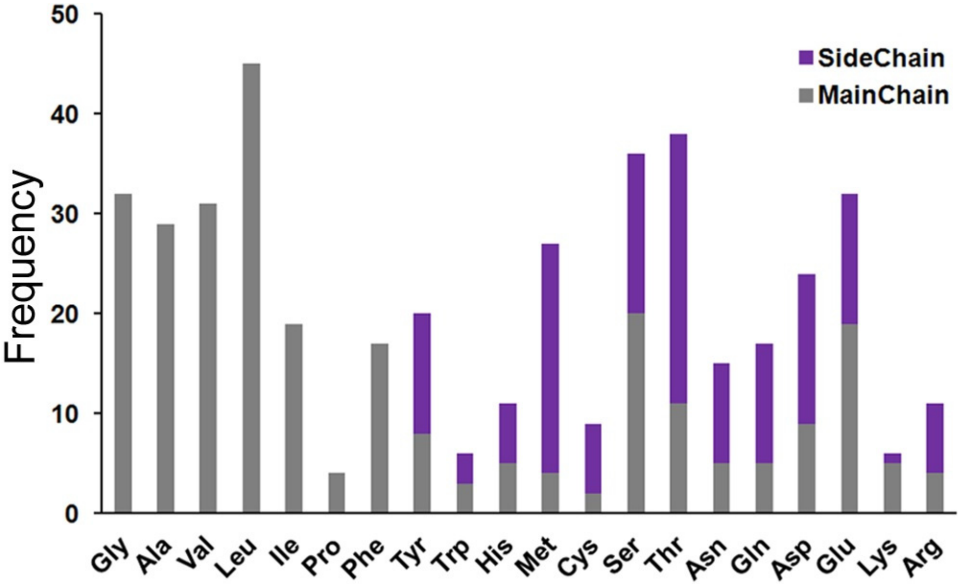


Figure 14: Frequency of C−X•••Y halogen bonds involving the twenty amino acids.

Among the 778 short contacts involving halogens found in the PDB, a significant amount (16.6%) consists in interactions with water molecules. This type of interaction should not be overlooked, especially considering pharmacokinetics. Since the ADME processes (absorption, distribution, metabolism, and excretion) take place in aqueous environment, halogen bonds involving water would play an important role in the performance and activity of a drug. Furthermore, halogen-water contacts are more thermodynamically stable than other water-involved interactions, so their overall contribution cannot be ignored. In crystallography, the difficulty to resolve water molecules at high resolution (< 2.0 Å) suggests that the number of XBs formed with water in the real biological world can be much higher.[95] A typical motif by which water molecules interact with halogen atoms is the halogen–water–hydrogen bridge (XWH-bridge). A water molecule makes a XB with a halogen atom, forming simultaneously a hydrogen bond with an HB donor/acceptor atom.[96] This kind of configuration possesses stronger stability and directionality than the classical water-mediated HB motif, giving a significant contribution in stabilizing biomolecular conformations and mediating protein–protein, protein–nucleic acid, and receptor–ligand recognition and binding.

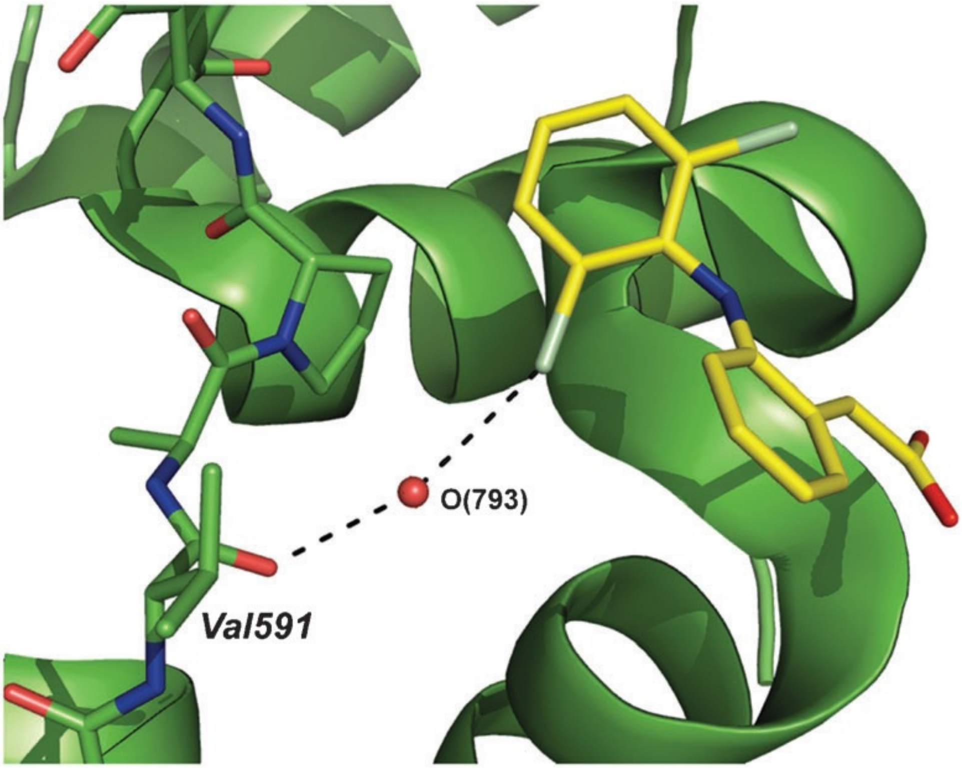


Figure 15: A halogen–water–hydrogen bridge (XWH-bridge) contributes to the interaction between diclofenac and lactoferrin.

1.4 Amyloid fibrils

Once the impact of halogenation stress on biomolecules and the remarkable role of halogen bonding in enhancing molecular stability has been highlighted, it is important to focus on a relevant class of proteins that is a major subject for many disciplines including chemistry, material science and medicine: amyloids. The term 'amyloid' – introduced for the first time by the German scientist Rudolph Virchow[97] – does not properly define a class, but a protein state. The amyloid state of a protein defines the occurrence of thread-like material (amyloid fibrils) that are rich in β-sheet structures. Amyloid fibrils are the result of the self-assembly process involving soluble peptides or proteins that undergo misfolding or denaturation, leading to insoluble polymeric aggregates that are resistant to degradation.[98] Amyloids are highly ordered, close packed and elongated structures; their diameter is in the range of nanometers while their length can reach the order of microns. A variable number of protofilaments (earlier stage of aggregation) twist around each other forming a mature amyloid fibril (Fig. 16). Such a compact architecture, which immediately recalls the structure of a rope, confers impressive chemical stability[99] and strength comparable to steel,[100,101] making amyloids a pioneer subject in material science.

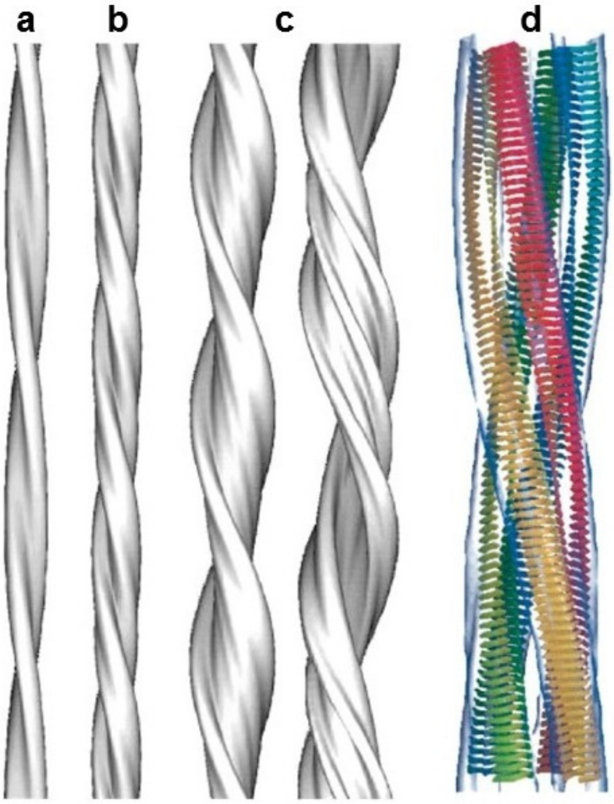


Figure 16: 3D maps of insulin fibril structures obtained using cryoEM.[102] The fibrils can contain two a), four b) or six c) protofilaments. β-strand model for insulin fibrils containing four protofilaments d).

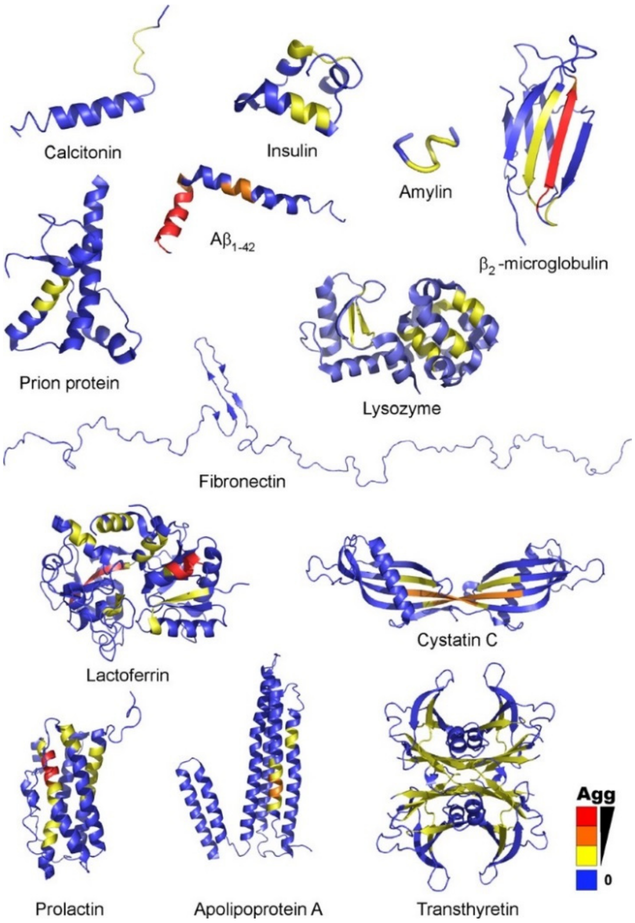


Figure 17: Cartoon representation of proteins involved in amyloid-related diseases. The different colors indicate the protein regions having higher aggregation propensity. Blue means no propensity to aggregate, red means maximum propensity.[103]

Unlike the native state of proteins, where each supramolecular architecture is unique according to the protein function, the amyloid state is almost independent of the amino acidic sequence and it shares the structural features just mentioned above. Recent studies demonstrated that the thermodynamic stability of the fibrillar state is higher than the native one,[104] suggesting that in principle all peptides and proteins can adopt the amyloid structural conformation.[105] However, high activation barriers hinder the functional native state – that is a metastable state – to evolve spontaneously into amyloids. Although all amino acidic sequences can potentially arrange into fibrils, a limited number of peptides and proteins shows a remarkable amyloidogenic behavior[103] (Fig. 17). The amyloid formation from these sequences having high aggregation propensity is often associated with some pathologies.[106] This is the main reason for the growing activity and interest in the amyloid phenomenon, especially considering the social and economic incidence of some amyloid-related diseases. The social burden of Alzheimer’s disease in the world is around $400 billion[107] per year, while type II diabetes is expected to arrest the rising life expectancy in the most developed countries.[108]

1.4.1 Amyloid structure

The definition of amyloid has continuously evolved over the last two centuries according to the growing understanding of its structure. Starting from the introduction of the term (1854), when scientists erroneously considered carbohydrates as the main components of amyloids instead of proteins, nowadays it is possible to have a structural insight up to the atomic level. At macromolecular level, Transmission Electron Microscopy (TEM) and Atomic Force Microscopy (AFM) allow to obtain high-resolution images of long, relatively straight, unbranched filaments (Fig 18). Their width varies in the range 6-15 nm, depending on the number of bundled subunits (protofilaments of 2-3 nm in diameter) constituting the fibril.[102,109,110]

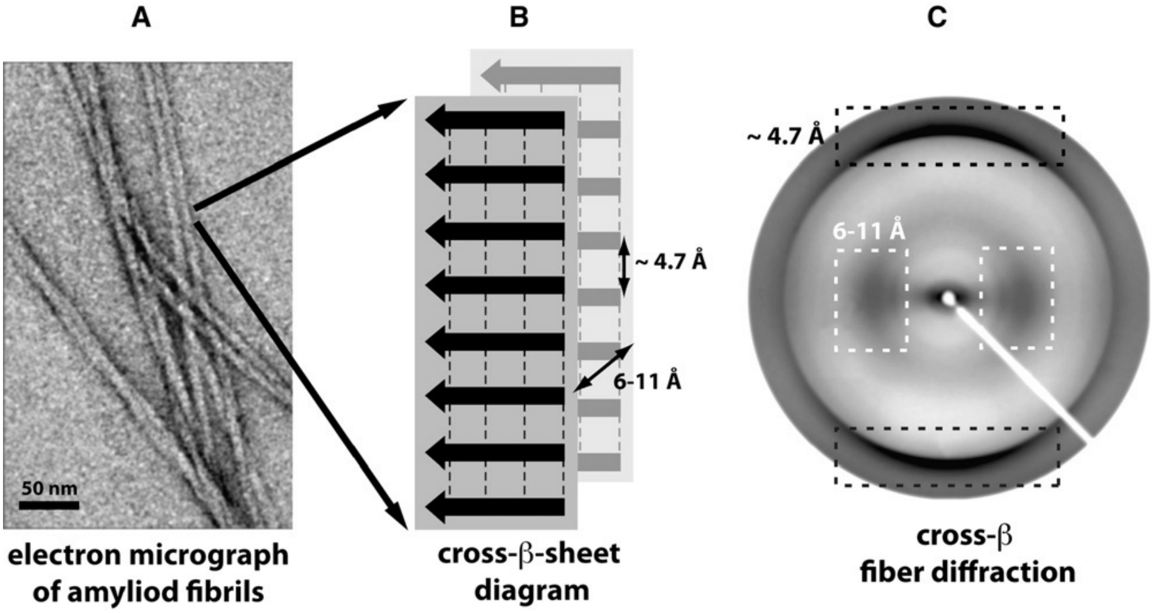


Figure 18: Enlightening amyloid structure from macroscopic to molecular level. Electron micrograph revealing the macroscopic appearance of an amyloid fibril a) diagram showing the peculiar cross-β structure b) typical X-ray fiber diffraction pattern giving the geometrical features of the cross-β structure c).

Several AFM and TEM studies[110,111] revealed that protofilaments – which are clearly distinguishable in micrographs – could arrange in different mode and number under the same experimental conditions. All amyloid fibrils have a pure translational symmetry element lying parallel to the fibril axis; however, a combination between translational and rotational elements is more frequent, resulting in a helical or screw symmetry.[112] These symmetrical features imply that well aligned fibrils could possess a peculiar diffraction pattern, giving additional details about the core structure of amyloids. X-ray fiber diffraction[109,113,114] reveals a meridional reflection at 4.7 Å and an equatorial reflection at 6–11 Å (Fig. 18). The first one is consistent with the spacing among β-strands, which form a β-sheet structure running along the fiber axis. The latter corresponds to the distance between adjacent β-sheets, interacting orthogonally respect to the fiber axis.[115–117] These characteristic distances define the typical cross-β spine structure, which is common to all amyloid fibrils. Although the core structure is the same, with hydrogen bonding playing a key role in stabilizing the β-strands, a degree of polymorphism in the amino acid sequence exists.[118] This polymorphism results in different β-strand orientation, different number of β-sheets in a protofibril and – as already mentioned – variable number of protofibrils forming a mature fibril. The side chain interactions are the driving force for these molecular changes, since varying the amino acid sequence affects how the side chains interact each other and with the surrounding solvent. In order to obtain detailed information about the non-covalent interactions determining amyloid polymorphism, high-resolution structural methods – like single crystal X-ray diffraction – are essential.

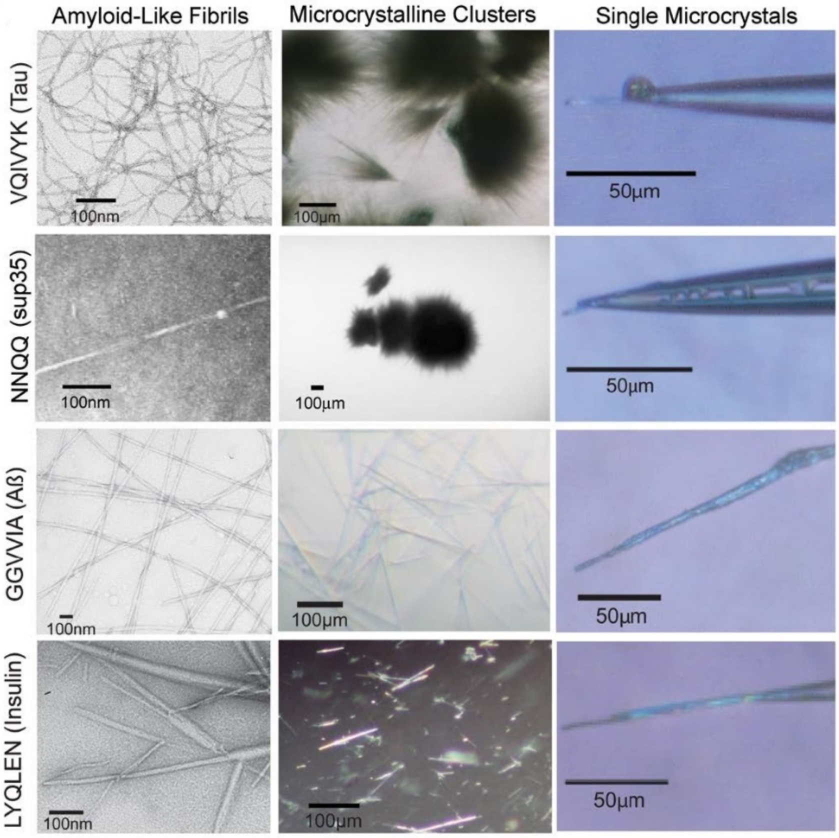


Figure 19: Electron micrographs of representative amyloid-like fibrils (left), magnified images of microcrystalline clusters (middle) and single microcrystals mounted for X-ray diffraction (right).[123]

The main challenge to apply this technique is to obtain crystals of amyloid fibrils. Despite having highly ordered structures, the one-dimensional nature of fibrils makes them poor candidates for crystallization. However, several fragments of amyloidogenic sequences, short enough to pack in a three-dimensional lattice, form very small crystals (Fig. 19). Since these microcrystals are invariably needle-shaped, having a cross section of few nanometers, non-conventional X-ray source as synchrotron radiation are necessary for their structure determination. Several short fragments of amyloid-related proteins were successfully crystallized; structures revealed by these microcrystals nicely reflect the cross-β structure of amyloid fibrils. These short protein segments also form amyloid-like fibers having the same features of amyloid fibers formed by their parent protein; diffraction studies showing that the long axis of the microcrystals correspond to the fibril axis confirmed this similarity. Microcrystals and fibrils often grow in the same conditions, suggesting that their structures are closely related. The capability of microcrystals to seed the growth of amyloid fibrils is further evidence for this similarity. The first amyloid-forming segment determined by single-crystal X-ray diffraction was the sequence GNNQQNY,[119] which is part of the yeast prion protein.[120] Bioinformatics studies identified the amyloidogenic potential of this sequence. After crystal structure determination of this peptide, it was possible to develop a structure-based algorithm[121,122] that allowed identifying new amyloidogenic segments from other amyloid-related proteins. More than 100 crystal structures of amyloid-forming fragments have been solved, representing a widespread survey of the types of interactions that support the cross-β spine motif. This crystallographic collection reveals a tight packing among spinal β-sheets, which occurs among the amino acid side-chains protruding from the backbone of each β-strand. The side-chains of adjacent β-sheets seems to interdigitate like the meshing teeth of a zipper. For this reason, the stacking interface resulting from the pairing of spinal β-sheets is named ‘steric zipper’. The ‘steric zippers’ can be distinguished in ‘homosteric’ (consisting of β-strands of a single sequence, and ‘heterosteric’ (β-strands of different sequences). Sawaya *et al.*[123] classified the known ‘homosteric zippers’ considering three different features: (I) orientation of the β-strands in the same β-sheet (parallel or antiparallel); (II) orientation of the paired β-sheet (parallel or antiparallel) with respect to one another; (III) packing of the paired β-sheet (identical or opposite surfaces of the facing β-sheets creating the zipper interface). Considering these criteria, eight different classes of steric zippers have been defined[124] (Fig. 20), showing that a degree of polymorphism in amyloid structures exists.[125] Besides crystallography, solid-state nuclear magnetic resonance (ssNMR)[126–128] and X-Ray powder diffraction[129,130] are alternative high-resolution structural methods for amyloid fibers.

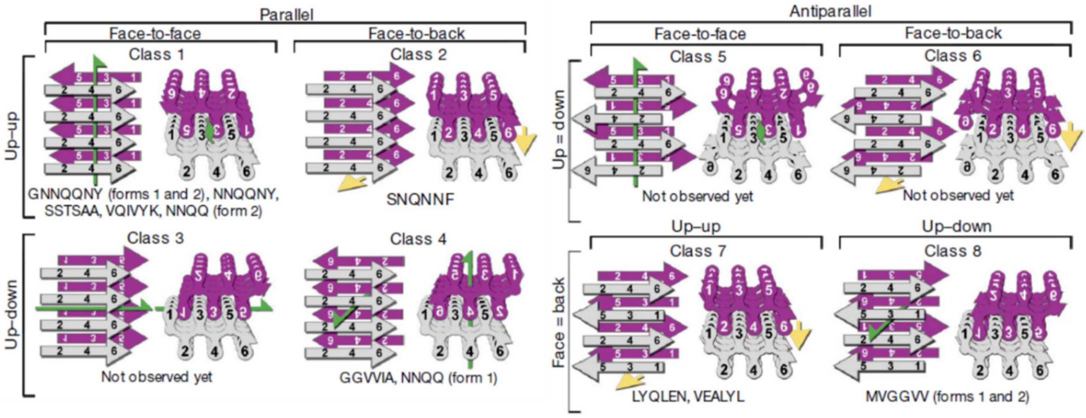


Figure 20: The eight classes of steric zippers. Two spinal β-sheets can be classified by: the orientation of their faces ('face-to-face' or 'face-to back'), the orientation of their strands (with both sheets having the same edge of the strand 'up', or one 'up' and the other 'down'), and the strands orientation within each sheet (parallel or antiparallel). Both side views (left) and top views (right) show which residues of the sequence point into the zipper and which point outward. Green arrows show two-fold screw axes, yellow arrows show translational symmetry. Below each class are listed some examples of protein fragments belonging to that class.[123]

1.4.2 Amyloid formation pathways

The process driving the conversion of proteins from their native states to amyloid fibrils follows a complex mechanism involving many intermediate species and pathways. Native proteins generally undergo a misfolding process, leading to partially or fully unfolded states that start to self-assemble. Such misfolded species are intrinsically amyloidogenic due to the exposure of chemical groups (mainly hydrophobic) that in the native state are confined into restricted regions and contribute to the stability of the folded form through non-covalent interactions. The self-assembly of the misfolded states consists in a structural rearrangement into species stabilized by β-sheet structure. This reorganization, which is the final stage leads to an amyloid fibril, leads to a progressive increase in size, compactness, regularity of the β-sheet structures and rationalization of the hydrophobic groups' location (randomly distributed in the unfolded state).

Due to the complexity and high number of species involved, different approaches can describe the amyloid formation process, each of them emphasizing a specific aspect of the aggregation mechanism. Although many graphical representations there exist, the simplest way to describe the formation of amyloid fibrils is a plot showing the amyloid concentration present in a sample as a function of time (Fig. 21). From the experimental point of view, the amyloid growth can be monitored by using a fluorescent dye like Thioflavin T (ThT), the fluorescence of which increases upon binding to amyloid material.[131–134] Three different stages characterize the aggregation process.[135–137] The first step, named lag-phase or nucleation (grey region), consists in a multiplicity of events representing the embryonic phase of amyloid formation. In particular, the early self-assembly involving the misfolded states results in a complex equilibrium among a series of multi-sized oligomers, the dimensions of which increase over time. At this stage, the amount of amyloid material is under the experimental detection range and the oligomer elongation proceeds until reaching a size limit. Beyond this limit, the rate of the association process prevails on the rate of dissociation,[138] thus the equilibrium moves toward amyloid growth. An aggregated species reaching this size limit is defined as a nucleus. The lag-phase duration depends on different factors. The first one is the nucleus size;[139,140] the smaller is the nucleus size – ranging from monomer to long oligomers – the shorter will be the lag-phase. Another important factor is the impact of fragmentation, a secondary process consisting in the breakage of the forming fibrils. Each breakage event doubles the number of growing fibrils, hence speeds up the overall kinetics, since each fibril fragment acts as seed for further aggregation.[141,142] An atypical case, which is a total bypass of the lag-phase (blue line), occurs in the presence of nuclei from the very beginning of the process.

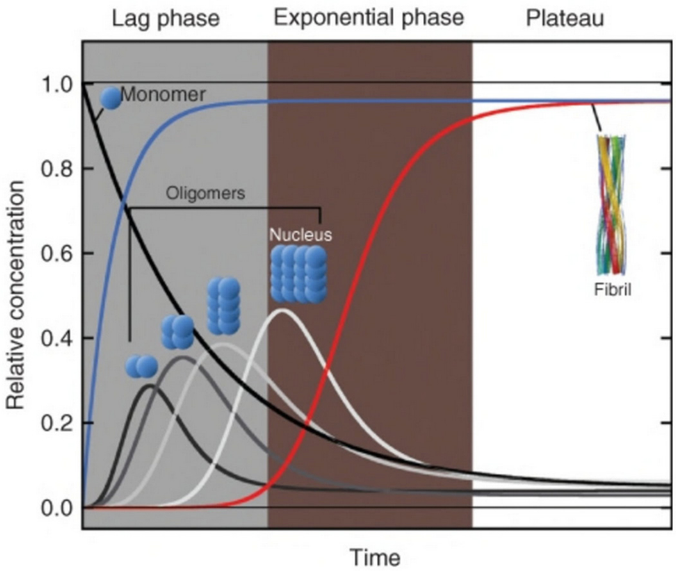


Figure 21: Relative concentration of a protein in different aggregation states as a function of time, starting from the beginning of the aggregation reaction. The black line represents monomer concentration. The red line refers to the concentration of protein in amyloid form. The gray lines refer to oligomers of different size. The kinetics of amyloid formation can be divided in three parts: lag phase (gray region), exponential phase (brown region), and plateau (white region). The lag phase is often suppressed (blue trace) in the presence of seeding.

The end of the lag-phase coincides with the formation of nuclei, which are the starting point of fibril assembly. At this stage, named exponential-growth or elongation (brown region), the nucleating species and other intermediates associate further into filamentous aggregates called protofibrils. Then the protofibrils grow in length and complexity through the binding of monomers and other intermediate structural units, evolving progressively into a mature amyloid fibril (Fig. 22). Once nuclei form, the aggregation proceeds rapidly and results in an evident increase of the fluorescence signal.

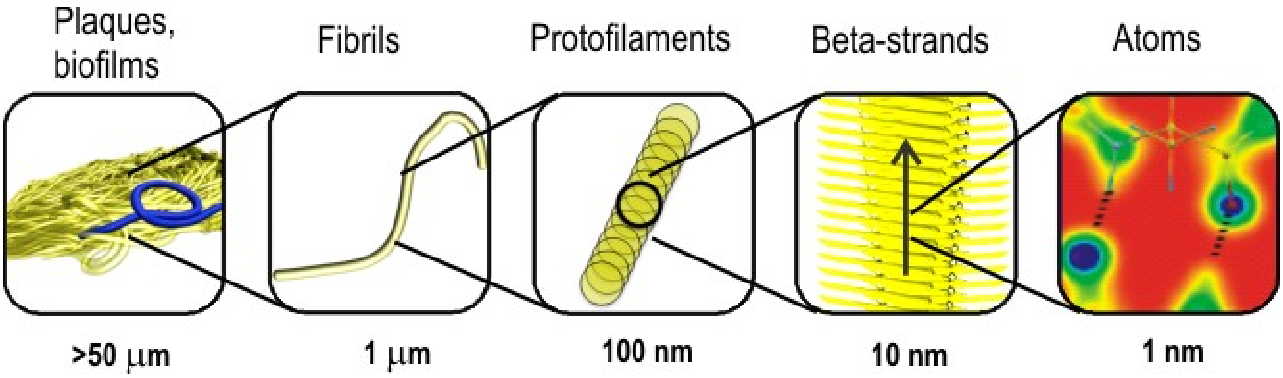


Figure 22: Different levels of hierarchy in the structure of amyloid materials.[143]

At the end of the exponential phase, the signal associated to amyloid concentration does not increase any more. In the last stage, named plateau (white region), the system reaches an equilibrium, with the concentration of fibrils, oligomers[144,145] and monomers[146] remaining constant over time.

Monitoring the amyloid content over time gives several insights about the mechanisms of protein aggregation. The nucleus size, the rates of nucleation, fragmentation, and elongation are standard parameters to assess the amyloidogenic propensity of a protein sequence. Moreover, running the same experiment on different protein variants allows evaluating which amino acid residues mostly affect the aggregation process. In this way, it is possible to map out the most critical protein segments in amyloid formation.[147]

In addition to ThT fluorescence, microscopy[114,148–154] and other spectroscopic techniques such as Dynamic Light Scattering (DLS),[155–159] Small Angle X-ray Scattering (SAXS),[160,161] Circular Dichroism (CD),[162–168] Fourier Transform Infrared Spectroscopy (FTIR),[169] solid state NMR[170–172] (ssNMR) and solution NMR[173] enlighten different aspects of amyloid assembly. However, because of the complexity of the aggregation process, no single technique can fully elucidate all the factors affecting it.

1.5 Objectives

Among the wide range of noncovalent interactions involving halogen atoms, the potential role of halogen bonding in biomolecular self-assembly processes has seldom been explored. The general aim of this thesis is to assess the impact of halogenation on the self-assembly of biomolecules having amyloidogenic propensity.

Since it has been demonstrated both theoretically and experimentally that aromatic residues are of great importance in the aggregation process leading to amyloid formation,[174–178] the aromatic amino acid phenylalanine (Phe) was selected as a scaffold for studying the effect of halogenation on single amino acids and on more complex peptides. In particular, the modified amino acids *p*-iodo-phenylalanine (*p*-I-Phe), *p*-bromo-phenylalanine (*p*-Br-Phe), and *p*-chloro-phenylalanine (*p*-Cl-Phe) were studied in comparison to the wild type (WT) amino acid Phe.

The aim of the first part of this thesis was to demonstrate that halogenation may dramatically influence fibrillation properties of 9-fluorenyl-methoxycarbonyl (Fmoc) functionalized Phe, a well-known Low Molecular-mass Organic Gelator (LMOG).[179–181] The corresponding Fmoc-4-X-Phenylalanines (X = F, Cl, Br, I) were selected to be evaluated against the WT-peptide for their gelation properties. Moreover, halogenation may promote crystallization, therefore providing a tool to correlate halogen bond occurrence with observed rheological properties.

After assessing the role of halogenation on the self-assembly of a simple amino acidic system, the second part of this thesis will focus on the effect of halogenation on the self-assembly of a small amyloidogenic pentapeptide containing a pair of following phenylalanine residues: KLVFF.[182] This sequence was chosen for its biological relevance, since it is the most fibrillogenic region of the peptide Amyloid beta (Aβ),[183,184] commonly associated to Alzheimer's disease.[185] The change of halogen atom nature, number, and position in the sequence was studied in order to correlate the applied modification effect to the peptide nanostructure.

Finally, it was expected that halogenation may promote crystallization and phase determination[187-193] in single crystal X-ray analysis of amyloidogenic peptides, which is usually quite difficult. The obtained halogenated sequences were crystallized with the two-fold purpose to obtain high resolution details of their amyloid structure, as well as to establish the role of halogen bonding in the molecular self-assembly.

1.6 References

(1) Greenwood, N. N.; Earnshaw, A. *Chemistry of the elements*; Butterworth Heinemann, 1997.

(2) Gribble, G. W. *Environ. Chem.* **2015**, *12* (4), 396.

(3) Gribble, G. W. *Naturally occurring Organohalogens Compounds - A Comprehensive Update*; Springer Wien New York, 2010.

(4) Gribble, G. W. *J. Chem. Educ.* **2004**, *81* (10), 1441.

(5) Wang, L.; Zhou, X.; Fredimoses, M.; Liao, S.; Liu, Y. *RSC Adv.* **2014**, *4* (101), 57350.

(6) Gribble, G. W. *Natural Production of Organohalogen Compounds*; Springer, 2003.

(7) Drechsel, E. *Z. Biol.* **1896**, *33*, 85.

(8) Shaw, P. D.; Hager, L. P. *J. Am. Chem. Soc.* **1959**, *88*, 1012.

(9) Wagner, C.; Omari, M. El; Ko, G. M. *J. Nat. Prod.* **2009**, *72* (February), 540.

(10) Yeh, E.; Garneau, S.; Walsh, C. T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (11), 3960.

(11) Seibold, C.; Schnerr, H.; Rumpf, J.; Kunzendorf, A.; Hatscher, C.; Wage, T.; Ernyei, A. J.; Dong, C.; Naismith, J. H.; Van Pée, K.-H. *Biocatal. Biotransformation* **2006**, *24* (6), 401.

(12) Piraee, M.; White, R. L.; Vining, L. C. *Microbiology* **2004**, *150* (1), 85.

(13) Pacholec, M.; Sello, J. K.; Walsh, C. T.; Thomas, M. G. *Org. Biomol. Chem.* **2007**, *5*, 1692.

(14) Krebs, C.; Fujimori, D. G.; Walsh, C. T.; Bollinger, J. M. *Acc. Chem. Res.* **2007**, *40* (Iv), 484.

(15) Vaillancourt, F. H.; Yin, J.; Walsh, C. T. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (29), 10111.

(16) Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L. *Nature* **2006**, *440* (7082), 368.

(17) Hofrichter, M.; Ullrich, R. *Appl. Microbiol. Biotechnol.* **2006**, *71* (3), 276.

(18) Butler, A.; Carter-Franklin, J. N. *Nat. Prod. Rep.* **2004**, *21* (1), 180.

(19) Ohshiro, T.; Littlechild, J.; Garcia-Rodriguez, E.; Isupov, M. N.; Iida, Y.; Kobayashi, T.; Izumi, Y. *Protein Sci.* **2004**, *13* (6), 1566.

(20) Colin, C.; Leblanc, C.; Wagner, E.; Delage, L.; Leize-Wagner, E.; Van Dorsselaer, A.; Kloareg, B.; Potin, P. *J. Biol. Chem.* **2003**, *278* (26), 23545.

(21) Fournier, J. B.; Rebuffet, E.; Delage, L.; Grijol, R.; Meslet-Cladiere, L.; Rzonca, J.; Potin, P.; Michel, G.; Czjzek, M.; Leblanc, C. *Appl. Environ. Microbiol.* **2014**, *80* (24), 7561.

(22) Murali Manoj, K. *Biochim. Biophys. Acta - Proteins Proteomics* **2006**, *1764* (8), 1325.

(23) Barnham, K. J.; Masters, C. L.; Bush, A. I. *Nat. Rev. Drug Discov.* **2004**, *3*, 205.

(24) Thanan, R.; Oikawa, S.; Hiraku, Y.; Ohnishi, S.; Ma, N.; Pinlaor, S.; Yongvanit, P.; Kawanishi, S.; Murata, M. *Int. J. Mol. Sci.* **2015**, *16* (1), 193.

(25) Babior, B. M. *Curr. Opin. Immunol.* **2004**, *16* (1), 42.

(26) Hurst, J. K.; Lymar, S. V. *Acc. Chem. Res.* **1999**, *32* (6), 520.

(27) Harrison, J.E., and Schultz, J. *J. Biol. Chem.* **1976**, *151* (5), 1371.

(28) Weiss, S. J.; Klein, R.; Slivka, A.; Wei, M. *J. Clin. Invest.* **1982**, *70* (3), 598.

(29) Weiss, S. J.; Test, S. T.; Eckmann, C. M.; Roos, D.; Regiani, S. *Science* **1986**, *234*, 200.

(30) Mayeno, A. N.; Curran, A. J.; Roberts, R. L.; Foote, C. S. *J. Biol. Chem.* **1989**, *264* (10), 5660.

(31) Thomas, E. L.; Bozeman, P. M.; Jefferson, M. M.; King, C. C. *J. Biol. Chem.* **1995**, *270* (7), 2906.

(32) Fernández, E. *Austin J. Clin. Neurol.* **2015**, *2* (10), 1082.

(33) Hazen, S. L.; Heinecke, J. W. *J. Clin. Invest.* **1997**, *99* (9), 2075.

(34) Henderson, J. P.; Byun, J.; Heinecke, J. W. *J. Biol. Chem.* **1999**, *274* (47), 33440.

(35) Henderson, J. P.; Byun, J.; Williams, M. V; McCormick, M. L.; Parks, W. C.; Ridnour, L. A.; Heinecke, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (4), 1631.

(36) Kumar, K.; Margerum, D. W. *Inorg. Chem.* **1987**, *26* (16), 2706.

(37) Hazen, S. L.; Hsu, F. F.; Mueller, D. M.; Crowley, J. R.; Heinecke, J. W. *J. Clin. Invest.* **1996**, *98* (6), 1283.

(38) Brennan, M. L.; Anderson, M. M.; Shih, D. M.; Qu, X. D.; Wang, X.; Mehta, A. C.; Lim, L. L.; Shi, W.; Hazen, S. L.; Jacob, J. S.; Crowley, J. R.; Heinecke, J. W.; Lusis, A. J. *J. Clin. Invest.* **2001**, *107* (4), 419.

(39) Gaut, J. P.; Yeh, G. C.; Tran, H. D.; Byun, J.; Henderson, J. P.; Richter, G. M.; Brennan, M. L.; Lusis, A. J.; Belaaouaj, A.; Hotchkiss, R. S.; Heinecke, J. W. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (21), 11961.

(40) Van Der Vliet, A.; Nguyen, M. N.; Shigenaga, M. K.; Eiserich, J. P.; Marelich, G. P.; Cross, C. E. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2000**, *279* (3), L537.

(41) Lamb, N. J.; Gutteridge, J. M. C.; Baker, C.; Evans, T. W.; Quinlan, G. J. *Crit. Care Med* **1999**, *27*, 1738.

(42) Wu, W.; Samoszuk, M. K.; Comhair, S. A. A.; Thomassen, M. J.; Farver, C. F.; Dweik, R. A.; Kavuru, M. S.; Erzurum, S. C.; Hazen, S. L. *J. Clin. Invest.* **2000**, *105* (10), 1455.

(43) van der Veen, B. S.; de Winther, M. P.; Heeringa, P. *Antioxidants Redox Signal.* **2009**, *11* (11), 2899.

(44) Martorana, A.; Koch, G. *Front. Aging Neurosci.* **2014**, *6* (SEP), 1.

(45) Yap, Y. W.; Whiteman, M.; Cheung, N. S. *Cell. Signal.* **2007**, *19* (2), 219.

(46) García-Moreno, J.-M.; Martín de Pablos, A.; García-Sánchez, M.-I.; Méndez-Lucena, C.; Damas-Hermoso, F.; Rus, M.; Chacón, J.; Fernández, E. *Antioxid. Redox Signal.* **2013**, *18* (11), 1296.

(47) Green, P. S.; Mendez, A. J.; Jacob, J. S.; Crowley, J. R.; Growdon, W.; Hyman, B. T.; Heinecke, J. W. *J. Neurochem.* **2004**, *90* (3), 724.

(48) Choi, D.-K. *J. Neurosci.* **2005**, *25* (28), 6594.

(49) Henderson, J. P.; Byun, J.; Williams, M. V.; Mueller, D. M.; McCormick, M. L.; Heinecke, J. W. *J. Biol. Chem.* **2001**, *276* (11), 7867.

(50) Wu, W.; Chen, Y.; Avignon, A.; Hazen, S. L. *Biochem.* **1999**, 3538.

(51) Cascorbi, I.; Henning, S.; Brockmoller, J.; Gephart, J.; Meisel, C.; Muller, J. M.; Loddenkemper, R.; Roots, I. *Cancer Res.* **2000**, *60* (3), 644.

(52) Kantarci, O. H.; Lesnick, T. G.; Yang, P.; Meyer, R. L.; Hebrink, D. D.; McMurray, C. T.; Weinshenker, B. G. *Mayo Clin.Proc.* **2002**, *77* (1), 17.

(53) Metrangolo, P.; Pilati, T.; Resnati, G. *CrystEngComm* **2006**, *8*, 946.

(54) Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimägi, A.; Resnati, G.; Terraneo, G. *Chem. Rev.* **2016**, *116* (4), 2478.

(55) Colin, M. M.; Gaultier de Claubry, H. *Ann. Chim.* **1814**, *90*, 87.

(56) Colin, M. M. *Ann. Chim.* **1814**, *91*, 252.

(57) Guthrie, F. *J. Chem. Soc.* **1863**, *16*, 239.

(58) Desiraju, G. R.; Ho, P. S.; Kloo, L.; Legon, A. C.; Marquardt, R.; Metrangolo, P.; Politzer, P.; Resnati, G.; Rissanen, K. *Pure Appl. Chem.* **2013**, *85* (8), 1711.

(59) Metrangolo, P.; Meyer, F.; Pilati, T.; Resnati, G.; Terraneo, G. *Angew. Chem. Int. Ed.* **2008**, *47* (33), 6114.

(60) Murray, J. S.; Lane, P.; Politzer, P. *J. Mol. Model.* **2009**, *15* (6), 723.

(61) Politzer, P.; Murray, J. S.; Clark, T. *Phys. Chem. Chem. Phys.* **2013**, *15* (27), 11178.

(62) Brinck, T.; Murray, J. S.; Politzer, P. *Int. J. Quantum Chem.* **1992**, *44* (S19), 57.

(63) Brinck, T.; Murray, J. S.; Politzer, P. *Int. J. Quantum Chem.* **1993**, *48* (2), 73.

(64) Sedlak, R.; Kolár, M. H.; Hobza, P. *J. Chem. Theory Comput.* **2015**, *11* (10), 4727.

(65) Clark, T.; Hennemann, M.; Murray, J. S.; Politzer, P. *J. Mol. Model.* **2007**, *13* (2), 291.

(66) Politzer, P.; Murray, J. S.; Clark, T. *Phys. Chem. Chem. Phys.* **2010**, *12* (28), 7748.

(67) Kolář, M.; Hostaš, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9987.

(68) Murray, J. S.; Macaveiu, L.; Politzer, P. *J. Comput. Sci.* **2014**, *5* (4), 590.

(69) Politzer, P.; Murray, J. S. *ChemPhysChem* **2013**, *14* (2), 278.

(70) Riley, K. E.; Murray, J. S.; Fanfrlík, J.; Řezáč, J.; Solá, R. J.; Concha, M. C.; Ramos, F. M.; Politzer, P. *J. Mol. Model.* **2013**, *19* (11), 4651.

(71) Xu, K.; Ho, D. M.; Pascal, R. A. *J. Am. Chem. Soc.* **1994**, *116* (1 3), 105.

(72) Muller, M.; Albrecht, M.; Gossen, V.; Peters, T.; Hoffmann, A.; Raabe, G.; Valkonen, A.; Rissanen, K. *Chem. Eur. J.* **2010**, *16* (41), 12446.

(73) Bent, H. A. *Chem. Rev.* **1968**, *68*, 587.

(74) Alkorta, I.; Blanco, F.; Solimannejad, M.; Elguero, J. *J. Phys. Chem. A* **2008**, *112* (43), 10856.

(75) Ding, H.; Lu, Y.; Wu, W.; Liu, H. *Chem. Phys.* **2014**, *441*, 30.

(76) Politzer, P.; Murray, J. S.; Lane, P. *Int. J. Quantum Chem.* **2007**, *107*, 3046.

(77) Corradi, E.; Meille, S. V.; Messina, M. T.; Metrangolo, P.; Resnati, G. *Angew. Chem. Int. Ed.* **2000**, *39* (10), 1782.

(78) Shields, Z. P.; Murray, J. S.; Politzer, P. *Int. J. Quantum Chem.* **2010**, *110*, 2823.

(79) Legon, A. *Angew. Chem. Int. Ed. Engl.* **1999**, *38* (18), 2686.

(80) Metrangolo, P.; Resnati, G. *Halogen Bonding I*; Springer, 2015.

(81) Aakeroy, C. B.; Fasulo, M.; Schultheiss, N.; Desper, J.; Moore, C. *J. Am. Chem. Soc.* **2007**, *129* (45), 13772.

(82) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acid Res.* **2000**, *28* (1), 235.

(83) Voth, A. R.; Khuu, P.; Oishi, K.; Ho, P. S. *Nat. Chem.* **2009**, *1* (1), 74.

(84) Auffinger, P.; Hays, F. a; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (48), 16789.

(85) Parisini, E.; Metrangolo, P.; Pilati, T.; Resnati, G.; Terraneo, G. *Chem. Soc. Rev.* **2011**, *40* (5), 2267.

(86) Gentry, C. L.; Egleton, R. D.; Gillespie, T.; Abbruscato, T. J.; Bechowski, H. B.; Hruby, V. J.; Davis, T. P. *Peptides* **1999**, *20* (10), 1229.

(87) Wilcken, R.; Zimmermann, M. O.; Bauer, M. R.; Rutherford, T. J.; Fersht, A. R.; Joerger, A. C.; Boeckler, F. M. *ACS Chem. Biol.* **2015**, *10* (12), 2725.

(88) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. *J. Med. Chem.* **2009**, *52*, 2854.

(89) Lu, Y.; Wang, Y.; Zhu, W. *Phys. Chem. Chem. Phys.* **2010**, *12* (18), 4543.

(90) Cody, V. *J. Mol. Struct.* **1984**, *112*, 189.

(91) Eneqvist, T.; Lundberg, E.; Karlsson, A.; Huang, S.; Santos, C. R. A.; Power, D. M.; Sauer-Eriksson, A. E. *J. Biol. Chem.* **2004**, *279* (25), 26411.

(92) Wilcken, R.; Zimmermann, M. O.; Lange, A.; Joerger, A. C.; Boeckler, F. M. *J. Med. Chem.* **2013**, *56*, 1363.

(93) Howard, E. I.; Sanishvili, R.; Cachau, R. E.; Mitschler, A.; Chevrier, B.; Barth, P.; Lamour, V.; Van Zandt, M.; Sibley, E.; Bon, C.; Moras, D.; Schneider, T. R.; Joachimiak, A.; Podjarny, A. *Proteins Struct. Funct. Genet.* **2004**, *55* (4), 792.

(94) Muzet, N.; Guillot, B.; Jelsch, C.; Howard, E.; Lecomte, C. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (15), 8742.

(95) Xu, Z.; Yang, Z.; Liu, Y.; Lu, Y.; Chen, K.; Zhu, W. *J. Chem. Inf. Mod.* **2014**, *54*, 69.

(96) Zhou, P.; Lv, J.; Zou, J.; Tian, F.; Shang, Z. *J. Struct. Biol.* **2010**, *169* (2), 172.

(97) Sipe, J. D.; Cohen, A. S. *J. Struct. Biol.* **2000**, *130* (2–3), 88.

(98) Rambaran, R. N.; Serpell, L. C. *Prion* **2008**, *2* (3), 112.

(99) Schleeger, M.; vandenAkker, C. C.; Deckert-gaudig, T.; Deckert, V.; Velikov, K. P.; Koenderink, G.; Bonn, M.; Corianne, C.; Deckert-gaudig, T.; Deckert, V.; Velikov, K. P.; Koenderink, G.; Bonn, M. *Polymer* **2013**, *54* (10), 2473.

(100) Smith, J. F.; Knowles, T. P. J.; Dobson, C. M.; Macphee, C. E.; Welland, M. E. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (43), 15806.

(101) Knowles, T. P.; Fitzpatrick, A. W.; Meehan, S.; Mott, H. R.; Vendruscolo, M.; Dobson, C. M.; Welland, M. E. *Science* **2007**, *318* (5858), 1900.

(102) Jiménez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V; Dobson, C. M.; Saibil, H. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (14), 9196.

(103) Jahn, T. R.; Radford, S. E. *Arch. Biochem. Biophys.* **2008**, *469* (1), 100.

(104) Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. *J. Am. Chem. Soc.* **2011**, 14160.

(105) a) Hamley, I. W. *Angew. Chem. Int. Ed.* **2007**, *46*, 8128. b) Dobson, C. M. *Nature* **2003**, *426*, 884.

(106) a) Hamley, I. W. *Chem.Rev.* **2012**, *112*, 5147. b) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75* (1), 333.

(107) Otzen, D. E. *Amyloid Fibrils and Prefibrillar aggregates_ Molecular and Biological Properties*; Wiley-VCH, 2013.

(108) Layden, J.; Carnes, B. A.; Ph, D.; Brody, J.; Hayflick, L.; Butler, R. N.; Allison, D. B.; Ludwig, D. S. *N. Engl. J. Med.* **2005**, 1138.

(109) Sunde, M.; Blake, C. *Adv. Protein Chem.* **1997**, *50*, 123.

(110) Khurana, R.; Ionescu-Zanetti, C.; Pope, M.; Li, J.; Nielson, L.; Ramirez-Alvarado, M.; Regan, L.; Fink, A. L.; Carter, S. A. *Biophys. J.* **2003**, *85* (2), 1135.

(111) Makin, O. S.; Serpell, L. C. *J. Mol. Biol.* **2004**, *335* (5), 1279.

(112) Greenwald, J.; Riek, R. *Structure* **2010**, *18* (10), 1244.

(113) Sunde, M.; Serpell, L. C.; Bartlam, M.; Fraser, P. E.; Pepys, M. B.; Blake, C. C. . *J. Mol. Biol.* **1997**, *273* (3), 729.

(114) Jimenez, J. L.; Guijarro, J. I.; Orlova, E.; Saibil, H. R. *EMBO J.* **1999**, *18* (4), 815.

(115) Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. *Proc Natl Acad Sci U S A* **2005**, *102* (2), 315.

(116) Makin, O. S.; Serpell, L. C. *FEBS J.* **2005**, *272* (23), 5950.

(117) Serpell, L. C. *Biochim. Biophys. Acta* **2000**, *1502*, 16.

(118) Fändrich, M.; Dobson, C. M. *EMBO J.* **2002**, *21* (21), 5682.

(119) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nat. Cell Biol.* **2005**, *435* (7043), 773.

(120) Balbirnie, M.; Grothe, R.; Eisenberg, D. S. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (5), 2375.

(121) Thompson, M. J.; Sievers, S. a; Karanicolas, J.; Ivanova, M. I.; Baker, D.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (11), 4074.

(122) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (8), 3487.

(123) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg, D. *Nature* **2007**, *447* (7143), 453.

(124) Stroud, J. C. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (4), 540.

(125) Colletier, J.-P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A. B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16938.

(126) Ritter, C.; Adrian, M.; Riek-loher, D.; Bohrmann, B.; Do, H.; Schubert, D.; Riek, R. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (48), 17342.

(127) Colvin, M. T.; Silvers, R.; Ni, Q. Z.; Can, T. V.; Sergeyev, I.; Rosay, M.; Donovan, K. J.; Michael, B.; Wall, J.; Linse, S.; Griffin, R. G. *J. Am. Chem. Soc.* **2016**, *138* (30), 9663.

(128) 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. *Nat. Struct. Mol. Biol.* **2016**, *23* (February), 1.

(129) Stroud, J. C.; Liu, C.; Teng, P. K.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (20), 7717.

(130) Diaz-Avalos, R.; Long, C.; Fontano, E.; Balbirnie, M.; Grothe, R.; Eisenberg, D.; Caspar, D. L. D. *J. Mol. Biol.* **2003**, *330* (5), 1165.

(131) Krebs, M. R. H.; Bromley, E. H. C.; Donald, A. M. *J. Struct. Biol.* **2005**, *149* (1), 30.

(132) Khurana, R.; Coleman, C.; Ionescu-Zanetti, C.; Carter, S. A.; Krishna, V.; Grover, R. K.; Roy, R.; Singh, S. *J. Struct. Biol.* **2005**, *151* (3), 229.

(133) Biancalana, M.; Koide, S. *Biochim. Biophys. Acta - Proteins Proteomics* **2010**, *1804* (7), 1405.

(134) Biancalana, M.; Makabe, K.; Koide, A.; Koide, S. *J. Mol. Biol.* **2009**, *385* (4), 1052.

(135) Morris, A. M.; Watzky, M. A.; Finke, R. G. *Biochim. Biophys. Acta* **2009**, *1794* (3), 375.

(136) Eichner, T.; Radford, S. E. *Mol. Cell* **2011**, *43* (1), 8.

(137) Lomakin, A.; Chung, D. S.; Benedek, G. B.; Kirschner, D. A.; Teplow, D. B. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (3), 1125.

(138) Ferrone, F. *Methods Enzymol.* **1999**, *309* (1), 256.

(139) Chen, S.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (18), 11884.

(140) Collins, S. R.; Douglass, A.; Vale, R. D.; Weissman, J. S. *PLoS Biol.* **2004**, *2* (10).

(141) 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. *Science* **2009**, *326* (5959), 1533.

(142) Xue, W.-F.; Homans, S. W.; Radford, S. E. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (26), 8926.

(143) Knowles, T. P. J.; Buehler, M. J. *Nat. Nanotechnol.* **2011**, *6* (8), 469.

(144) Koffie, R. M.; Meyer-Luehmann, M.; Hashimoto, T.; Adams, K. W.; Mielke, M. L.; Garcia-Alloza, M.; Micheva, K. D.; Smith, S. J.; Kim, M. L.; Lee, V. M.; Hyman, B. T.; Spires-Jones, T. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (10), 4012.

(145) Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O. *Nat. Struct. Mol. Biol.* **2010**, *17* (5), 561.

(146) Carulla, N.; Caddy, G. L.; Hall, D. R.; Zurdo, J.; Gairí, M.; Feliz, M.; Giralt, E.; Robinson, C. V; Dobson, C. M. *Nature* **2005**, *436* (7050), 554.

(147) Chiti, F.; Taddei, N.; Baroni, F.; Capanni, C.; Stefani, M.; Ramponi, G.; Dobson, C. M. *Nat. Struct. Biol.* **2002**, *9* (2), 137.

(148) Schmidt, A.; Annamalai, K.; Schmidt, M.; Grigorieff, N.; Fändrich, M. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (22), 201522282.

(149) Khurana, R.; Ionescu-Zanetti, C.; Pope, M.; Li, J.; Nielson, L.; Ramírez-Alvarado, M.; Regan, L.; Fink, A. L.; Carter, S. a. *Biophys. J.* **2003**, *85* (2), 1135.

(150) Gosal, W. S.; Morten, I. J.; Hewitt, E. W.; Smith, D. A.; Thomson, N. H.; Radford, S. E. *J. Mol. Biol.* **2005**, *351* (4), 850.

(151) Podestà, A.; Tiana, G.; Milani, P.; Manno, M. *Biophys. J.* **2006**, *90* (2), 589.

(152) Jansen, R.; Dzwolak, W.; Winter, R. *Biophys. J.* **2005**, *88* (2), 1344.

(153) Kad, N. M.; Myers, S. L.; Smith, D. P.; Smith, D. A.; Radford, S. E.; Thomson, N. H. *J. Mol. Biol.* **2003**, *330* (3), 785.

(154) Yagi, H.; Ban, T.; Morigaki, K.; Naiki, H.; Goto, Y. *Biochemistry* **2007**, *46* (51), 15009.

(155) Smith, M. I.; Sharp, J. S.; Roberts, C. J. *Biophys. J.* **2008**, *95* (7), 3400.

(156) Streets, A. M.; Sourigues, Y.; Kopito, R. R.; Melki, R.; Quake, S. R. *PLoS One* **2013**, *8* (1), 1.

(157) Georgalis, Y.; Starikov, E. B.; Hollenbach, B.; Lurz, R.; Scherzinger, E.; Saenger, W.; Lehrach, H.; Wanker, E. E. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (11), 6118.

(158) Shen, C. L.; Scott, G. L.; Merchant, F.; Murphy, R. M. *Biophys J* **1993**, *65* (6), 2383.

(159) Manno, M.; Craparo, E. F.; Martorana, V.; Bulone, D.; San Biagio, P. L. *Biophys. J.* **2006**, *90* (12), 4585.

(160) Ahmad, A.; Millett, I. *Biochemistry* **2003**, *42*, 11404.

(161) Oliveira, C. L. P.; Behrens, M. A.; Pedersen, J. S.; Erlacher, K.; Otzen, D.; Pedersen, J. S. *J. Mol. Biol.* **2009**, *387* (1), 147.

(162) Ahmad, A.; Uversky, V. N.; Hong, D.; Fink, A. L. *J. Biol. Chem.* **2005**, *280* (52), 42669.

(163) Walsh, D. M.; Lomakin, A.; Benedek, G. B.; Condron, M. M.; Teplow, D. B. *J. Biol. Chem.* **1997**, *272* (35), 22364.

(164) Hess, S.; Lindquist, S. L.; Scheibel, T. *EMBO Rep.* **2007**, *8* (12), 1196.

(165) Bouchard, M.; Zurdo, J.; Nettleton, E. J.; Dobson, C. M.; Robinson, C. V. *Protein Sci.* **2000**, *9* (10), 1960.

(166) Juárez, J.; Taboada, P.; Mosquera, V. *Biophys. J.* **2009**, *96* (6), 2353.

(167) Cukalevski, R.; Yang, X.; Meisl, G.; Weininger, U.; Bernfur, K.; Frohm, B.; Knowles, T. P. J.; Linse, S. *Chem. Sci.* **2015**, *6* (7), 4215.

(168) Daly, S.; Kulesza, A.; Poussigue, F.; Simon, A.-L.; Choi, C. M.; Knight, G.; Chirot, F.; MacAleese, L.; Antoine, R.; Dugourd, P. *Chem. Sci.* **2015**, *6* (8), 5040.

(169) Seo, J.; Hoffmann, W.; Warnke, S.; Huang, X.; Gewinner, S.; Schöllkopf, W.; Bowers, M. T.; von Helden, G.; Pagel, K. *Nat. Chem.* **2017**, *9*, 39.

(170) Kamihira, M.; Naito, a; Tuzi, S.; Nosaka, a Y.; Saitô, H. *Protein Sci.* **2000**, *9*, 867.

(171) Scheidt, H. A.; Morgado, I.; Huster, D. *J. Biol. Chem.* **2012**, *287* (27), 22822.

(172) Jaronieci, C. P.; MacPhee, C. E.; Bajaj, V. S.; McMahon, M. T.; Dobson, C. M.; Griffin, R. G. *Proc Natl Acad Sci U S A* **2004**, *101* (3), 711.

(173) Karamanos, T. K.; Kalverda, A. P.; Thompson, G. S.; Radford, S. E. *Prog. Nucl. Magn. Reson. Spectrosc.* **2015**, *88–89*, 86.

(174) Gazit, E. *FASEB J.* **2002**, *16* (1), 77.

(175) Gazit, E. *Chem. Soc. Rev.* **2007**, *36* (8), 1263.

(176) Kuang, Y.; Gao, Y.; Zhang, Y.; Gao, P.; Xu, B. *J Am Chem Soc* **2010**, No. 5, 805.

(177) Reches, M.; Gazit, E. *Curr. Nanosci.* **2006**, *2* (2), 105.

(178) Martinez, C. R.; Iverson, B. L. *Chem. Sci.* **2012**, *3* (7), 2191.

(179) Abdallah, D. J.; Weiss, R. G. *Adv. Mater.* **2000**, *12* (17), 1237.

(180) Terech, P.; Weiss, R. G. *Chem. Rev.* **1997**, *97* (8), 3133.

(181) Weiss, M. G.; Weiss, R. G. *Acc. Chem. Res.* **2006**, *39* (8).

(182) Krysmann, M. J.; Castelletto, V.; Kelarakis, A.; Hamley, I. W.; Hule, R. a; Pochan, D. J. *Biochemistry* **2008**, 4597.

(183) Scheuner D., Eckman C., Jensen M., Song X., Citron M., Suzuki N., Bird T.D., Hardy J., Hutton M., Kukull W., Larson E., Levy-Lahad E., Viitanen M., Peskind E., Poorkaj P., Schellenberg G., Tanzi R., Wasco W., Lannfelt L., Selkoe D., Y. S. *Nat Med.* **1996**, 2, 864.

(184) Tanzi, R. E.; Gusella, J. F.; Watkins, P. C.; Bruns, G. A. P.; St George-Hyslop, P.; Van Keuren, M. L.; Patterson, D.; Pagan, S.; Kurnit, D. M., Neve, R. L. *Science* **1987**, *235* (4791), 880.

(185) Pike, C. J.; Burdick, D.; Walencewicz, a J.; Glabe, C. G.; Cotman, C. W. *J. Neurosci.* **1993**, *13* (4), 1676.

(186) Bertolani, A.; Pirrie, L.; Houbenov, N.; Haataja, J.; Stefan, L.; Catalano, L.; Terraneo, G.; Giancane, G.; Valli, L.; Milani, R.; Ikkala, O.; Resnati, G.; Metrangolo, P. *Nat. Commun.* **2015**, *6* (7574), 1.

(187) Perutz, M. F. *Acta Crystallogr.* **1956**, *9* (11), 867.

(188) Blow, D. M.; Rossmann, M. G. *Acta Crystallogr.* **1962**, *15* (10), 1060.

(189) Kreutzer, A. G.; Hamza, I. L.; Spencer, R. K.; Nowick, J. S. *J. Am. Chem. Soc.* **2016**, *138* (13), 4634.

(190) Pham, J. D.; Spencer, R. K.; Chen, K. H.; Nowick, J. S. *J. Am. Chem. Soc.* **2014**, *136* (36), 12682.

(191) Bertolani, A.; Pizzi, A.; Pirrie, L.; Gazzera, L.; Morra, G.; Meli, M.; Colombo, G.; Genoni, A.; Cavallo, G.; Terraneo, G.; Metrangolo, P. *Chem. Eur. J.* ***2017**, 23, 2051*.

(192) Pham, J. D.; Chim, N.; Goulding, C. W.; Nowick, J. S. *J Am Chem Soc* **2013**, *135* (33), 12460.

(193) Liu, C.; Sawaya, M. R.; Cheng, P. N.; Zheng, J.; Nowick, J. S.; Eisenberg, D. *J. Am. Chem. Soc.* **2011**, *133* (17), 6736.

Chapter 2

Halogen bonding improves the hydrogelation properties of *N*-Fmoc-phenylalanines

2.1 Introduction

One of the most popular and effective ways to promote the self-assembly over length scales from the molecular level to a macroscopic scale to yield massive materials is to use small molecules that are able to form gels, namely LMOG (Low Molecular-Mass Organic Gelators).[1–3] Indeed, LMOGs are able to spontaneously form large aggregates with high aspect ratio. The resulting micrometer-length fibers form an entangled network hindering the flow of the aqueous or organic solvent, forming a material that exhibits properties in between the solid and the liquid phase, namely a hydro- or an organogel.[4] The intrinsic versatility of such kind of systems has given access to a plethora of applications,[5–11] allowing a rationalization of general structure-properties relationships. In this field, peptide and amino acid-based LMOGs are widely studied and particularly promising.[12] Indeed, their hydrogels showed unique merits thanks to their easy design and modification, biocompatibility[13] and versatility to adapt to multiple functions, with the additional advantage of a relatively low cost.[14] In fact, the above-mentioned LMOG hydrogelators had already found applications in different areas such as tissue engineering or scaffolding,[15–18] sensing of analytes,[19,20] drug delivery[21] or targeting.[22] In this contest, a commonly recognized way to enhance the hydrogelation performances of amino-acid based hydrogelators is to introduce hydrophobic and π-π interactions through an aromatic capping group.[23,24] Among these aromatic moieties, the 9-fluorenyl-methoxycarbonyl (Fmoc) group is so far the best established[25], especially for simple systems like phenylalanines.[26–28] In this latter case, systematic studies were focused on the tuning of the gelation properties through slight structural modifications, but a general scenario is still not defined.[29–31] The production of complex and functional self-assembled structures from simple, easy to handle and cheap building blocks is a desirable requirement in supramolecular chemistry. Since halogen bonding (XB) is a promising and regarded supramolecular tool also in the field of bioinspired materials,[32–34] the possibility to use the peculiarities of halogen atoms and XB to modulate hydrogelation properties of simple LMOGs was deeply investigated.

2.2 Materials

All the reagents were purchased from Sigma-Aldrich and used without any further purification.

2.3 Methods

2.3.1 Transmission Electron Microscopy (TEM). TEM bright field images were acquired using a Philips CM200 electron microscope operating at 200 kV equipped with a Field Emission Gun filament. A Gatan US 1000 CCD camera was used and 2048x2048 pixels images with 256 grey levels were recorded. The suspension were dropped onto a 200 mesh carbon-coated copper grid and air dried for several hours before analysis.

2.3.2 Rheology. All rheology tests were performed using a Physica-Paar UDS 200 rheometer, equipped with a plate–plate geometry measuring system (diameter of the upper plate 4.0 cm, measuring gap: 500 µm). Temperature was controlled with a Peltier device and maintained at 25°C. All the oscillatory measurements were performed within the linear viscoelastic range of gels. The instrumental setups for the rheology tests are the following: Frequency sweep test: frequency range 100 Hz to 0.001 Hz; strain 1%. All the measurements were repeated a minimum of three times.

2.3.3 Fibrillation kinetics. Measurements were performed on a Tecan GENios plate reader employing as a sample holder a full transparent 96 wellplate. Each well was filled with 500 µL of the proper peptide solution. Each sample was prepared in triplicate. Method: (i) shaking 30 s; (ii) wait 150 s; (iii) shake 2 s; (iv) measure absorption at 405 nm; (v) repeat stages ii-iv overnight. Each sample was subtracted with the blank solution, i.e. 5% DMSO in PBS Buffer solution (pH = 7.4).

Each measurement was fitted accordingly to equation 1 (Verhulst logistic function):[35]

$$Y = y_i \frac{y_f}{1 + e^{-(t-t_0)/\tau}}$$

Where: Y = absorbance, y_i and y_f being the initial and final absorbance values respectively;

t_0 = time to 50% of maximal signal or the time of inflection point;

τ = time constant of fibril growth

From the fitting, two parameters that are related to the fibril growth process can be extrapolated:

$k_{app} = 1/\tau$ apparent growth rate of fibrils

$T = t_0 - 2\tau$ delay time

The higher is k_{app}, the faster is the fibrillation, while the higher is the delay time, the slower is the kinetic fibrillation.

2.3.4 Gelation experiments. Hydrogels were prepared in small (1 mL) capped vials by adding PBS buffer (pH = 7.4) to a proper volume of a DMSO stock solution of the amino acid to reach the desired concentration and the desired percentage of DMSO in the final mixture. The resulting suspensions (upon addition of PBS to the DMSO stock solution, the amino acid precipitates), was vortexed for 30 s at 20x100 rpm, and heated using a heat gun until complete dissolution. The resulting solution were allowed to cool to r.t. and tested for the gel formation through the "tube inversion test".

2.3.5 Thermal stability. The vials containing the hydrogels at 5 mM or 1 mM concentration were inverted and fixed at the bottom of an oil bath with stirring. The temperature was kept at 25 °C for 10 min to equilibrate the system and then gradually increased from 25 to 120 °C (at 1 °C min^{-1}). The range temperature within the gels break and fall down is regarded as melting point range.

2.3.6 Crystallization assays. Fmoc-4-F-Phe crystals were obtained by dissolving the compound in a mixture of DMSO/water 95:5. Crystals suitable for XRD analysis formed after 1 month of slow evaporation at room temperature. Fmoc-4-Cl-Phe crystals were obtained by dissolving the compound in N-methyl acetamide. Crystals suitable for XRD analysis formed after 1 month of slow evaporation at 45 °C. Fmoc-4-Br-Phe crystals were obtained by dissolving the compound in a mixture of hexafluoro-2-propanol/water 95:5. Crystals suitable for XRD analysis formed after 1 month of slow evaporation at room temperature. Fmoc-4-I-Phe crystals were obtained by dissolving the compound in a mixture of dichloromethane/hexafluoro-2-propanol 1:1. Crystals suitable for XRD analysis formed after 1 month of slow evaporation at room temperature. Data collection was performed at the X-ray diffraction beamline (XRD1) of the Elettra Synchrotron, Trieste (Italy), proposal numbers 20150542 and 20155517.

2.3.7 Small Angle X-ray Scattering (SAXS). Synchrotron X-ray scattering data were collected at the Austrian SAXS beamline at ELETTRA (Trieste, Italy), proposal number 20160276. The measurements were carried out at a wavelength of 0.154 nm and a sample-detector distance of 1.1 m. The photon energy was 8 keV. The data were recorded with a Pilatus detector (PILATUS 100 K, DECTRIS Ltd., Villigen PSI, Switzerland), calibrated with silver behenate. The scattering intensity was measured as a function of the scattering vector (q)

$$q = 4\pi (\sin\vartheta)/\lambda,$$

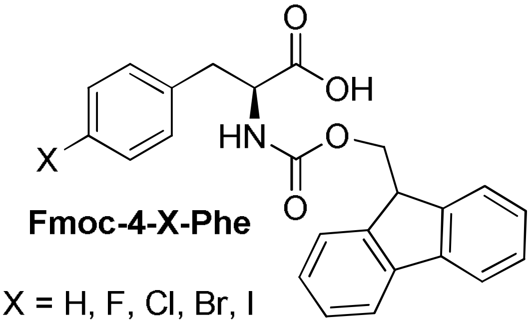
where 2ϑ is the scattering angle and λ is the wavelength. Data were analyzed with Igor Pro (Version 6.37, WaveMetrics Inc., USA). Experimental intensities were normalized to sample transmission and corrected for background.

2.3.8 Single crystal X-ray Diffraction (XRD). Data collections were performed at the X-ray diffraction beamline (XRD1) of the Elettra Synchrotron, Trieste (Italy).[36] The crystals were dipped in perfluoropolyether vacuum oil (Fomblin) and mounted on the goniometer head with a nylon loop. Complete datasets were collected at 100 K (nitrogen stream supplied through an Oxford Cryostream 700) through the rotating crystal method. Data were acquired using a monochromatic wavelength of 0.700 Å for **Fmoc-4-Cl-Phe, Fmoc-4-Br-Phe**, **Fmoc-4-I-Phe** and 0.800 Å for **Fmoc-4-F-Phe**, on a Pilatus 2M hybrid-pixel area detector. The diffraction data were indexed and integrated using XDS.[37] Scaling has been done using CCP4-Aimless code.[38,39] Crystals appear as very thin colorless needles or plates prone to radiation damage, as previously reported for other halogenated molecules.[40,41] None of the crystals tested diffracted better than 0.85 Å, with an average dataset resolution of ~0.9 Å. For the brominated aminoacid two different datasets, collected from different crystals randomly oriented, have been merged. The structures were solved by the dual space algorithm implemented in the SHELXT code.[42] Fourier analysis and refinement were performed by the full-matrix least-squares methods based on F^2 implemented in SHELXL-2014.[43] The Coot program was used for modeling.[44] Anisotropic thermal motion refinement have been used for all atoms with full occupancy. Geometric and thermal motion parameters restraints (DFIX, DANG, SIMU or DELU) have been applied on disordered and poorly defined fragments. Hydrogen atoms were included at calculated positions with isotropic $U_{factors} = 1.2\ U_{eq}$ or $U_{factors} = 1.5$ Ueq for hydroxyl groups (Ueq being the equivalent isotropic thermal factor of the bonded non-hydrogen atom). Refined Flack parameters[45] (Table 1S) have quite high e.s.d., especially for fluorine derivative, as a consequence of crystal radiation damage but are in agreement with expected phenylalanine $C\alpha$ configuration.

2.4 Results and discussion

Herein we report the production and thorough characterization of *N*-Fmoc-4X-Phenylalanine (**Fmoc-4-X-Phe**, X = H, F, Cl, Br, I, Fig. 1) hydrogels. Even if such systems have already been studied,[26,27,29–31] with conflicting results between authors, in the proposed experimental conditions an unprecedented behaviour was detected, *i.e.* the dependence of all the observable parameters related to self-assembly efficiency is in agreement with the commonly expected efficiency of halogen atoms in the contest of XB interactions.

Due to the large set of non-covalent interactions given by highly polarizable halogen atoms, we predicted that the self-assembly of iodinated and brominated derivatives of **Fmoc-4-X-Phe** should be more efficient. In particular, we expected that the anisotropic behavior of iodine and bromine should allow them to act both as electrophilic and nucleophilic species, enlarging their potential interaction pattern.



Fmoc-4-X-Phe

Figure 1: Molecular formula of the LMOG used in this study, namely *N*-Fmoc-4-substituted-phenylalanines (Fmoc-4-X-Phe).

2.4.1 Fibrillation propensity and hydrogelation

In order to confirm our hypothesis, we monitored the fibril growth of each halogenated compound over time by measuring the UV-Vis absorbance of **Fmoc-4-X-Phe** diluted solutions (0.1 mM, 5% DMSO in PBS) under controlled shaking conditions.[46] The occurrence of the fibrillation process results in a sigmoidal curve, showing the three characteristic stages of fibril formation: a lag phase, a rapid growth representing the elongation process and a final stage of equilibrium. An empirical function, already reported in the literature for similar studies has been selected to fit the experimental data[46]. In this way, it was possible to extrapolate two different kinetic parameters: the delay time (T) and the apparent rate constant of fibril growth (k_{app}).

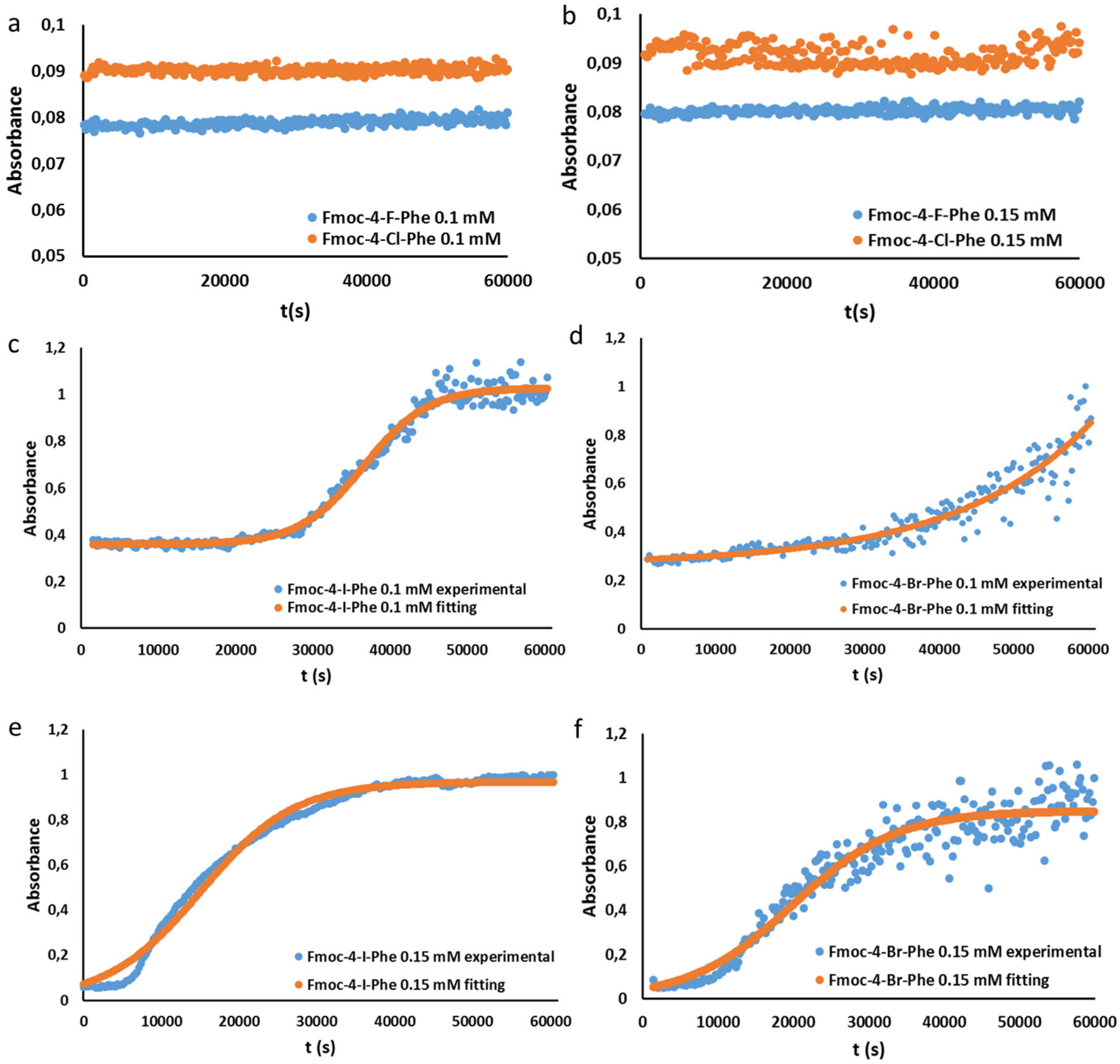
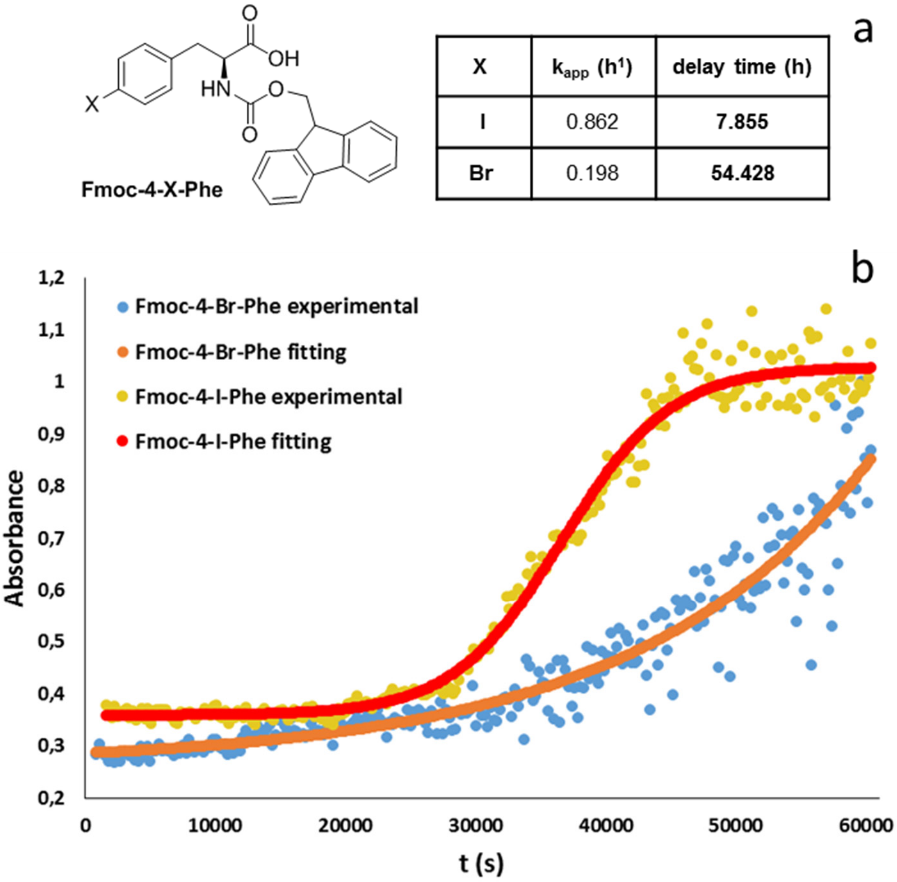


Figure 2: Fibrillation Kinetics of Fmoc-4-X-Phe. Fluorinated and chlorinated amino acids at 0.1 mM concentration (a) do not show the sigmoidal curve typical of fibrillation process. Increasing the concentration (b) has no effect on the kinetic. Fmoc-4-I-Phe 0.1 mM kinetics (c) follows the expected trend. Fmoc-4-Br-Phe 0.1 mM kinetics (d) is still in the growing phase in the considered time interval. Increasing the concentration of Fmoc-4-I-Phe leads to a faster kinetic (e). Increasing the concentration of Fmoc-4-Br-Phe leads the fibrillation process to completeness.

These parameters indicate that the most fibrillogenic compound is the iodinated one, followed by the bromine derivative, having a delay time one order of magnitude greater with respect to **Fmoc-4-I-Phe**. The higher fibrillation propensity of the iodinated compound is evident also graphically, since in the same time period the kinetics of the iodinated compound is complete, while for the bromo- derivative it is still in the growth phase (Fig. 3).



X	k_{app} (h^{1})	delay time (h)
I	0.862	**7.855**
Br	0.198	**54.428**

Figure 3: a) Table with the parameters extrapolated for the Fmoc-4-Br-Phe and Fmoc-4-I-Phe samples from the fitting of the UV kinetic measurements b) fibrillation kinetic parameters for 0.1 mM Fmoc-4-Br-Phe and Fmoc-4-I-Phe samples. The data with fitted profiles is also reported.

While it was possible to calculate kinetic parameters for iodo- and bromo- compounds, it was not the case for both fluorinated and chlorinated amino acids whose absorbance does not change over time. Further confirmation of the trend resulting from the kinetic study are the minimum gelation concentration (MGC) values (Table 1). As expected, the best gelator is the iodinated amino acid (MGC = 0.25 mM), followed in order by bromo-, chloro- and fluoro-compounds (0.5, 0.75, 3.0 mM respectively). The same working concentration of 5 mM was used to characterize the mechanical properties of the gels by rheology (Fig. 4). In addition to the halogenated derivatives, also the gel formed by the non-modified amino acid was studied for comparison. In general, the storage modulus value increases as a function of halogen atom polarizability (Fig. 5a), confirming the same trend observed with both fibrillation kinetics and MGC values.

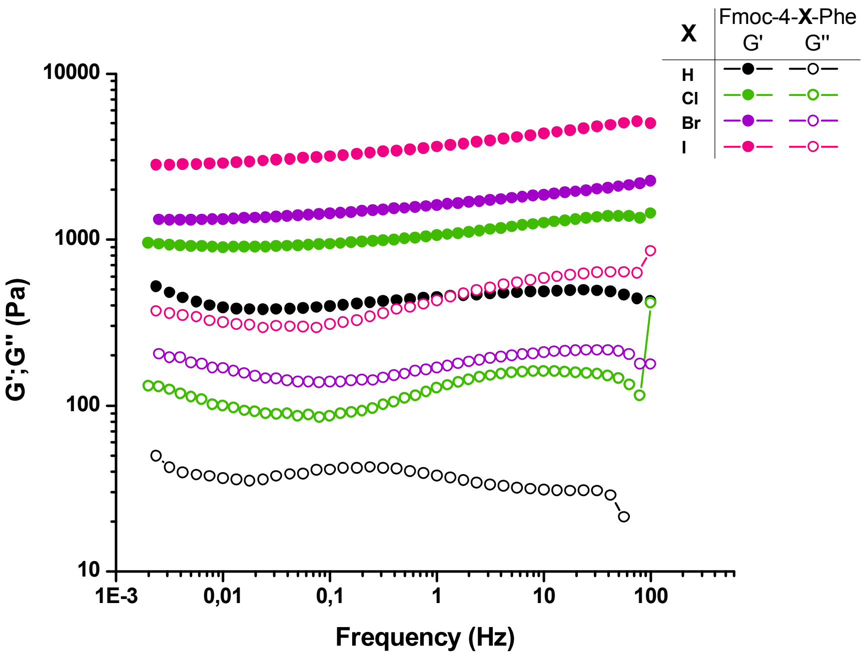


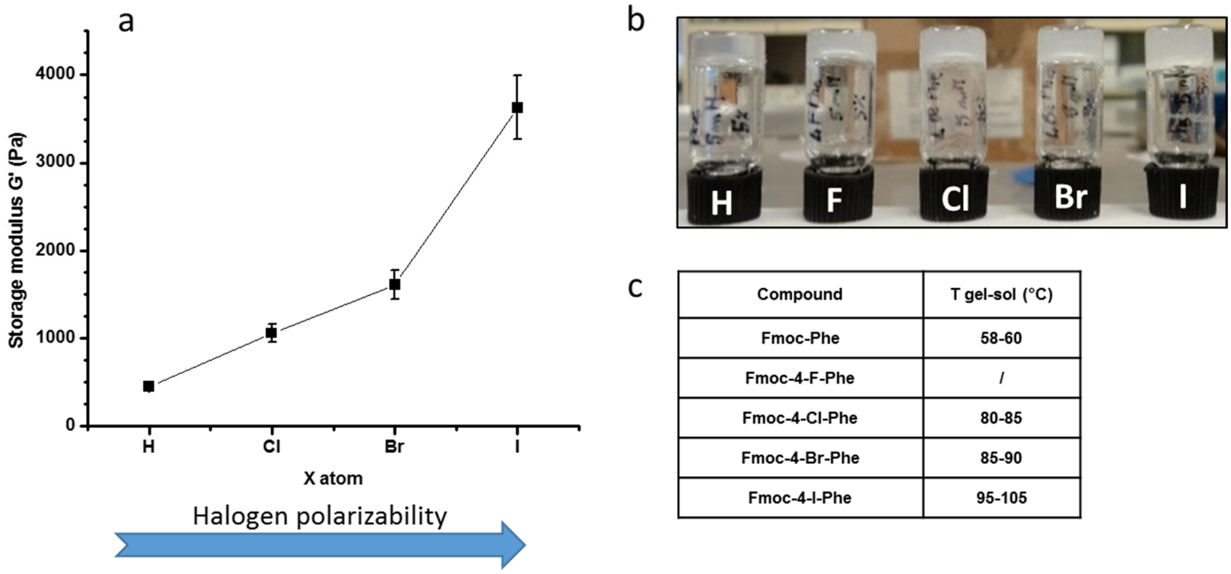
Figure 4: Rheology measurements on Fmoc-4-X-Phe 5mM gel samples in 5% DMSO/PBS. The storage (G') and the loss (G'') moduli are reported as a function of the oscillatory frequency. The measurements are performed in the linear viscoelastic range of the gels. G' values are systematically higher than G'', confirming that the observed samples were true gels. The gel were quite robust, being the one with X=I the stiffest one.

This finding demonstrates that fibrillation rate and hydrogel rigidity are intimately related in this kind of systems. It was not possible to run the analysis on the fluorinated molecule, since it forms a very weak gel, prohibiting reliable and reproducible measurements. We evaluated the gel strength also in term of gel-sol transition temperature, finding the same order of stability (Fig. 5).

Table 1: Minimum Gelation Concentration (MGC) for Fmoc-4-X-Phe in 5% DMSO/PBS.

AMINO ACID	MGC (mM)
Fmoc-Phe	5
Fmoc-4-F-Phe	3 *
Fmoc-4-Cl-Phe	0.75
Fmoc-4-Br-Phe	0.50
Fmoc-4-I-Phe	0.25

* Poorly stable towards inversion



Compound	T gel-sol (°C)
Fmoc-Phe	58-60
Fmoc-4-F-Phe	/
Fmoc-4-Cl-Phe	80-85
Fmoc-4-Br-Phe	85-90
Fmoc-4-I-Phe	95-105

Figure 5: Storage modulus G’ (at ω = 1 Hz) as a function of the X atom of the Fmoc-4-X-Phe LMOGs a). A picture of the Fmoc-4-X-Phe 5mM gels in 5% DMSO/PBS b). A table showing the gel-sol transition ranges of each compound.

Transmission electron microscopy (TEM) allowed assessing the morphology of the supramolecular structures constituting the hydrogel network. All the compounds showed fibrillar networks (Fig. 6), with increasing entanglement and fibril dimensions (demonstrated by statistical analysis of the TEM images, Fig. 7) moving from fluorine to iodine.

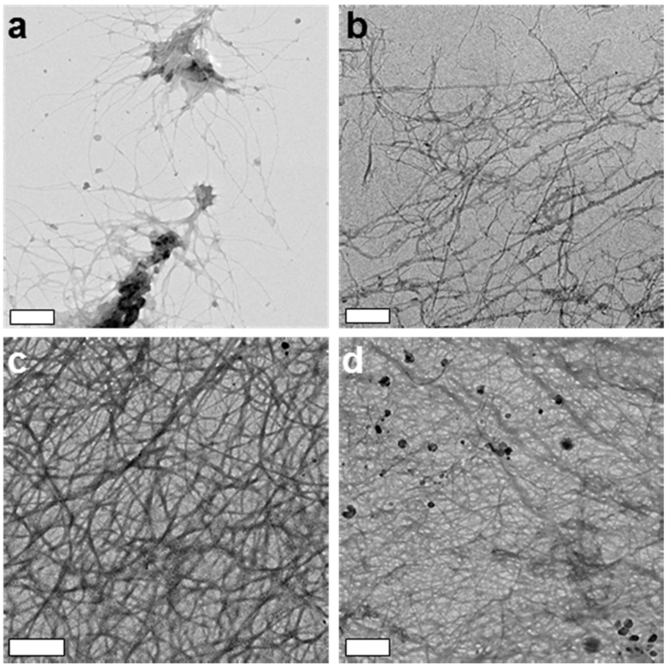


Figure 6: TEM images of 1 mM samples of **Fmoc-4-X-Phe** showing an increasing entanglement of the fibrillar network. Fmoc-4-F-Phe a) Fmoc-4-Cl-Phe b) Fmoc-4-Br-Phe c) Fmoc-4-I-Phe d) Scale bar 0.5 μm.

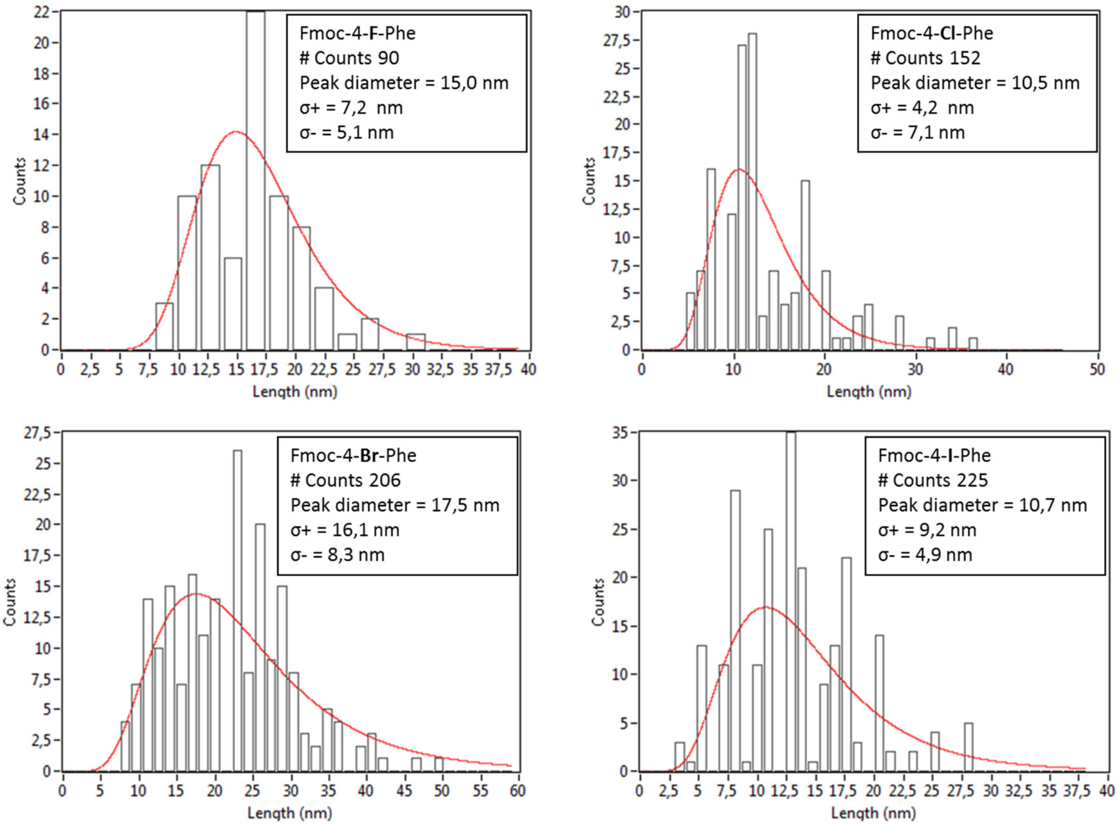


Figure 7: Statistical analysis of diameter of fibers observed in the TEM micrographs of 1 mM gel sample of Fmoc-4-X-Phe in 5% DMSO/PBS. The analysis was performed with a home-made routine programs (NPStat) that has already been used in the statistical analysis of dimension of nanostructured materials observed by TEM.[47] The analysis shows that the dimension of fibers is similar amongst the different sample, the bromo derivative having a slightly bigger and less monodisperse diameter range.

It has already been demonstrated how single-crystal X-Ray diffraction analysis of gelling molecules can be helpful to elucidate the main interactions that are then translated in the gel aggregates, being thus responsible for supramolecular assembly in the gel phase.[48–51] In particular, the correlation among crystals and gels has been proved also for XB active gelators.[51] In accordance with these findings, we analysed the crystal structures of the **Fmoc-4-X-Phe** series by single crystal X-ray diffraction (XRD), with the purpose to find out a connection between the hydrogelation properties of Fmoc-Phe derivatives and the role of halogen atoms in the self-assembly. Since all the halogenated amino acids form small and poorly diffracting crystals, a very powerful X-ray source like synchrotron radiation was necessary to solve the structures. The XRD experiments further support the hypothesis that the most polarizable halogen atoms give rise to more stable and efficient supramolecular assemblies. The crystal structures of the brominated and iodinated compounds show in fact the halogen atoms acting as electrophilic species, forming short contacts with the electron density of the Fmoc moiety of neighboring amino acid molecules

(Fig 8a; Fig 8b). These halogen-π interactions – which are halogen bonds – lead to the formation of closed repetition units constituted by three molecules of **Fmoc-4-X-Phe** (X=Br, I). These three-membered units stack along the *c* crystallographic axis through N–H···O hydrogen bonds between the carbonyl group of the Fmoc moiety and the N–H group of an adjacent amino acid molecule, leading to close-packed and elongated assemblies (Fig. 8c; Fig 8d).

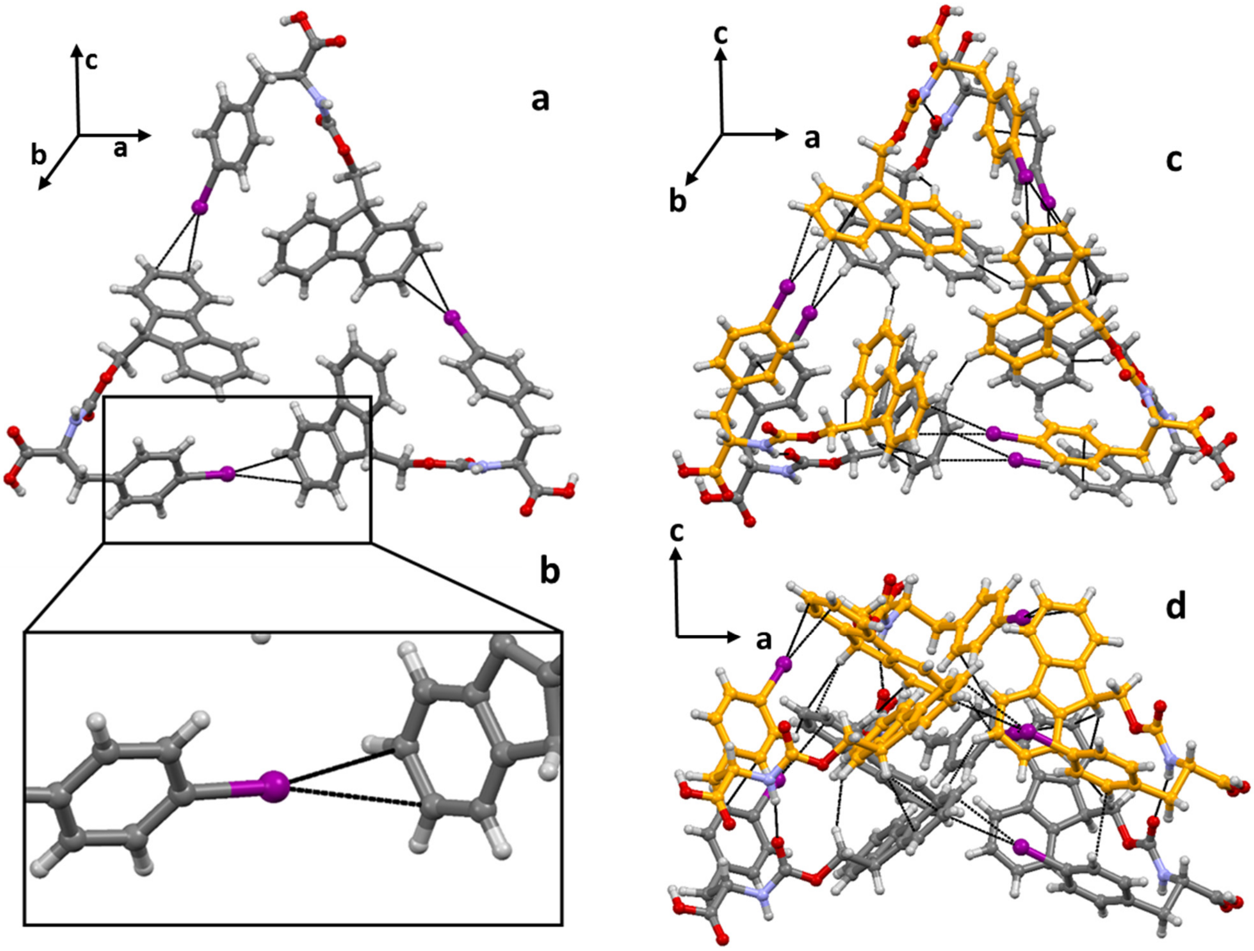


Figure 8: Crystal structure of Fmoc-4-I-Phe. Halogen···π contacts lead to formation of three-membered units. a) View along *c* crystallographic axis. b) Magnification of halogen···π contacts involving iodine and the electron density of the Fmoc moiety. c) d) Lateral views of stacked repetition units, with hydrogen bonds stabilizing the resulting supramolecular structure.

This supramolecular structure can further grow in diameter by lateral association of multiple three-membered units lying on the same plane (Figures 12c, 13c). Such a compact architecture, which is the results of the contribution of iodine and bromine displaying their peculiar amphoteric behavior, has not been noticed for the chlorinated and fluorinated compounds. As they are less polarizable, chlorine and fluorine do not act as electrophilic species in the crystal structure of **Fmoc-4-Cl-Phe** and **Fmoc-4-F-Phe**.

Chlorine shows weak intermolecular contacts Cl···C and Cl···H, while fluorine does not give rise to any interaction. Although the compounds of the series were crystallised avoiding the use of the saline solution of PBS, which is mandatory to have gel formation, the higher tendency of brominated and iodinated amino acids to arrange in more stable assemblies is evident and unambiguous. This suggests that, even in the presence of possible XB-interfering species, *i.e.* the solvent, the halogen atoms have a strong tendency to profit by the self-complementary nature of the studied LMOG, giving rise to interactions that reinforce their self-assembled superstructures, to whose robustness also hydrogen bonds and π-stacking interactions contribute. Considering the frequent occurrence of XBs in several crystallization conditions, it cannot be overlooked as key interaction for hydrogel formation in the Fmoc-Phe system. Therefore, the crystal structures of the **Fmoc-4-X-Phe** series can represent a reliable model explaining the reason for growing gel strength, growing thermal stability and higher fibrillogenic propensity by increasing halogen atom polarizability. More details on the crystal structures of the **Fmoc-4-X-Phe** series are shown below.

2.4.2 Crystal structures of Fmoc-4-X-Phe

The four amino acids crystallize in chiral space groups with one molecule in the asymmetric unit (ASU; Figure 9). Crystal packing of all the halogenated amino acids show stacked molecules with strong hydrogen bonds involving neighboring peptide bonds and terminal protonated carboxylic groups. Furthermore, aromatic interactions keep Fmoc moieties and phenylalanine rings significantly overlapped.

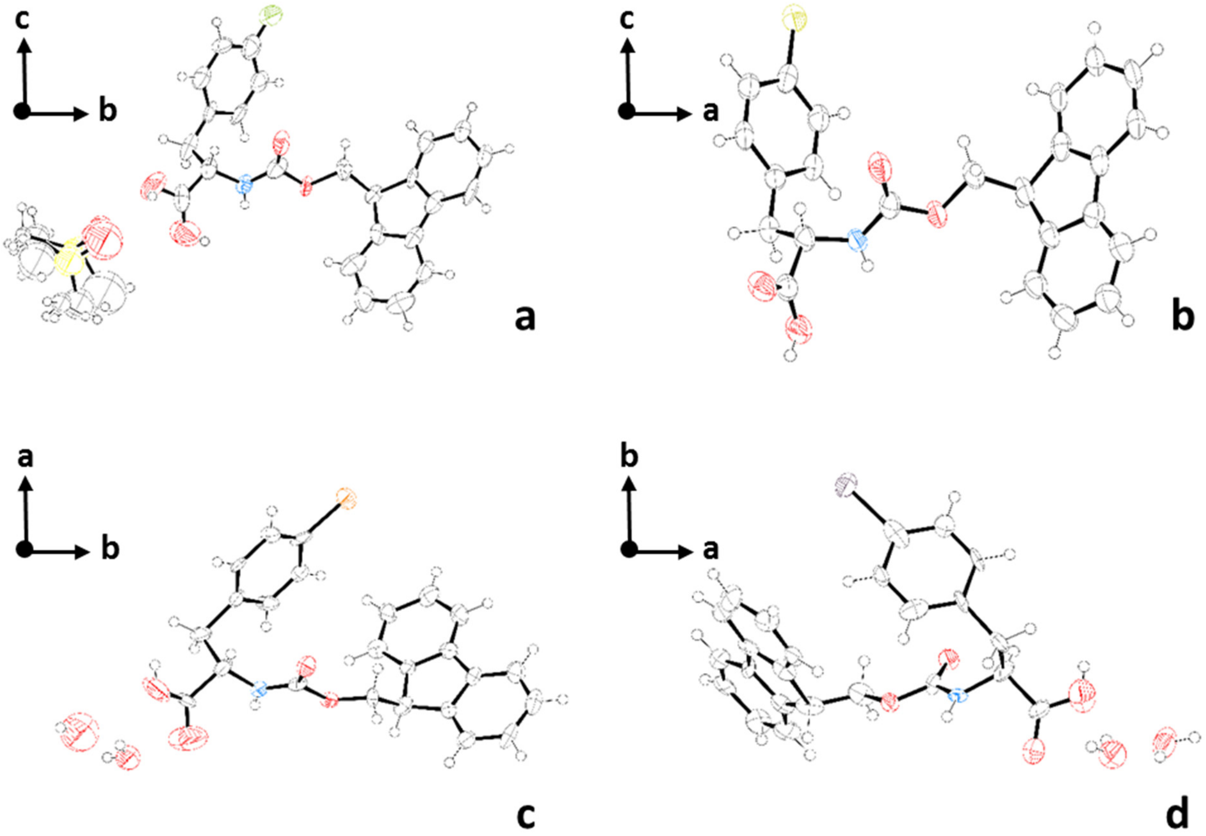


Figure 9: Ellipsoids representation of ASU contents (50% probability) for: **Fmoc-4-F-Phe** a), **Fmoc-4-Cl-Phe** b), **Fmoc-4-Br-Phe** c), and **Fmoc-4-I-Phe** d). Pictures were prepared using Ortep3 software.

The molecule that contains a fluorinated amino acid shows an orthorhombic unit cell that traps a disordered DMSO solvent molecule inside channels parallel to the crystallographic *a* axis.

Bromine and iodine molecules show the same hexagonal crystal packing, trapping water molecules in channels aligned with crystallographic *c* axis. The water molecules lie on crystallographic 6_3-screw axes and keep tightly connected hexamer of molecules exposing their carboxylic groups in the channels. The cell volume is slightly bigger for the iodinated peptide, as expected from comparison of halogens atomic radius.

Essential crystal and refinement data are reported below (Table 2).

Table 2: Crystallographic data and refinement details for compounds **Fmoc-4-F-Phe**, **Fmoc-4-Cl-Phe**, **Fmoc-4-Br-Phe** and **Fmoc-4-I-Phe.**

	Fmoc-4-F-Phe·C_2H_6SO [$C_{24}H_{20}FNO_4$·C_2H_6SO]	**Fmoc-4-Cl-Phe** [$C_{24}H_{20}ClNO_4$]	**Fmoc-4-Br-Phe**·⅓H_2O [$C_{24}H_{20}BrNO_4$·⅓H_2O]	**Fmoc-4-I-Phe**·⅓H_2O [$C_{24}H_{20}INO_4$·⅓H_2O]
CCDC Number	-	-	-	-
Chemical Formula	$C_{26}H_{26}FNO_5S$	$C_{26}H_{26}FNO_5S$	$C_{24}H_{20.67}BrNO_{4.33}$	$C_{24}H_{20.67}INO_{4.33}$
Formula weight (g/mol)	483.54	421.86	472.31	519.31
Temperature (K)	100(2)	100(2)	100(2)	100(2)
Wavelength (Å)	0.800	0.700	0.700	0.700
Crystal system	Orthorhombic	Monoclinic	Hexagonal	Hexagonal
Space Group	$P\ 2_12_12_1$	$P\ 2_1$	$P\ 6_3$	$P\ 6_3$
Unit cell dimensions	a = 4.920(1) Å b = 13.034(3) Å c = 36.789(7) Å α = 90° β = 90° γ = 90°	a = 13.168(3) Å b = 4.839(1) Å c = 17.400(4) Å α = 90° β = 111.41(3)° γ = 90°	a = 26.744(4) Å b = 26.744(4) Å c = 4.997(1) Å α = 90° β = 90° γ = 120°	a = 26.878(4) Å b = 26.878(4) Å c = 4.999(1) Å α = 90° β = 90° γ = 120°
Volume (Å^3)	2359.2(8)	1032.2(4)	3095.2(11)	3127.6(11)
Z	4	2	6	6
Density (calculated) (g·cm^{-3})	1.633	1.357	1.520	1.654
Absorption coefficient (mm^{-1})	0.246	0.205	1.893	1.458
F(000)	1016	440	1448	1556
Crystal size (mm^3)	0.09 x 0.01 x 0.01	0.10 x 0.05 x 0.02	0.05 x 0.05 x 0.02	0.09 x 0.01 x 0.01
Crystal habit	Colorless thin needles	Colorless thin needles	Colorless thin plates	Colorless thin needles
Theta range for data collection	1.63° to 21.61°	1.24° to 24.31°	1.50° to 22.86°	1.49° to 21.61°
Index ranges	$0 \leq h \leq 5$ $0 \leq k \leq 14$ $0 \leq l \leq 40$	$-15 \leq h \leq 15$ $-5 \leq k \leq 5$ $-20 \leq l \leq 20$	$-25 \leq h \leq 25$ $-25 \leq k \leq 25$ $-5 \leq l \leq 3$	$-26 \leq h \leq 13$ $-25 \leq k \leq 28$ $-5 \leq l \leq 3$
Reflections collected	5407	5367	4309	4718
Independent reflections (data with I>2σ(I))	1904 (1362)	3196 (2068)	2529 (1639)	2162 (1263)
Data multiplicity (max resltn)	2.71 (2.67)	2.68 (2.94)	2.57 (2.34)	2.71 (2.67)
I/σ(I) (max resltn)	8.01 (1.74)	10.30 (1.88)	7.85 (1.61)	8.01 (1.74)
R_{merge} (max resltn)	0.0784 (0.4584)	0.0740 (0.4376)	0.1128 (0.4091)	0.0784 (0.4584)
Data completeness (max resltn)	96.4% (93.3%)	97.3% (100.0%)	98.3% (100.0%)	96.4% (93.3%)
Refinement method	Full-matrix least-squares on F^2	Full-matrix least-squares on F^2	Full-matrix least-squares on F^2	Full-matrix least-squares on F^2
Data / restraints / parameters	1904 / 17 / 277	3196 / 1 / 272	2529 / 10 / 230	2162 / 61 / 231
Goodness-of-fit on F^2	1.095	1.016	1.066	1.021
Δ/σ_{max}	0.008	0.000	0.008	0.001
Final R indices [I>2σ(I)]	$R_1 = 0.1081$, $wR_2 = 0.2778$	$R_1 = 0.0665$, $wR_2 = 0.1645$	$R_1 = 0.0691$, $wR_2 = 0.1693$	$R_1 = 0.0736$, $wR_2 = 0.1546$
R indices (all data)	$R_1 = 0.1391$, $wR_2 = 0.2994$	$R_1 = 0.1171$, $wR_2 = 0.1977$	$R_1 = 0.1207$, $wR_2 = 0.2006$	$R_1 = 0.1426$, $wR_2 = 0.1883$
Flack x parameter	0.384(565)	0.030(202)	0.047(45)	0.006(125)
Largest diff. peak and hole (eÅ^{-3})	0.461 and -0.394	0.418 and -0.506	0.569 and -0.702	0.920 and -0.761
R.M.S. deviation from mean (eÅ^{-3})	0.088	0.063	0.110	0.138

$R_1 = \Sigma\ ||F_o|-|F_c||\ /\ \Sigma\ |F_o|$

$wR_2 = \{\Sigma\ [w(F_o^2 - F_c^2)^2]\ /\ \Sigma\ [w(F_o^2)^2]\}^{½}$

2.4.2.1 Crystal structure of Fmoc-4-F-Phe

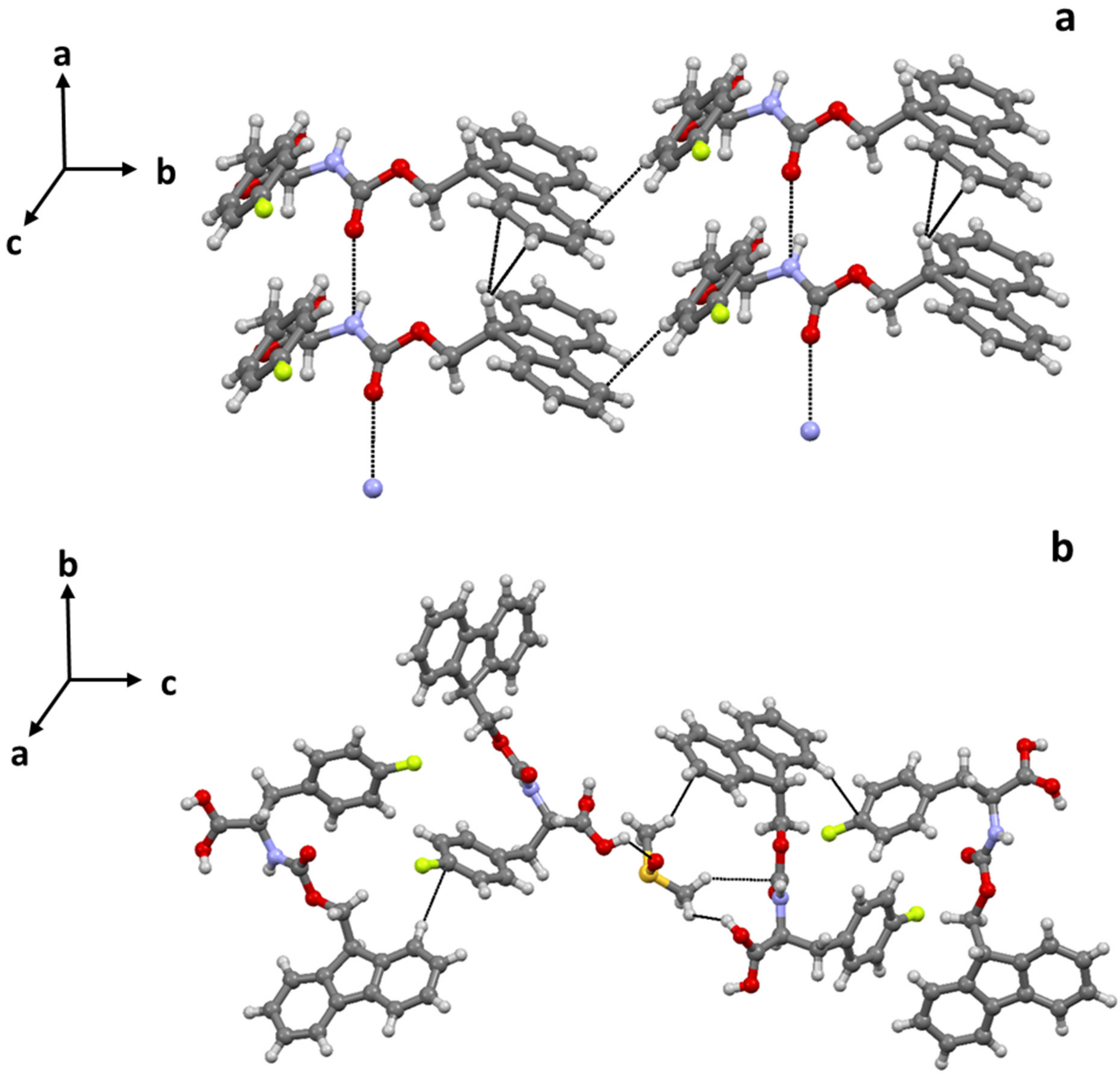


Figure 10: Crystal structure of Fmoc-4-F-Phe. a) Short contacts N···O lead to the formation of infinite chains along the crystallographic axis *a*. Neighboring chains interact with each other through C–H···π contacts involving the Fmoc moiety and the phenyl ring of different molecules. b) C–H···π contacts and other interactions involving DMSO stabilize the crystal packing along the *c* crystallographic axis.

Fmoc-4-F-Phe crystallizes in the $P\ 2_1 2_1 2_1$ space group. The asymmetric unit consists of a single molecule of amino acid solvated by one molecule of DMSO. The crystal structure grows along the *a* crystallographic axis through short contacts N···O between the carbonyl group of the Fmoc moiety and the nitrogen of an adjacent amino acid molecule, forming an infinite chain (N···O distance 2.85 Å; Fig 10a). The crystal packing is further stabilized by C–H···π interactions between the Fmoc moieties. Along the *b* crystallographic axis neighboring chains are kept connected through C–H···π contacts involving the phenyl ring and the Fmoc moiety of different amino acid molecules (Fig 10a). Along the crystallographic axis c the structure grows through C–H···π contacts between the Fmoc moiety and the phenyl ring of a near molecule. In the same direction, DMSO molecules join independent chains through intermolecular hydrogen bonding and short contacts C···H (Fig 10b).

2.4.2.2 Crystal structure of Fmoc-4-Cl-Phe

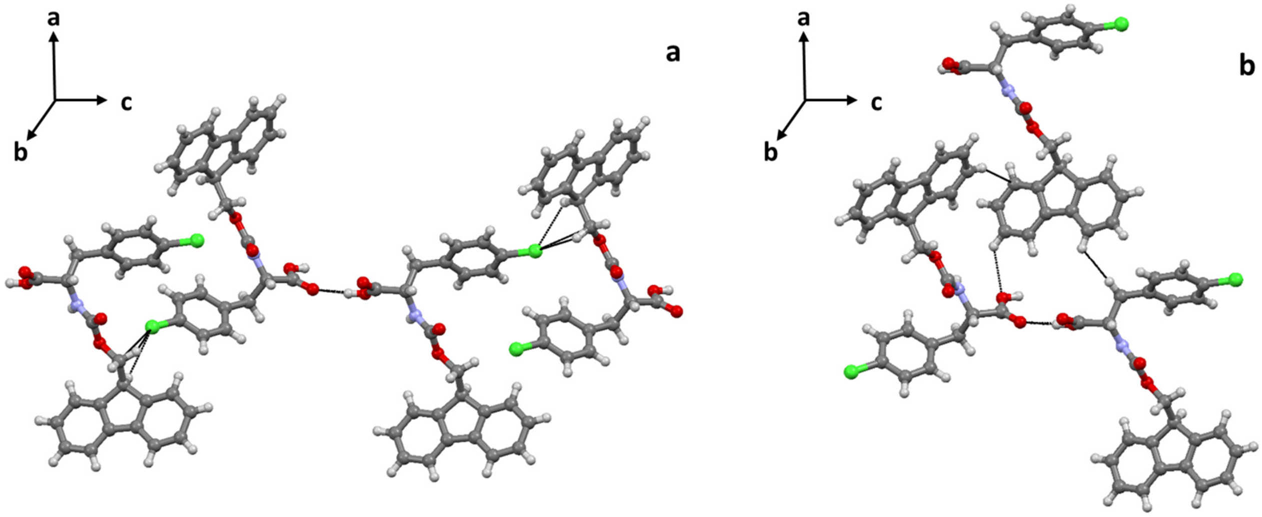


Figure 11: Crystal structure of Fmoc-4-Cl-Phe. a) Cl···C and Cl···H contacts confer stability to the crystal packing along the *a* crystallographic axis. b) C–H···π contacts among the Fmoc moieties further stabilize the crystal structure along axis *c*.

Fmoc-4Cl-Phe crystallizes in the *P* 2_1 space group. The asymmetric unit consists of a single molecule of amino acid. Chlorine shows short contacts Cl···C and Cl···H with the methylene group of the Fmoc moiety of a neighboring amino acid molecule (C···Cl distance 3.365 Å). The crystal structure develops along the crystallographic axis *a* through hydrogen bonding between the C-terminal groups of neighboring molecules (Fig 11a). Along the crystallographic axis *c* the structure grows through C–H···π contacts between the Fmoc moieties of adjacent molecules. In this direction, intermolecular short contacts H···H and C–H···O hydrogen bonding stabilize the crystal packing (Fig 11b).

2.4.2.3 Crystal structure of Fmoc-4-Br-Phe

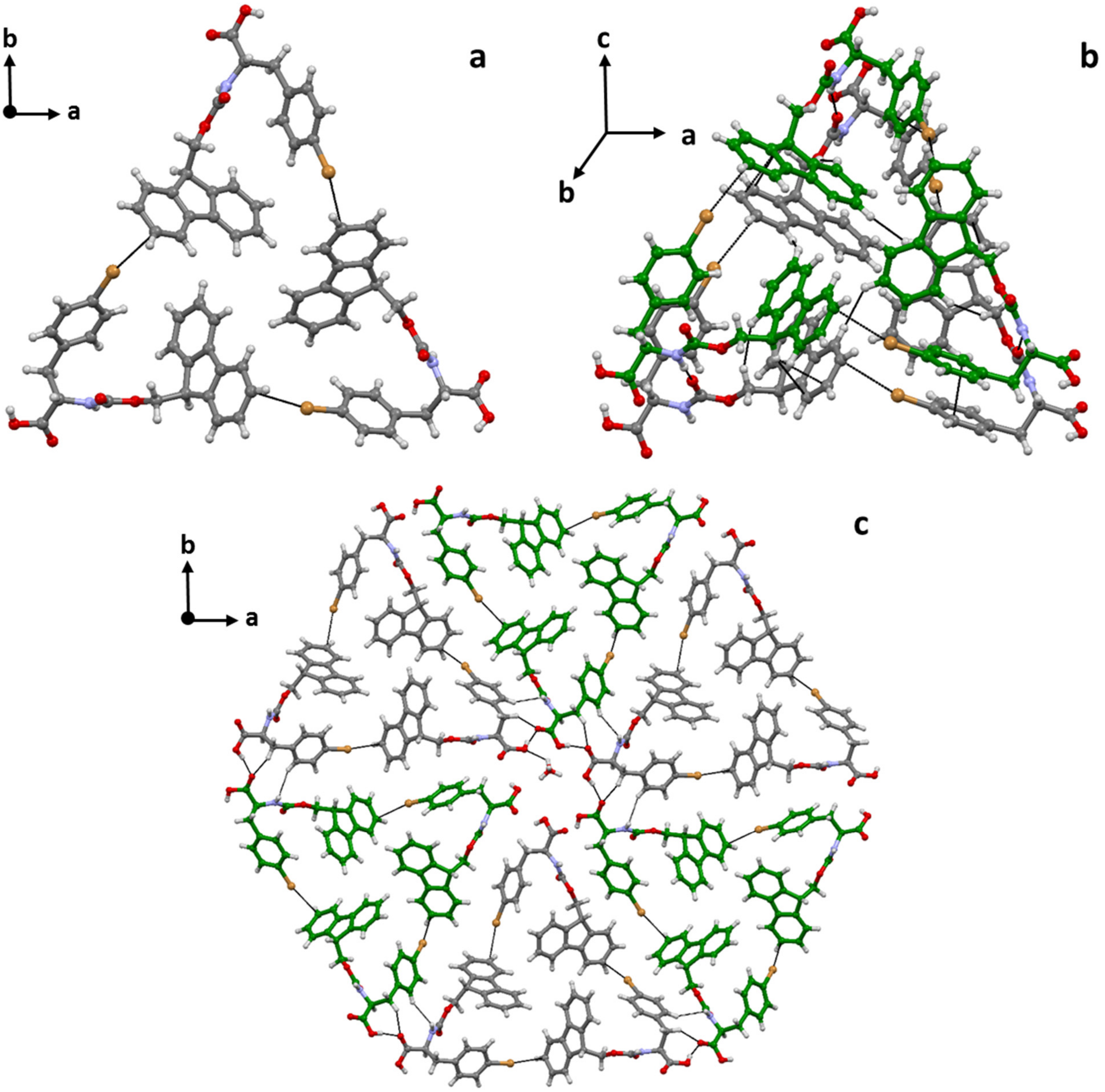


Figure 12: Crystal structure of Fmoc-4-Br-Phe. Halogen···π contacts lead to formation of three-membered units, view along *c* crystallographic axis a). Lateral view of stacked repetition units, with hydrogen bonds stabilizing the resulting supramolecular structure b) Hexamer of three-membered repetition units, forming a channel in the centre. This channel runs along the *c* crystallographic axis and hosts water molecules c).

Fmoc-4Br-Phe crystallizes in the *P* 6_3 space group. The asymmetric unit consists of a single molecule of amino acid solvated by two water molecules. Bromine interacts with the Fmoc moiety of a neighboring amino acid molecule (Br···π distance 3.371 Å), resulting in a closed repetition unit formed by three molecules of **Fmoc-4Br-Phe** (Fig 12a). This unit develops along the *c* crystallographic axis through N–H···O hydrogen bonds between the carbonyl group of the Fmoc moiety and the N–H group of an adjacent amino acid molecule lying on another plane (Fig 12b). Along this direction, the crystal packing is further stabilized by intermolecular short contacts H···H, π···π (among phenyl rings) and C–H···π (among Fmoc moieties). The crystal structure grows along both the crystallographic axis *a* and *b* via hydrogen bonding between the C-terminal groups of neighboring molecules, belonging to independent three-membered units. Lateral association of these repetition units leads to hexamers arranged in such a way to form channels running along the crystallographic axis *c*. (Fig 12c). These channels host water molecules, which further stabilize this arrangement by hydrogen bonding with the C-terminal groups of the amino acids.

2.4.2.4 Crystal structure of Fmoc-4-I-Phe

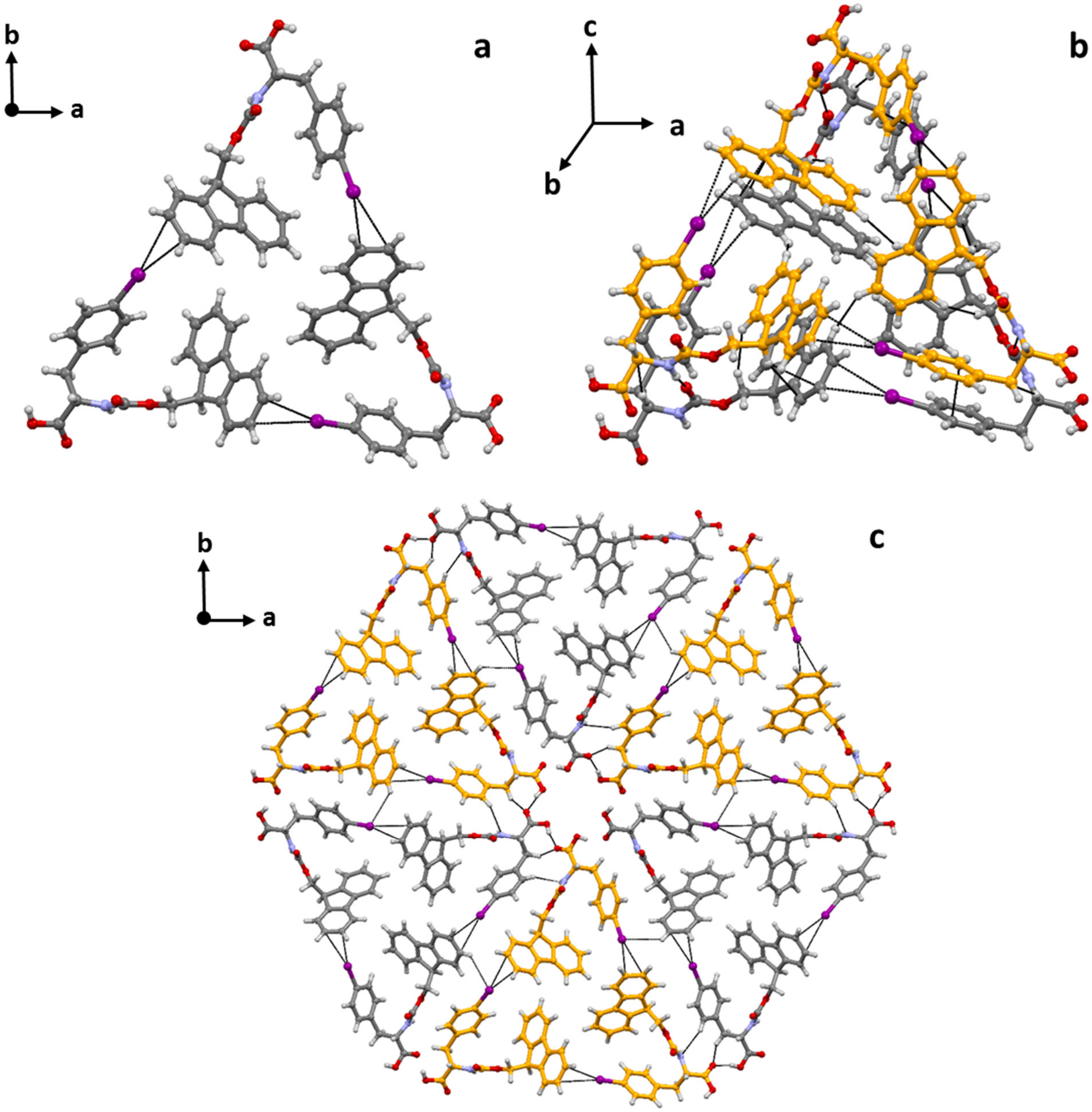


Figure 13: Crystal structure of Fmoc-4-I-Phe. Halogen···π contacts lead to formation of three-membered units, view along *c* crystallographic axis a). Lateral view of stacked repetition units, with hydrogen bonds stabilizing the resulting supramolecular structure b). Hexamer of three-membered repetition units, forming a channel in the centre. This channel runs along the *c* crystallographic axis and hosts water molecules c).

Fmoc-4I-Phe crystallizes in the *P* 6_3 space group. The asymmetric unit consists of a single molecule of amino acid solvated by two water molecules. Iodine interacts with the Fmoc moiety of a neighboring amino acid molecule (I···π distance 3.392 Å), resulting in a closed repetition unit formed by three molecules of **Fmoc-4I-Phe** (Fig 13a). This unit develops along the *c* crystallographic axis through N–H···O hydrogen bonds between the carbonyl group of the Fmoc moiety and the N–H group of an adjacent amino acid molecule lying on another plane (Fig. 13b). Along this direction, the crystal packing is further stabilized by intermolecular short contacts H···H, π···π (among phenyl rings) and C–H···π (among Fmoc moieties). The crystal structure grows along both the crystallographic axis *a* and *b* via hydrogen bonding between the C-terminal groups of neighboring molecules, belonging to independent three-membered units. Lateral association of these repetition units leads to hexamers arranged in such a way to form channels running along the crystallographic axis *c*. (Fig 13c). These channels host water molecules, which further stabilize this arrangement by hydrogen bonding with the C-terminal groups of the amino acids.

2.4.3 CCDC Analysis

The consistency of the halogen-π contacts noticed in **Fmoc-4-I-Phe** and **Fmoc-4-Br-Phe** found confirmation through a survey in the CCDC database (Table 3, Table 4). The geometrical parameters of these contacts – typical of halogen bonds – are within the range of similar halogen-π interactions found in the database (Fig. 14, Fig. 15). Considering the frequent occurrence of halogen bond in several crystallization conditions, it cannot be overlooked as key interaction for hydrogel formation in the Fmoc-Phe system. Therefore, the crystal structures of the **Fmoc-4-X-Phe** series can represent a reliable model explaining the reason for growing gel strength, growing thermal stability and higher fibrillogenic propensity by increasing halogen atom polarizability.

Table 3a: Geometrical parameters set-up used to find in the CCDC database I···π interactions similar to the one found in Fmoc-4-I-Phe.

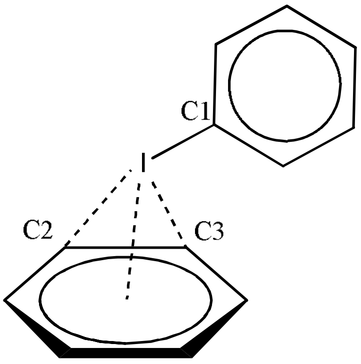
	**Parameter**	**Range**
	Distance I (I···Centroid)	0 – 4.68 Å
	Distance II (I···C_2)	0 – 3.68 Å
	Distance III (I···C_3)	0 – 3.68 Å
	Angle I (C_1– I – C_2)	0 – 180°
	Angle II (C_1– I – C_3)	0 – 180°
	Angle III (C_1– I – Centroid)	90 – 180°

Table 3b: Average values of the geometrical parameters of I···π interactions found in the CCDC database (299 hits) and comparison with Fmoc-4-I-Phe.

Parameter	Average	Fmoc-4-I-Phe
Distance I (I···Centroid)	3.69 Å	3.83 Å
Distance II (I···C_2)	3.54 Å	3.59 Å
Distance III (I···C_3)	3.54 Å	3.34 Å
Angle I (C_1–I–C_2)	157.05°	169.87°
Angle II (C_1–I–C_3)	158.60°	166.39°
Angle III (C_1–I–Centroid)	156.35°	165.23°

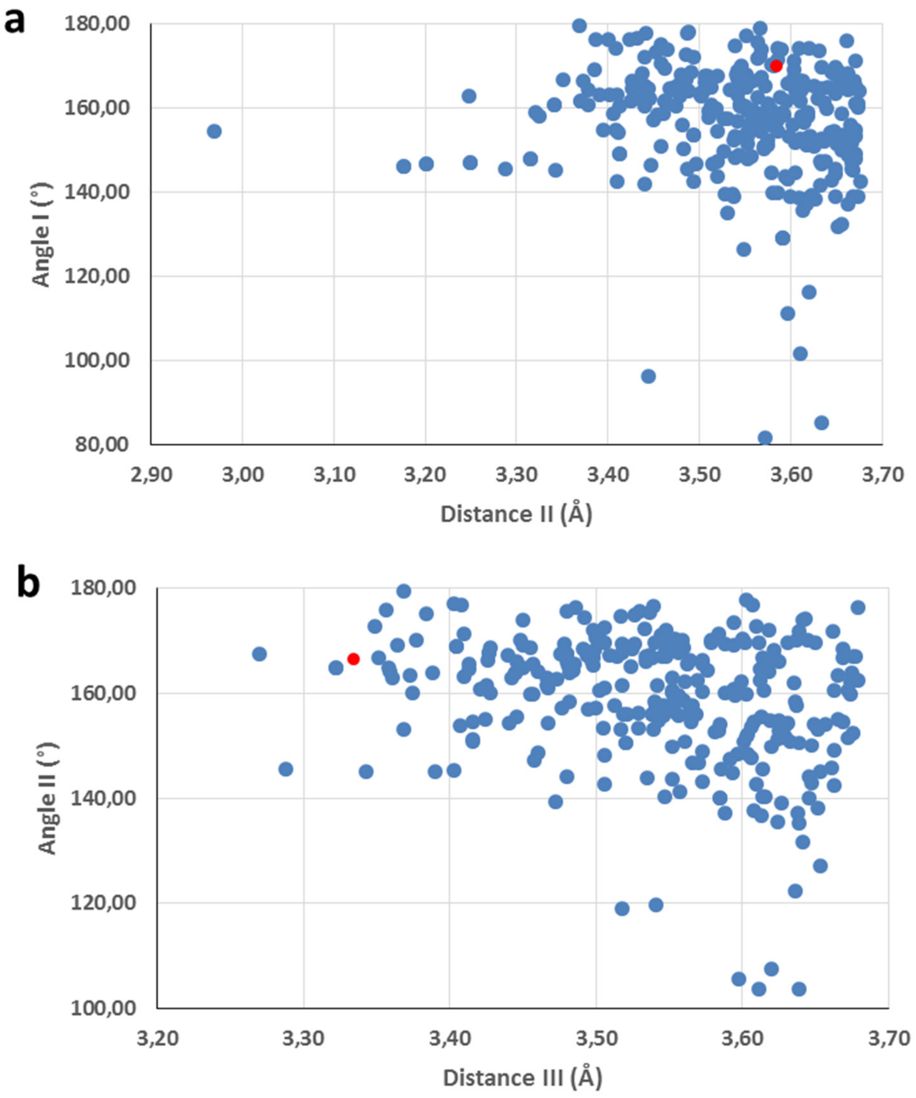


Figure 14: Distance II plotted as a function of its corresponding angle a). Distance III plotted as a function of its corresponding angle b). The red spot is Fmoc-4-I-Phe.

Table 4a: Geometrical parameters set-up used to find in the CCDC database Br···π interactions similar to the one found in Fmoc-4-Br-Phe.

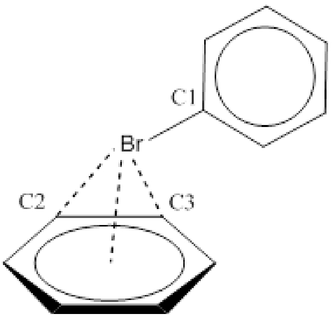
	Parameter	**Range**
	Distance I (Br···Centroid)	0 – 4.53 Å
	Distance II (Br···C_2)	0 – 3.53 Å
	Distance III (Br···C_3)	0 – 3.53 Å
	Angle I (C_1 – Br– C_2)	0 – 180°
	Angle II (C_1 –Br– C_3)	0 – 180°
	Angle III (C_1 –Br– Centroid)	90 – 180°

Table 4b: Average values of the geometrical parameters of Br···π interactions found in the CCDC database (609 hits) and comparison with Fmoc-4-Br-Phe. Distance III and angle II are not reported for Fmoc-4-Br-Phe since in the crystal structure bromine shows a contact with a unique carbon atom of the Fmoc moiety.

Parameter	**Average**	**Fmoc-4-Br-Phe**
Distance I (Br···Centroid)	3.64 Å	3.76 Å
Distance II (Br···C_2)	3.43 Å	3.26 Å
Distance III (Br···C_3)	3.42 Å	/
Angle I (C_1 –Br– C_2)	155.85°	167.95°
Angle II (C_1– Br– C_3)	157.69°	/
Angle III (C_1– Br– Centroid)	154.93°	164.19°

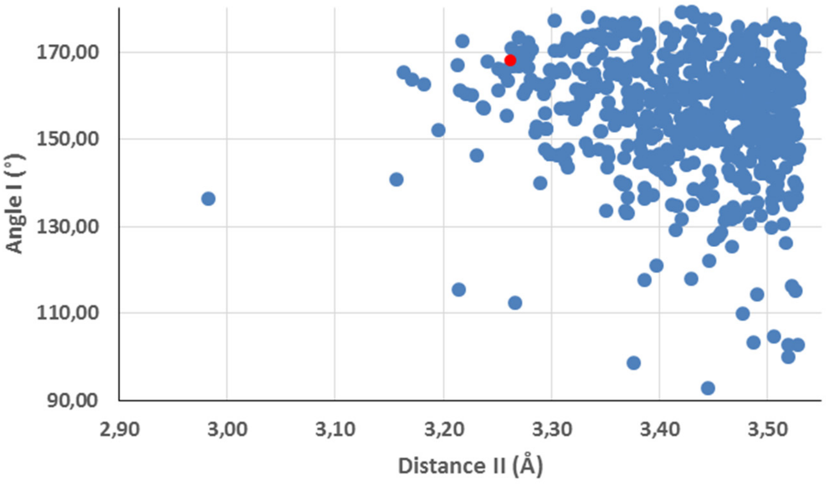


Figure 15. Distance II plotted as a function of its corresponding angle. The red spot is Fmoc-4-Br-Phe.

2.4.4 SAXS experiments

Small Angle X-ray scattering measurements on the 5 mM **Fmoc-4-X-Phe** hydrogels (5% DMSO/PBS) are consistent with an increasing complexity of the structures constituting the gels with halogen atom polarizability. While the non-modified amino acid does not show a well-defined form factor (Fig. 17), the halogenated compounds present higher level of order, which increases in parallel with the self-assembly efficiency of the amino acids. Indeed, a more complex scattering signal is associated to the compounds forming the strongest hydrogels (see rheology and thermal stability results). The iodinated and the brominated amino acids show very similar SAXS signatures, corroborating the results of single crystal XRD, with the two compounds that are almost isostructural. In particular, the scattering associated to **Fmoc-4-I-Phe** hydrogel (Fig 16a) results in a series of broad peaks, located in the region of high scattering vector values (q >1). The scattering of **Fmoc-4-Br-Phe** leads to similar signals, although less evident (Fig 16b). This finding is consistent with the hydrogel network formed by the brominated amino acid, less packed and organized compared to that of the iodinated compound. Interesting, the SAXS spectra of **Fmoc-4-Cl-Phe** and **Fmoc-4-F-Phe** show a Bragg peak at 2.7 nm^{-1} corresponding to a repetitive feature of about 2.5 nm. This feature might be related to an arrangement of the amino acid in the fibril cross-section. This is also observed in the SAXS spectrum of the bromine amino acid, although less pronounced. In fact, for brominated and iodinated derivatives, the formation of longer fibrils – that entangle forming a complex network – results in a form factor probably dominating the signal, thus covering the Bragg Peak.

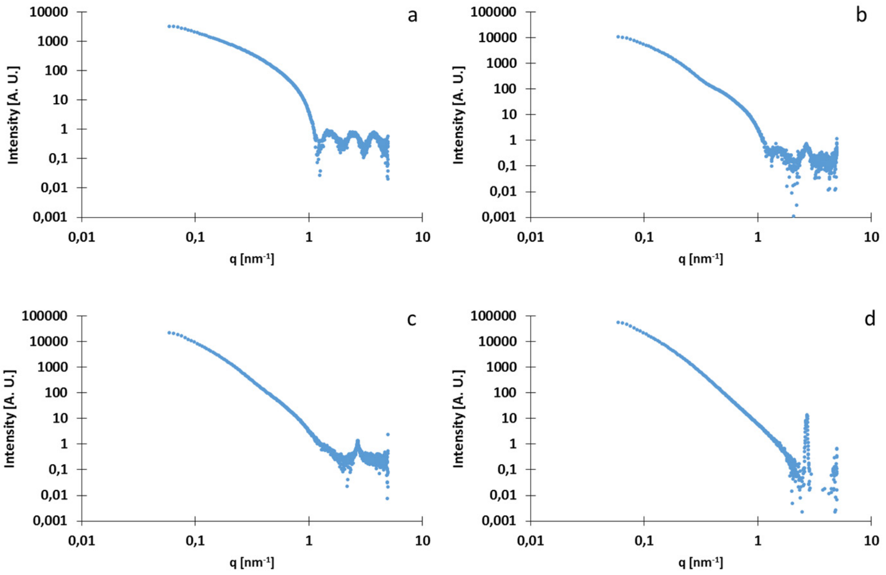


Figure 16. SAXS measurements of 5 mM hydrogels formed by: Fmoc-4-I-Phe a) Fmoc-4-Br-Phe b) Fmoc-4-Cl-Phe c) Fmoc-4-F-Phe d).

Although a detailed study focused on the fitting of the SAXS data is still necessary to get a full structural description of the self-assemblies constituting the hydrogels, it is anyway evident that there is a connection among increasing halogen atom polarizability, structural complexity and periodicity.

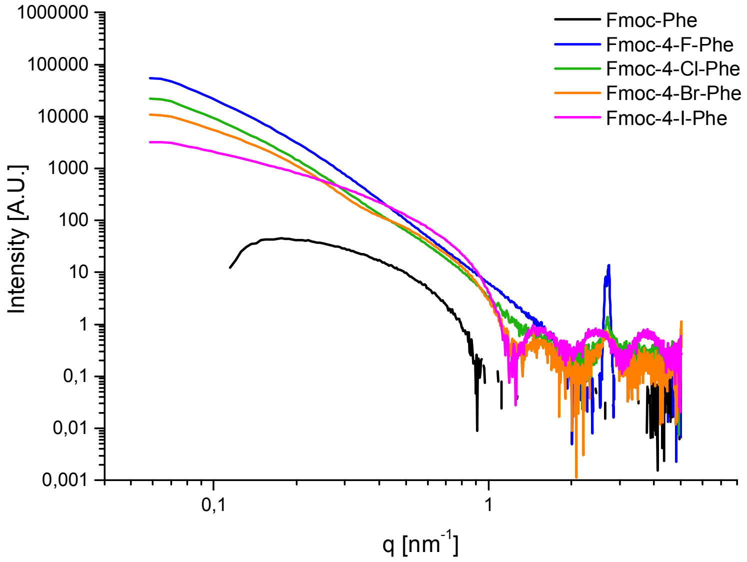


Figure 17. Plot of the SAXS intensity profiles of Fmoc-4-X-Phe 5 mM hydrogels. The signal associated to the gel formed by the non-modified amino acid is plotted for a direct comparison with the halogenated compounds.

2.5 Conclusions

Halogenation is a minimal structural modification; however, due to the large set of non-covalent interactions given by halogen atoms, it deeply changes the properties of a molecule. In this work, we demonstrated with several experimental techniques that the capability of highly polarizable halogen atoms to act as electrophilic species improves the fibrillation properties of a well-known system like Fmoc-Phe. The crystallographic study on the **Fmoc-4-X-Phe** series gives an enlightening molecular insight of the non-covalent interactions driving the self-assembly process of these compounds. Considering the overall interaction pattern, halogen bond results decisive in conferring enhanced efficiency to the supramolecular arrangement and higher stability to the resulting assemblies reinforcing their hydrogen bonds/π-stacking based backbone. For this reason, halogen bonding can be considered a powerful supramolecular tool enriching the landscape of biomolecular and hydrogelator self-assembly.

2.6 References

(1) Abdallah, D. J.; Weiss, R. G. *Adv. Mater.* **2000**, *12* (17), 1237.

(2) Weiss, M. G., Weiss, R. G. *Acc. Chem. Res.* **2006**, *39* (8).

(3) Raeburn, J.; Adams, D. J. *Chem. Commun.* **2015**, *51* (25), 5170.

(4) Terech, P.; Weiss, R. G. *Chem. Rev.* **1997**, *97* (8), 3133.

(5) Duan, P.; Yanai, N.; Nagatomi, H.; Kimizuka, N. *J. Am. Chem. Soc.* **2015**, *137* (5), 1887.

(6) Praveen, V. K.; Ranjith, C.; Bandini, E.; Ajayaghosh, A.; Armaroli, N. *Chem. Soc. Rev.* **2014**, 4222.

(7) Tao, L.; Huo, Z.; Dai, S.; Ding, Y.; Zhu, J.; Zhang, C.; Zhang, B.; Yao, J.; Nazeeruddin, M. K.; Gra, M. *J. Phys. Chem. C* **2014**, *118*, 16718.

(8) Hemamalini, A.; Mohan Das, T. *New J. Chem.* **2013**, *37* (8), 2419.

(9) Schön, E. M.; Marqués-Lõpez, E.; Herrera, R. P.; Alemán, C.; Díaz, D. D. *Chem. Eur. J.* **2014**, *20* (34), 10720.

(10) Oh, H.; Yaraghi, N.; Raghavan, S. R. *Langmuir* **2015**, *31* (19), 5259.

(11) Foster, J. A.; Piepenbrock, M. M.; Lloyd, G. O.; Clarke, N.; Howard, J. A. K.; Steed, J. W. *Nat. Chem.* **2010**, *2* (12), 1037.

(12) Nanda, J.; Biswas, A.; Banerjee, A. *Soft Matter* **2013**, *9* (16), 4198.

(13) Yang, C.; Chu, L.; Zhang, Y.; Shi, Y.; Liu, J.; Liu, Q.; Fan, S.; Yang, Z.; Ding, D.; Kong, D.; Liu, J. *ACS Appl. Mater. Interfaces* **2015**, *7* (4), 2735.

(14) Tomasini, C.; Castellucci, N. *Chem. Soc. Rev.* **2013**, *42* (1), 156.

(15) Boekhoven, J.; Stupp, S. I. *Adv. Mater.* **2014**, *26* (11), 1642.

(16) Knipe, J. M.; Peppas, N. A. *Regen. Biomater.* **2014**, *1* (1), 57.

(17) Dankers, P. Y. W.; Beek, D. J. M. Van; Cate, a T.; Rint, P.; Meijer, E. W. *Engineering* **2003**, *24*, 52.

(18) Nagai, Y.; Yokoi, H.; Kaihara, K.; Naruse, K. *Biomaterials* **2012**, *33* (4), 1044.

(19) Zhang, J.; Ou, C.; Shi, Y.; Wang, L.; Chen, M.; Yang, Z. *Chem. Commun.* **2014**, *50*, 12873.

(20) Ikeda, M.; Tanida, T.; Yoshii, T.; Kurotani, K.; Onogi, S.; Urayama, K.; Hamachi, I. *Nat. Chem.* **2014**, *6* (6), 511.

(21) Banerjee, A.; Basu, K.; Baral, A.; Basak, S.; Dehsorkhi, A.; Nanda, J.; Bhunia, D.; Ghosh, S.; Castelletto, V.; Hamley, I. *Chem. Commun.* **2016**, *1* (LII), 5.

(22) Wang, H.; Mao, D.; Wang, Y.; Wang, K.; Yi, X.; Kong, D.; Yang, Z.; Liu, Q.; Ding, D. *Sci. Rep.* **2015**, *5* (October), 16680.

(23) Martin, A. D.; Robinson, A. B.; Mason, A. F.; Wojciechowski, J. P.; Thordarson, P. *Chem. Commun.* **2014**, *50* (98), 15541.

(24) Martin, A. D.; Wojciechowski, J. P.; Warren, H.; In Het Panhuis, M.; Thordarson, P. *Soft Matter* **2016**, *12*, 2700.

(25) Draper, E. R.; Morris, K. L.; Little, M. A.; Raeburn, J.; Colquhoun, C.; Cross, E. R.; McDonald, T. O.; Serpell, L. C.; Adams, D. J. *CrystEngComm* **2015**, *17* (42), 8047.

(26) Liyanage, W.; Nilsson, B. L. *Langmuir* **2016**, *32* (3), 787.

(27) Ryan, D. M.; Doran, T. M.; Anderson, S. B.; Nilsson, B. L. *Langmuir* **2011**, *27* (7), 4029.

(28) Pappas, C. G.; Abul-Haija, Y. M.; Flack, A.; Frederix, P. W. J. M.; Ulijn, R. V. *Chem. Commun.* **2014**, *50* (73), 10630.

(29) Ryan, D. M.; Doran, T. M.; Nilsson, B. L. *Langmuir* **2011**, *27* (17), 11145.

(30) Ryan, D. M.; Anderson, S. B.; Nilsson, B. L. *Soft Matter* **2010**, *6* (14), 3220.

(31) Wang, Y.; Zhang, Z.; Xu, L.; Li, X.; Chen, H. *Colloids Surfaces B Biointerfaces* **2013**, *104*, 163.

(32) Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimägi, A.; Resnati, G.; Terraneo, G. *Chem. Rev.* **2016**, *116* (4), 2478.

(33) Gilday, L. C.; Robinson, S. W.; Barendt, T. A.; Langton, M. J.; Mullaney, B. R.; Beer, P. D. *Chem. Rev.* **2015**, *115* (15), 7118.

(34) Lu, Y.; Wang, Y.; Zhu, W. *Phys. Chem. Chem. Phys.* **2010**, *12* (18), 4543.

(35) S. R. Raghavan, B. H. C. *Molecular Gels, Materials with Self-Assembled Fibrillar Networks*; R. G. Weiss, P. T., Ed.; Springer, Dordrecht, The Netherlands, 2006.

(36) Lausi, A.; Polentarutti, M.; Onesti, S.; Plaisier, J. R.; Busetto, E.; Bais, G.; Barba, L.; Cassetta, A.; Campi, G.; Lamba, D.; Pifferi, A.; Mande, S. C.; Sarma, D. D.; Sharma, S. M.; Paolucci, G. *Eur. Phys. J. Plus* **2015**, *130* (3), 43.

(37) Kabsch, W. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (2), 125.

(38) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 235.

(39) Evans, P. R.; Murshudov, G. N. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (7), 1204.

(40) Zwart, P. H.; Banumathi, S.; Dauter, M.; Dauter, Z. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (11), 1958.

(41) McGeehan, J. E.; Carpentier, P.; Royant, A.; Bourgeois, D.; Ravelli, R. B. G. *J. Synchrotron Radiat.* **2007**, *14* (1), 99.

(42) Sheldrick, G. M. *Acta Crystallogr. Sect. A Found. Crystallogr.* **2015**, *71* (1), 3.

(43) Sheldrick, G. M. *Acta Crystallogr. Sect. C Struct. Chem.* **2015**, *71* (Md), 3.

(44) Emsley, P.; Cowtan, K. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12 I), 2126.

(45) Parsons, S.; Flack, H. D.; Wagner, T. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2013**, *69* (3), 249.

(46) Lee, C.-C.; Nayak, A.; Sethuraman, A.; Belfort, G.; McRae, G. J. *Biophys. J.* **2007**, *92* (10), 3448.

(47) Giorgetti, E.; Marsili, P.; Cicchi, S.; Lascialfari, L.; Albiani, M.; Severi, M.; Caporali, S.; Muniz-Miranda, M.; Pistone, A.; Giammanco, F. *J. Colloid Interface Sci.* **2015**, *442*, 89.

(48) Makarević, J.; Jokić, M.; Raza, Z.; Štefanić, Z.; Kojić-Prodić, B.; Žinić, M. *Chem. Eur. J.* **2003**, *9* (22), 5567.

(49) Martin, A. D.; Wojciechowski, J. P.; Bhadbhade, M. M.; Thordarson, P. *Langmuir* **2016**, *32* (9), 2245.

(50) Cicchi, S.; Ghini, G.; Lascialfari, L.; Brandi, A.; Betti, F.; Berti, D.; Ferrati, S.; Baglioni, P. *Chem. Commun.* **2007**, No. 14, 1424.

(51) Meazza, L.; Foster, J. A.; Fucke, K.; Metrangolo, P.; Resnati, G.; Steed, J. W. *Nat Chem* **2013**, *5* (1), 42.

Chapter 3

Halogenation as a powerful strategy to tune the self-assembly of an amyloid peptide fragment

3.1 Introduction

Amyloid fibrils are striking examples of self-assembly processes. They result from soluble peptides or proteins present in the human body that assemble into insoluble nanostructures, which are associated to pathologies like Alzheimer's, Parkinson's, Creutzfeldt-Jacob and Huntington's diseases.[1] The presence of these fibrillar aggregates is also significant in nature, covering a wide range of functional roles, such as being a key component of spider silk.[2] Because of their impressive mechanical properties and chemical stability, research in the development of new biomaterials takes inspiration from the self-assembly behavior of amyloid fibrils.[3] In order to rationalize the architecture of new amyloid-based materials, a fundamental requirement is the possibility to tune the self-assembly process. A recent work from our group exploits partial iodination of an amino acidic residue as a new strategy to promote fibril formation of the amyloidogenic peptide fragment DFNKF.[4] In particular, iodination of phenylalanine residues strongly amplifies the fibrillation propensity of the modified peptide compared to the wild type sequence. In this context, to demonstrate the generality of this finding, we decided to apply the same strategy to another known peptide fragment – KLVFF – (H_2N-Lys-Leu-Val-Phe-Phe-COOH)[5–8] that is part of the peptide Aβ40, whose spontaneous assembly into amyloid fibrils is associated to Alzheimer's disease.[9] Since many experimental results and modelling studies[10–12] suggested that this fragment (Aβ $_{(16\text{-}20)}$) is critical for the fibrillation of the full-length Aβ peptide, we selected this segment for our investigation. The peculiar cross-β-sheet structure of amyloid fibrils[13,14] is stabilized by several non-covalent interactions, such as hydrogen bonding, π-π stacking and hydrophobic interactions.[15] In this regard, Gazit and coworkers demonstrated both theoretically and experimentally that π-π interactions involving aromatic residues have a central role in self-assembly processes that lead to amyloid formation.[16] In particular, the diphenylalanine motif – FF – was deeply studied, showing that this dipeptide forms well ordered, tubular, and elongated assemblies.[17,18] The importance of aromatic interactions in peptide assembly can be inferred also from a recent paper by Frederix *et al.,*[19] where 8000 possible combinations of tripeptides were computationally screened for the aqueous self-assembly propensity. This screening showed that all the top aggregating tripeptides contain a pair of adjacent aromatic amino acids and at least one phenylalanine residue. In particular, the sequence VFF turns out to be one of the best in the aggregation propensity (AP) score. This result, combined with the study on the iodinated DFNKF,[4] hints on the effects of halogenation on larger sequences containing the FF motif, such as KLVFF. Although several modifications of amyloidogenic sequences have been investigated to tune their self-assembly properties,[20,21] halogenation has not yet been widely explored.

3.2 Materials

Congo red, Rhodamine B, NaCl, Phosphate buffered saline and D_2O were purchased from Sigma-Aldrich and used without further purification. Peptides with confirmed amino acid analysis (purity ≥ 98%), were purchased from Biopeptek (Malvern, USA). The integrity of all peptides was confirmed by ion spray mass spectrometry and the purity was determined by reverse phase high-pressure liquid chromatography (HPLC).

3.3 Methods

3.3.1 Peptide solutions and Gelation experiments. Peptide solutions were freshly prepared in deionized water (18.2 MΩ·cm), sonicated for 20 seconds, and gently warmed to reach 90 °C. The peptide solutions were stored in sealed vials at r.t. for varying time points before analysis.

Hydrogels were prepared by dissolving peptide powder (15 mM for halogenated peptides and 64 mM for the wild-type peptide KLVFF) in deionized water (18.2 MΩ·cm) or in D_2O (≥ 99.9 atom % deuterium). The glass vials containing the 500 µL solutions were sealed, sonicated for 20 seconds, heated using a heat gun until complete dissolution of the peptide powder, and eventually left cooling at r.t. All samples were stored at r.t. for 48 hours before analysis.

3.3.2 Rheology. Rheology experiments were performed using a TA instrument AR-G2 Rheometer. A 20 mm stainless steel, parallel-plate geometry was used with a gap distance of 1000 µm. Oscillatory frequency sweep studies were performed for a range of 0.1-100 rad/s, using a 0.5% strain. Oscillatory amplitude sweep studies were conducted from 0.01 to 100% strain with an angular frequency of 1 rad/s. The ring cast method was used for hydrogel preparation at a peptide concentration of 15 mM. The peptide solution was sonicated for 20 s in a sealed glass vial before heating to 90 °C to afford complete dissolution of the peptide. After cooling the solutions were transferred into ring-casts of 22 mm diameter and placed in tightly sealed tissue-culture dishes for 48 hours. All measurements were repeated a minimum of three times.

3.3.3 Circular Dichroism (CD) Spectroscopy. All the circular dichroism experiments were carried out in deionized water (18.2 MΩ·cm) in a 0.1 cm quartz cuvette, using a JASCO J-815 CD spectrometer. Acquisitions were performed between 190 and 250 nm with a 0.1 nm step size, 1 nm bandwidth, 100 nm min^{-1} scanning speed and 1 s response time. All the spectra are an average of 10 scans and were corrected from a reference solution, comprised of deionized water (18.2MΩ.cm) alone. Raw data (θ, in mdeg) were subsequently converted to mean residue ellipticity ([θ] in $deg.cm^2.dmol^{-1}$) for the sake of comparison, in accordance with the following formulae:[4]

$$[\theta] = \frac{\theta}{10 * l * c * (n-1)}$$

where θ is the observed ellipticity in mdeg, c is the concentration of the sample in $mol.L^{-1}$, (n-1) is the number of peptide bonds, and l is the pathlength of the cuvette in cm.

3.3.4 Confocal Microscopy. Hydrogels were imaged using a Zeiss LSM 710 microscope with a He/Ne laser (λ_{ex}= 543 nm). The fluorescent dye, Rhodamine B, was incorporated into an aged hydrogel (48 h) scaffold by addition of 10 μl of the dye solution (0.1% w/v). Following complete absorption of the dye, the sample was excited at 543 nm and emitted light recorded using the E570LP emission filter.

3.3.5 Congo Red Staining. All samples were monitored for green birefringence using an Olympus BX50 polarizing microscope with a SensiCam PCO camera used to display and enhance images. An 80% ethanol solution saturated with NaCl and Congo Red was freshly prepared before each measurement. A piece of each peptide hydrogel was placed on a glass microscope slide, allowed to air dry and then stained with Congo Red solution. Subsequently, excess Congo Red solution was blotted off the slide and the samples were analyzed using both bright and polarized light.

3.3.6 Infrared Spectroscopy (FT-IR). Infrared spectra were recorded at room temperature using a Nicolet iS50 FT-IR spectrometer equipped with a DTGS detector. Peptides were analyzed as solutions (after heating at 100 °C in order to break any pre-formed fibrils) or gels at 15 mM in D_2O. Spectra represent an average of 64 scans recorded in single beam mode with a 4 cm^{-1} resolution and corrected for the background. The second derivative analyses of the spectra were performed using the Nicolet FTIR software, Omnic 9.0®, with a 13-point and 3rd polynomial order Savitzky-Golay function.

Second derivative spectra generated negative bands as compared with the original spectra, thus for comparison all the second-derivative spectra were multiplied by -1.

3.3.7 Transmission Electron Microscopy (TEM). TEM bright field images were acquired using a Philips CM200 electron microscope operating at 200 kV equipped with a Field Emission Gun filament. A Gatan US 1000 CCD camera was used and 2048x2048 pixel images with 256 grey levels were recorded. The suspension were dropped onto a 200 mesh carbon-coated copper grid and air dried for several hours before analysis.

3.3.8 Cryogenic Transmission Electron Microscopy (cryo-TEM). The cryo-TEM images were collected using JEM 3200FSC field emission microscope (JEOL) operated at 300 kV in bright field mode with Omega-type Zero-loss energy filter. The images were acquired with Gatan digital micrograph software while the specimen temperature was maintained at -187 ^{o}C. The Cryo-TEM samples were prepared by placing 3 µL aqueous dispersion of nanoparticles/clusters on a 200 mesh copper grid with holey carbon support film (CF-Quantifoil) and plunge freezed using vitrobot with 2s blotting time under 100% humidity.

3.3.9 Single crystal X-ray diffraction (XRD). H_2N-Lys-Leu-Val-(*p*-IodoPhe)-(*p*-IodoPhe)-COOH *i.e.* **KLVF(I)F(I)**, H_2N-Lys-Leu-Val-(*p*-BromoPhe)-(*p*-BromoPhe)-COOH, *i.e.* **KLVF(Br)F(Br)** and H_2N-Lys-Leu-Val-(*p*-ChloroPhe)-(*p*-ChloroPhe)-COOH, *i.e.* **KLVF(Cl)F(Cl)** were obtained as solvated species by dissolving the peptide in a water/hexafluoro-2-propanol 90:10 mixture. Crystals suitable for XRD analysis were obtained after two months of slow evaporation. Data collections were performed at the X-ray diffraction beamline (XRD1) of the Elettra Synchrotron, Trieste (Italy).[22] The crystals were dipped in perfluoropolyether vacuum oil (Fomblin) and mounted on the goniometer head with a nylon loop. Complete datasets were collected at 100 K (nitrogen stream supplied through an Oxford Cryostream 700) through the rotating crystal method. Data were acquired using a monochromatic wavelength of 0.850 Å for **KLVF(I)F(I)** and 0.700 Å for **KLVF(Br)F(Br)** and **KLVF(Cl)F(Cl)** on a Pilatus 2M hybrid-pixel area detector. The diffraction data were indexed and integrated using XDS.[23] Scaling have been done using CCP4-Aimless code.[24,25] Crystals appear as very thin needles prone to radiation damage, as previously reported for other halogenated molecules.[26,27] For the brominated peptide we managed to collect a complete dataset from a unique crystal; for the iodine and chlorine derivatives, diffraction decayed even after small doses so four different datasets had to be merged for **KLVF(I)F(I)** and three datasets for **KLVF(Cl)F(Cl)** (collected from different crystals randomly oriented). Semi-empirical absorption correction and scaling was performed for the **KLVF(Br)F(Br)** dataset, exploiting multiple measures of symmetry-related reflections, using SADABS program.[28] The structures were solved by the dual space algorithm implemented in the SHELXT code.[29] Fourier analysis and refinement were performed by the full-matrix least-squares methods based on F^2 implemented in SHELXL-2014.[30] The Coot program was used for modeling.[31]

KLVF(Cl)F(Cl) peptide crystallized in a monoclinic unit cell ($P\,2_1$ space group). The model has been fully refined anisotropic as a 2-component non-merohedral twin. Crystal showed two domains related by a 180° rotation around the *c** reciprocal lattice direction (twin fraction refined to 17%). One peptide and four water molecules have been found in the asymmetric unit.

KLVF(Br)F(Br) and **KLVF(I)F(I)** peptides crystallized in equivalent conditions and showed the same $P\,2_12_12_1$ orthorhombic crystalline form. The cell volume is slightly bigger for the iodinated peptide, as expected from comparison of the halogen atomic radius. None of the crystals tested diffracted better than 1.1 Å, and considering radiation damage, the overall dataset resolution is not better than ~1.25 Å, for both the compounds. The number of data for model fitting was therefore limited and, to avoid over-refinement, anisotropic thermal motion modeling has been applied only to halogen atoms of the peptide (the heaviest atoms in the structures). Geometric restraints on bond lengths and angles (DFIX, DANG) have been used for all the residues and thermal motion parameters restrains (SIMU) have been applied on disordered and poorly defined fragments. Hydrogen atoms were included at calculated positions with isotropic $U_{factors} = 1.2\ U_{eq}$ or $U_{factors} = 1.5$ Ueq for methyl and hydroxyl groups (Ueq being the equivalent isotropic thermal factor of the bonded non-hydrogen atom). R_1(free)[32] values have been calculated for the brominated and iodinated models, omitting 5% of reflections (randomly selected) from refinement cycles. Reasonable agreement of final R_1(free) with R_1 values (Table 1S) exclude over-refinement issues, despite poor data/parameters ratios. A final refined Flack parameter 0.005(32)[33] for **KLVF(Br)F(Br)** confirms the reliability of the stereochemical configuration shown. The flack parameters for **KLVF(I)F(I)** and **KLVF(Cl)F(Cl)** are not reliable as a consequence of dataset merging in presence of significant radiation damage (Flack parameter 0.521(54) and 0.38(18)). Nevertheless, R_1 increases significantly inverting the structure (almost doubles) suggesting that the stereochemical configuration is the same as **KLVF(Br)F(Br)** (as expected from synthetic pathway). Pictures were prepared using Mercury[34] and Pymol software.[35] Essential crystal and refinement data (Table 1) are reported below.

3.3.10 Dynamic Light Scattering (DLS). Dynamic Light Scattering measurements were performed on an ALV/CGS-3 Platform-based Goniometer System equipped with an ALV-7004 correlator and an ALV / CGS-3 goniometer. The signal was detected by an ALV-Static and Dynamic Enhancer detection unit. The light source was the second harmonic of a diode-pumped Coherent Innova Nd:YAG laser (λ = 532 nm), linearly polarized in the vertical direction. Measurements were performed at 25 °C. Approximately 1 mL of sample solution was transferred into the cylindrical Hellma scattering cell.

The dynamic information on particles present in the peptide solutions were derived from the normalized autocorrelation function $g_2(q, \tau)$ of the scattered intensity, which is measured according to

$$g_2(q, \tau) = \frac{\langle I(q,t) I(q, t+\tau)\rangle}{\langle I(q,t)\rangle^2}$$

where q is the scattering vector, τ is the relaxation time and I is the scattered intensity. Data analysis has been performed with two different methods: in the first one, the autocorrelation functions have been analyzed through Laplace inversion (CONTIN algorithm),[36] which resulted in a double distribution of the decay rates associated to two distinct populations. In the second method, the same autocorrelation functions have been analyzed using a double exponential decay model, yielding two distinct decay rates. For each sample, at least three measurements were performed at four different angles (70°, 90°,110°,130°), corresponding to four different scattering vectors (q):

$$q = (4\pi n/\lambda)\, sin(\theta/2)$$

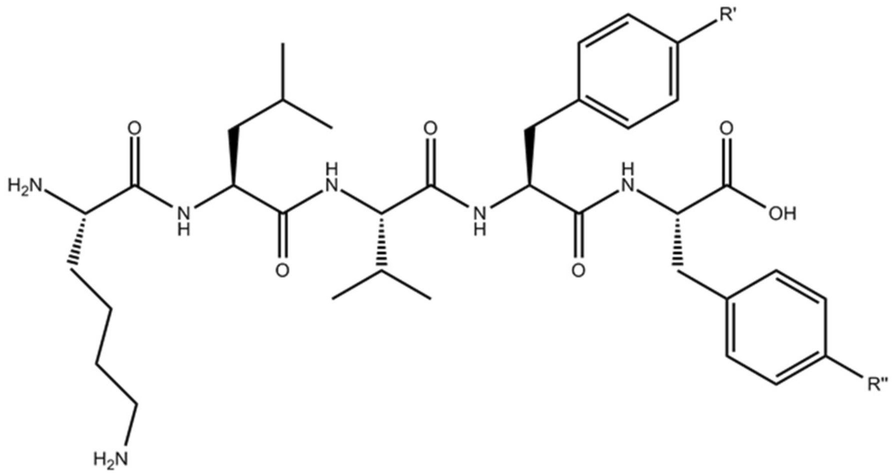
where n is the refractive index of the medium and θ is the scattering angle. The extracted decay rates were plotted versus the square scattering vectors (q^2) showing a linear dependence typical of Brownian motion. The slope of this curve represents the averaged translational diffusion coefficient (D_T), from which, through the Stokes-Einstein equation

$$R_H = k_B T/6\pi\eta D_T$$

(k_B is the Boltzmann constant, T is the temperature and η is the viscosity) it is possible to calculate the averaged hydrodynamic radius (R_H) associated to each population. This analysis, although neglecting polydispersity of the two populations, is a good control on the Laplace inversion method (CONTIN), which can produce artefacts if the autocorrelation functions are significantly noisy, with low amplitude to baseline ratio.

3.4 Results and discussion

Iodinated, brominated and chlorinated derivatives of KLVFF (Fig. 1) were prepared introducing halogen atoms on the *p*-position of the benzyl side chain of phenylalanine. The modified amino acids *p*-iodo phenylalanine (*p*-I-Phe), *p*-bromo phenylalanine (*p*-Br-Phe) and *p*-chloro phenylalanine (*p*-Cl-Phe) were introduced in the wild type sequence in place of natural phenylalanine (Phe). Since KLVFF contains two different halogenation targets (inner and terminal phenylalanine), we also studied the influence of halogenation on the peptide folding varying their position in the sequence.



	R'	R"
KLVFF	H	H
KLVF(I)F	I	H
KLVFF(I)	H	I
KLVF(I)F(I)	I	I
KLVF(Br)F	Br	H
KLVFF(Br)	H	Br
KLVF(Br)F(Br)	Br	Br
KLVF(Cl)F(Cl)	Cl	Cl

Figure 1: Chemical structure of KLVFF and its halogenated derivative.

Switching the halogen position on the different phenylalanine residues, we expect to generate isomers having different self-assembly behaviors, as recently reported by Cui *et al.*[37] by changing the position of two adjacent natural amino acids. Selective halogenation of the phenylalanine residues allowed us to investigate the folding behavior of different peptide isomers in new terms, *i.e.* using halogenation as a tool for generating isomerism. The main advantage of using a point mutation like halogenation is the minimal structural modification it implies. Despite this, it may induce striking changes in some properties compared to those of the wild-type analogue, due to the large set of non-covalent interactions given by halogen atoms.[38] Considering that the anisotropic properties of halogens strongly change along the periodic group, we expect to notice a different self-assembly behavior moving from chloro- to iodo- derivatives because of potentially different interaction patterns. Among the possible non-covalent interactions involving halogens also halogen bond (XB),[39] whose strength is related to the polarizability of the halogen atom,[40] was considered. Iodine and bromine, being more polarizable than chlorine, should have more propensity to act as XB donors than chlorine.

The different modified derivatives containing the halogenated amino acids were carefully studied both in the solid state and in solution. In order to get insights into the interaction pattern of halogen atoms in this amyloid fragment, we studied the crystal structures of the di-iodinated derivative **KLVF(I)F(I)**, the di-brominated derivative **KLVF(Br)F(Br)** and the di-chlorinated peptide **KLVF(Cl)F(Cl)**. Then we used these structures as a model to explain the different self-assembly properties of the halogenated KLVFF peptides in solution. We obtained also the crystal structures of **KLVFF(I)** and **KLVFF(Br)**, which features will be discussed in chapter 4. Indeed the latter, having a single halogen atom confined in peripheral position, represent good examples of the use of heavy atoms to help structure determination, and they could be reliable models for the native sequence KLVFF. These crystal structures do not show halogen atoms directly involved in specific interactions, unlike the ones that are presented below.

3.4.1 Crystal structures of KLVF(X)F(X) peptides

Amyloid-like fibrils, although arising by different amino acidic sequences, share a common structural peculiarity defined as 'cross-β' spine.[41] Because of their low crystallinity, the atomic structural determination of amyloid fibrils formed by full-length proteins is still a challenge. However, crystallographic studies performed on shorter fibril-forming segments can be considered a representative model of the complete sequences.[42–44] The spine consists of a pair of β-sheets running parallel to the fibril axis. Extensive hydrogen-bonded strands, perpendicular to the fibril axis, form the β-sheets. The facing side chains of pairing β-sheets form a self-complementary dry interface defined as 'steric zipper'.[45] This motif is common to all amyloid structures and defines the non-covalent interactions involving the interdigitated side chains, which in most cases are weak hydrophobic interactions. Each fibril-forming segment has a peculiar 'steric zipper' giving information that can elucidate the mechanism of aggregation of monomers into oligomers and fibrils. Here, we present the crystal structures of the halogenated derivatives **KLVF(I)F(I)**, **KLVF(Br)F(Br)** and **KLVF(Cl)F(Cl)** to show the impact of halogenation on the aggregation process of these peptide fragments. The use of iodination to amplify the self-assembly of the amyloidogenic sequence DFNKF[4,46] has been recently reported, demonstrating that halogenation affects the fibrillation propensity of the peptide. However, the role of halogenation on the final self-assembled structure has not been completely elucidated. The crystal structures of these derivatives will shed light on the halogen atom's interaction pattern.

3.4.1.1 Crystal structure of KLVF(I)F(I)

The crystal structure of **KLVF(I)F(I)** (experimental conditions reported in section 3.9) shows the amphoteric properties of iodine, behaving both as an electrophilic and a nucleophilic species. The asymmetric unit consists of a single molecule of **KLVF(I)F(I)** solvated by two water molecules and one molecule of hexafluoro-2-propanol (Figure 2a). The peptide monomers form extended hydrogen-bonded strands, resulting in parallel β-sheets. In detail, each **KLVF(I)F(I)** molecule forms eight backbone NH···O hydrogen bonds with the neighboring strands belonging to the same sheet (Fig 2b). The distance between each strand in the β-sheet is 4.91 Å. In addition to hydrogen bonds, also weak aromatic interactions among phenylalanine rings and C···H contacts involving the aliphatic residues contribute to the overall stability of the parallel β-sheets.

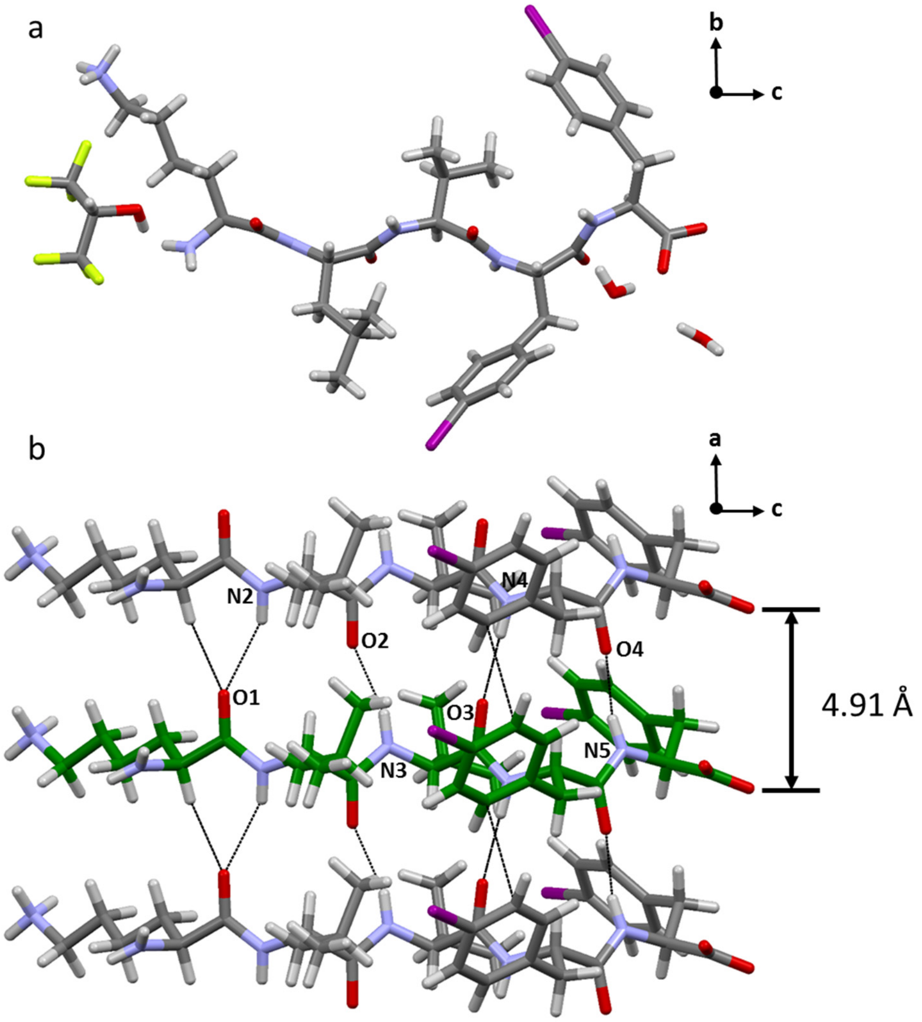


Figure 2: Crystal structure of **KLVF(I)F(I)**. Asymmetric unit consisting of a single peptide molecule solvated by two water molecules and one molecule of hexafluoro-2-propanol a). Hydrogen bonding contacts in a parallel β-sheet (Distances: N2H···O1 3.09 Å N3H···O2 2.98 Å N4H···O3 2.87 Å N5H···O4 2.84 Å) b). Color code: C, grey, green; O, red; N, violet; I, purple; F, yellow; H, white.

Paired β-sheets form the typical cross-β spine structure, with the facing side chains of the sheets interdigitated in a 'steric zipper', classified as FACE TO BACK – UP-DOWN – CLASS 4[42] (Fig 3a). This steric zipper, differently to those formed by wild type segments, is stabilized by intermolecular C=O···I contacts between strands belonging to adjacent β-sheets (Fig 3b). These inter-sheet interactions, involved in the mechanism of formation of the fibrils, are halogen bonds.[47,48] This is the first example showing the role of halogen bonding in stabilizing a steric zipper, instead of the common hydrophobic interactions. The iodine atom on the inner phenylalanine residue interacts with the amide oxygen of the leucine residue of a neighboring peptide molecule (I··· O distance 3.481 Å, C-I···O angle 158.89°). The iodine atom on the terminal phenylalanine residue interacts with the amide oxygen of the lysine residue of a near peptide fragment (I··· O distance 3.375 Å, C-I···O angle 165.62°).

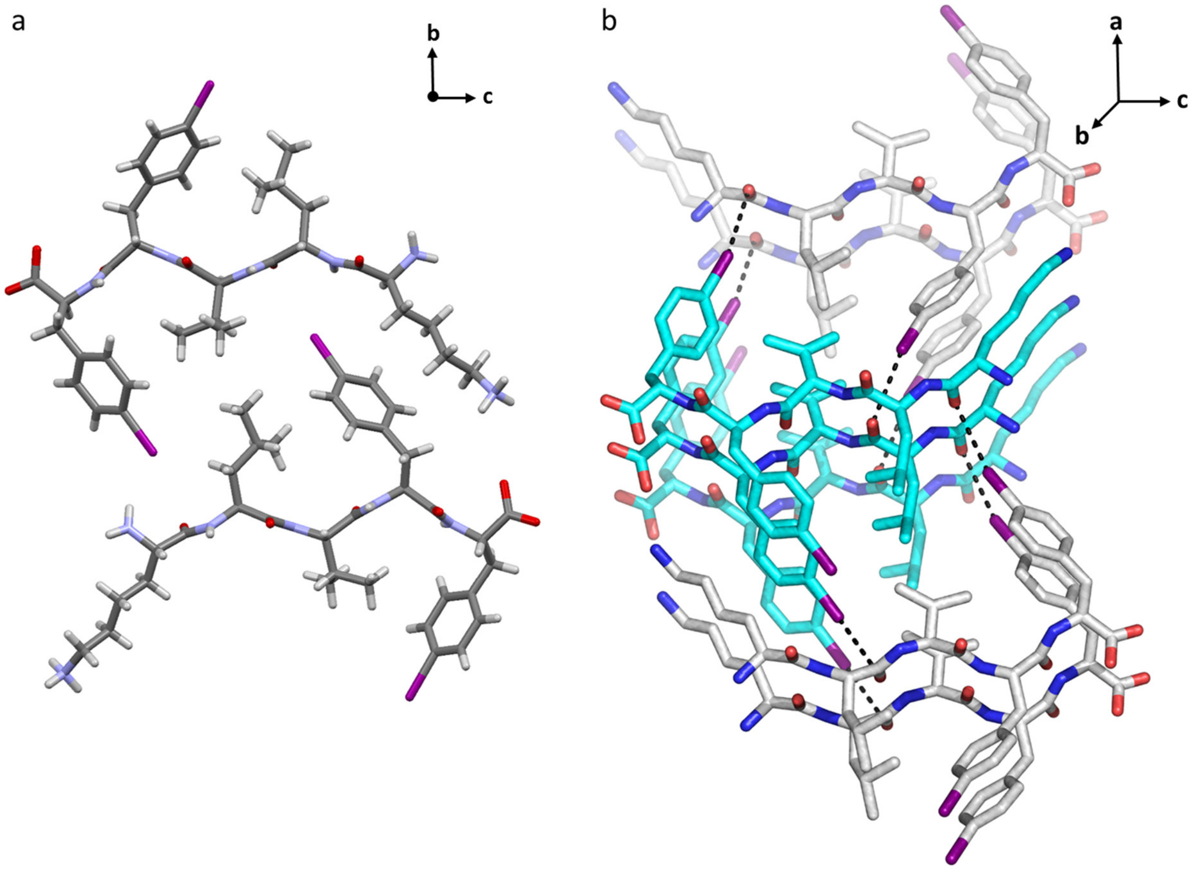


Figure 3: Overview of **KLVF(I)F(I)** steric zipper. View along the crystallographic *a* axis of two facing β-sheets showing the remarkable shape complementarity characteristic of the 'steric zipper' a). Staggered view of interdigitating β-sheets, with short contacts I··· O stabilizing the resulting steric zippers b). Color code: C, grey, sky blue; O, red; N, violet, blue; I, purple; H, white.

It is reported[49,50] that hydrogen (HB) and halogen bonds (XB) occur orthogonally when sharing a common bond acceptor. In this case the common acceptor is the carbonyl oxygen of the peptide bond, usually involved in classical NH···O hydrogen bond with the backbone of an adjacent strand. The orthogonality between HB and XB is confirmed by the angle I···O···N, representing the angle between HB (NH···O) and XB (C-I···O). The geometry is almost orthogonal (~88°) for both angles associated to the different iodine atoms of each monomer. Besides halogen bonding, the iodine atoms stabilize the steric zipper by hydrogen bonding with a paired strand. This interaction, involving the electronegative belt of the iodine atom, confirms its amphoteric properties due to the anisotropic distribution of the electron density. The lateral self-assembly of **KLVF(I)F(I)** is driven by electrostatic interactions: the positive charged lysine side chain stabilizes the C-terminal of another peptide monomer (N··· O distance 2.64 Å, Fig. 4).

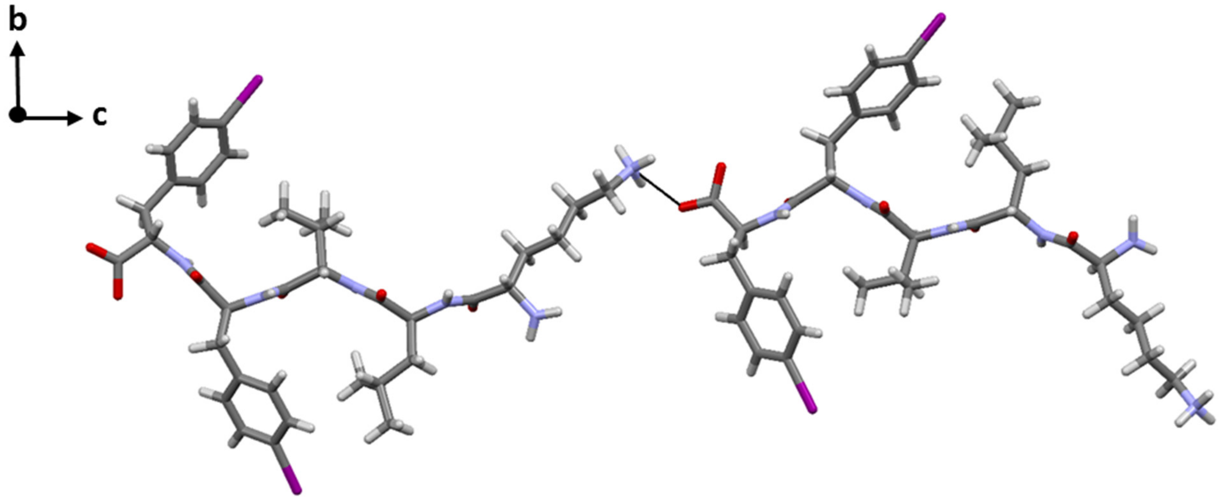


Figure 4: Lateral self-assembly of **KLVF(I)F(I)** driven by electrostatic interactions among the charged termini of the peptide strands (N··· O distance 2.64 Å). Color code: C, grey; O, red; N, violet; **I**, purple; H, white.

The crystal structure of the wild-type fragment KLVFFA has been previously determined,[51] showing polymorphism (PDB$_{\mathrm{ID}}$: 2Y2A, 2Y29, 3OW9). Comparing the structural properties of **KLVF(I)F(I)** and KLVFFA crystals striking differences can be observed. **KLVF(I)F(I)** monomers pair in parallel mode, with the halogenated phenyl rings lying on parallel planes and facing the aliphatic side chains of neighboring β-sheets. Differently, KLVFFA monomers pair in antiparallel mode, resulting in a different type of 'steric zipper' (UP-UP – FACE=BACK CLASS 7, Fig. 20 Chapter 1). While in **KLVF(I)F(I)** the overall inter-sheet stability is due to the presence of iodine, in KLVFFA the packing is stabilized by C-H···π contacts involving two facing phenyl rings belonging to adjacent sheets. Overall, the presence of iodine does not modify the amyloid nature of the sequence, rather it confers more robustness to the crystal packing. Thus, iodinated peptides are expected to self-assemble in solution in a more compact and stable amyloid fiber, thanks to the contribution of halogen bonds to the overall interaction pattern.

3.4.1.2 Crystal structure of KLVF(Br)F(Br)

The crystal structure of the di-brominated peptide **KLVF(Br)F(Br)** was also obtained and allowed a comparison of the interaction pattern as a function of the halogen atom polarizability. This pentapeptide is isostructural to the iodinated one (Fig. 5), with bromine showing the same amphoteric properties as iodine.

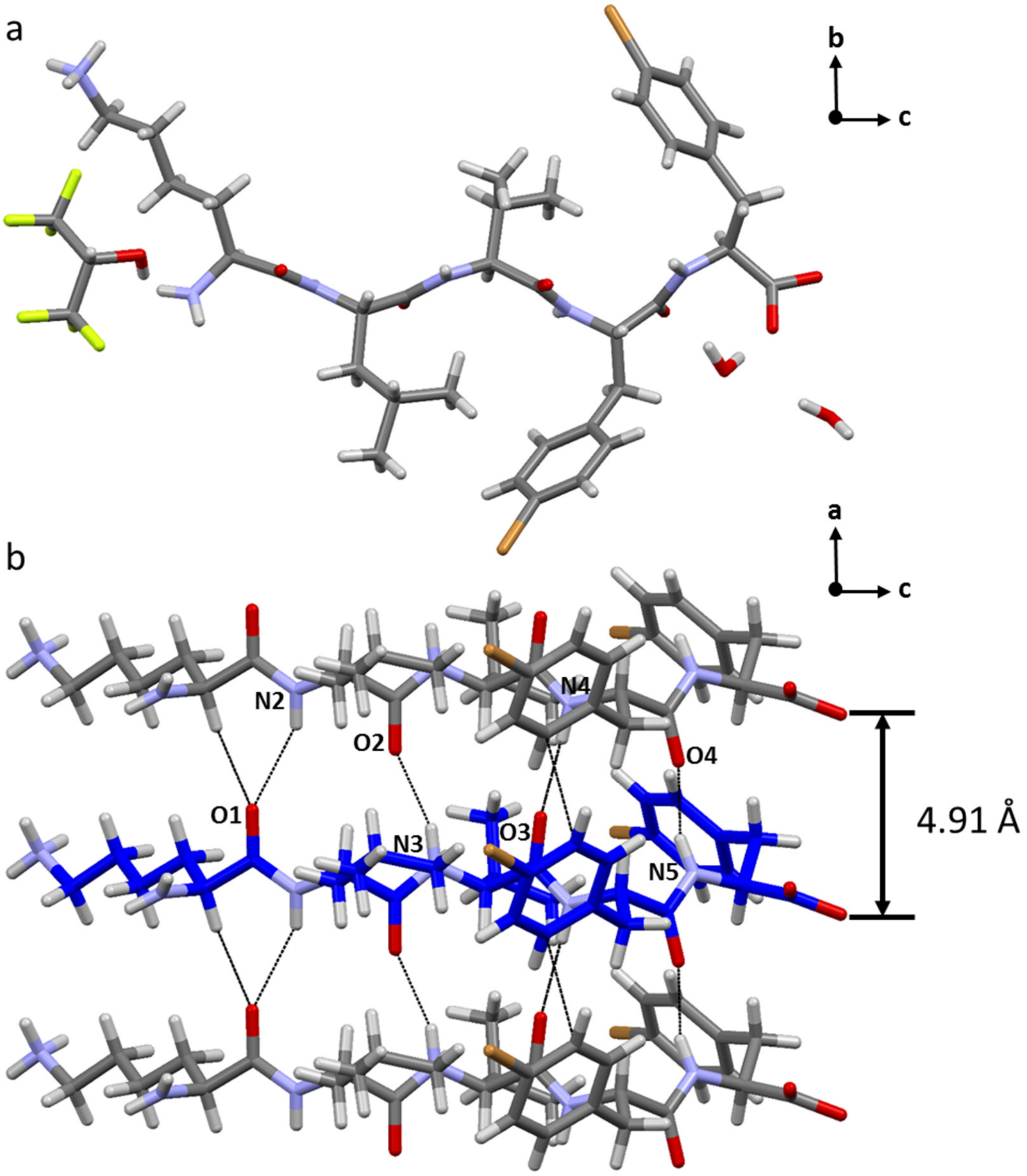


Figure 5: Crystal structure of **KLVF(Br)F(Br)**. Asymmetric unit consisting of a single peptide molecule solvated by two water molecules and one molecule of hexafluoro-2-propanol a). Hydrogen bonding contacts in a parallel β-sheet (Distances: N2H···O1 3.14 Å N3H···O2 3.01 Å N4H···O3 2.89 Å N5H···O4 2.84 Å) b). Color code: C, grey, blue; O, red; N, violet; Br, orange; F, yellow; H, white.

The electropositive region (σ-hole) of bromine is oriented towards the amide oxygen of a neighboring strand, although the distance between the halogen and the oxygen is a little bit longer than the sum of the Van der Waals radii. Although there is no a proper halogen-bond, the geometrical configuration of bromine in the structure shows anyway its propensity to act as a halogen-bond donor, resulting in the same steric zipper as the iodinated peptide (Fig. 6). Also in this case the angle Br···O···N is near to orthogonality (~85°& ~87°) for both bromine atoms of each monomer. Comparing the crystal structures of **KLVF(I)F(I)** and **KLVFF(Br)F(Br)**, we expect to obtain in solution a weaker self-assembly from the brominated peptide, as the contribution of halogen bonding on the stability of the steric zipper is less pronounced.

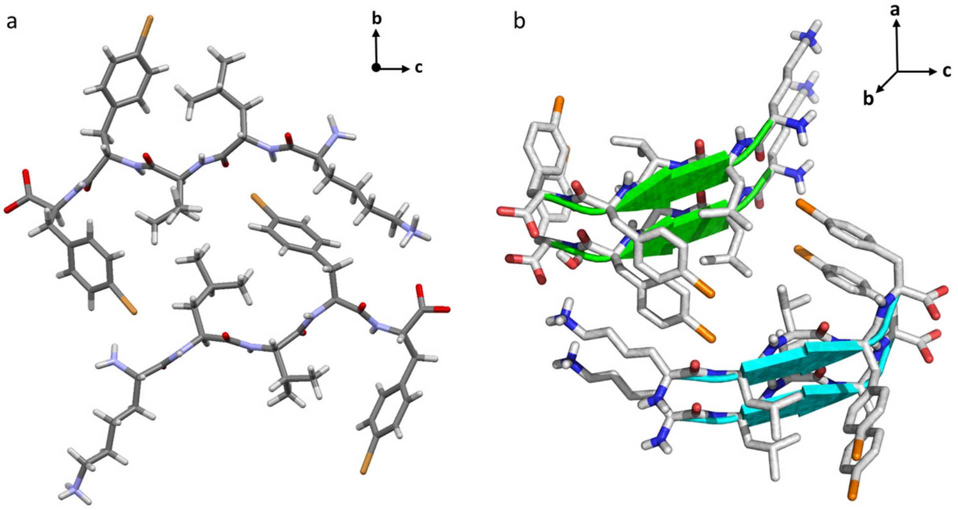


Figure 6: Overview of **KLVF(Br)F(Br)** steric zipper. View along the crystallographic *a* axis of two facing β-sheets showing the remarkable shape complementarity characteristic of the 'steric zipper' a). Staggered view of interdigitating β-sheets b). β-strands are depicted as cartoons. Color code: C, grey; O, red; N, violet, blue; Br, orange; H, white.

3.4.1.3 Crystal structure of KLVF(Cl)F(Cl)

Unlike the di-iodinated and the di-brominated peptides, the crystal structure of **KLVF(Cl)F(Cl)** shows chlorine acting only as nucleophilic species, as expected because of its lower polarizability. The di-chlorinated peptide crystallizes in a different space group ($P2_1$), with the asymmetric unit consisting of a single molecule of **KLVF(Cl)F(Cl)** solvated by four water molecules (Figure 7a). The peptide monomers form extended hydrogen-bonded strands, resulting in parallel β-sheets. In detail, each **KLVF(Cl)F(Cl)** molecule forms eight backbone NH···O hydrogen bonds with the neighboring strands belonging to the same sheet (Fig 7b). The distance between each strand in the β-sheet is 4.91 Å, the same as that of the other di-halogenated peptides.

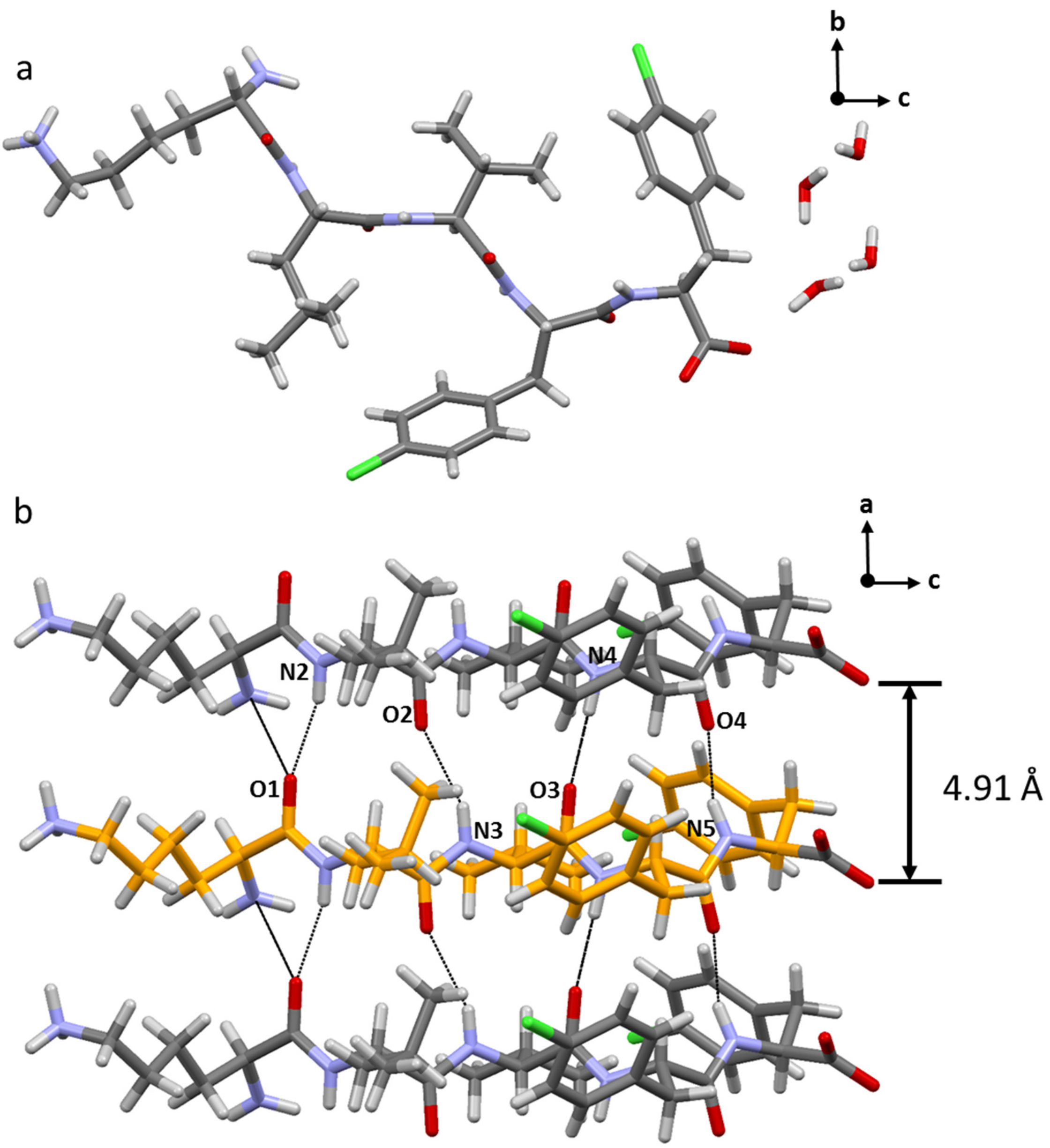


Figure 7: Crystal structure of **KLVF(Cl)F(Cl)**. Asymmetric unit consisting of a single peptide molecule solvated by four water molecules a). Hydrogen bonding contacts in a parallel β-sheet (Distances: N2H···O1 2.95 Å N3H···O2 3.09 Å N4H···O3 2.86 Å N5H···O4 2.83 Å) b). Color code: C, grey and orange; O, red; N, violet; Cl, green; H, white.

The ‘steric zipper’, resulting from the interdigitating side chains of facing β-sheets, can be classified as FACE TO BACK – UP-DOWN – CLASS 4 (Fig 8a).[42] The chlorine atom contributes to the steric zipper stability through weak intermolecular contacts. In particular, the chlorine atom on the inner phenylalanine residue interacts with a methyl hydrogen of the valine residue of a neighboring strand (Cl··· H distance 2.89 Å). The same halogen atom shows a short contact with the amine hydrogen of the N-terminal of a near peptide molecule (Cl··· H distance 2.87 Å). The lateral self-assembly of **KLVF(Cl)F(Cl)** is driven by electrostatic interactions: the carbonyl oxygen at the C-terminus of the peptide strands forms bifurcated salt bridges with the lysine side-chain of nearby peptide monomers (NZH··· O5: 2.99 Å, 2.74 Å). Short contacts between the charged terminus of the lysine side chain and the hydroxyl oxygen of the C-terminal (NZH··· OXT: 2.85 Å) give further stability along the same direction (Fig 8c). Comparing this crystal structure to those of the di-iodinated and di-brominated peptides, it is evident that chlorine shows a more limited interaction pattern than the more polarizable halogen atoms, its electrophilic component being reduced.

For this reason, we expect that chlorinated KLVFF derivatives give rise in solution to less stable self-assembled structures.

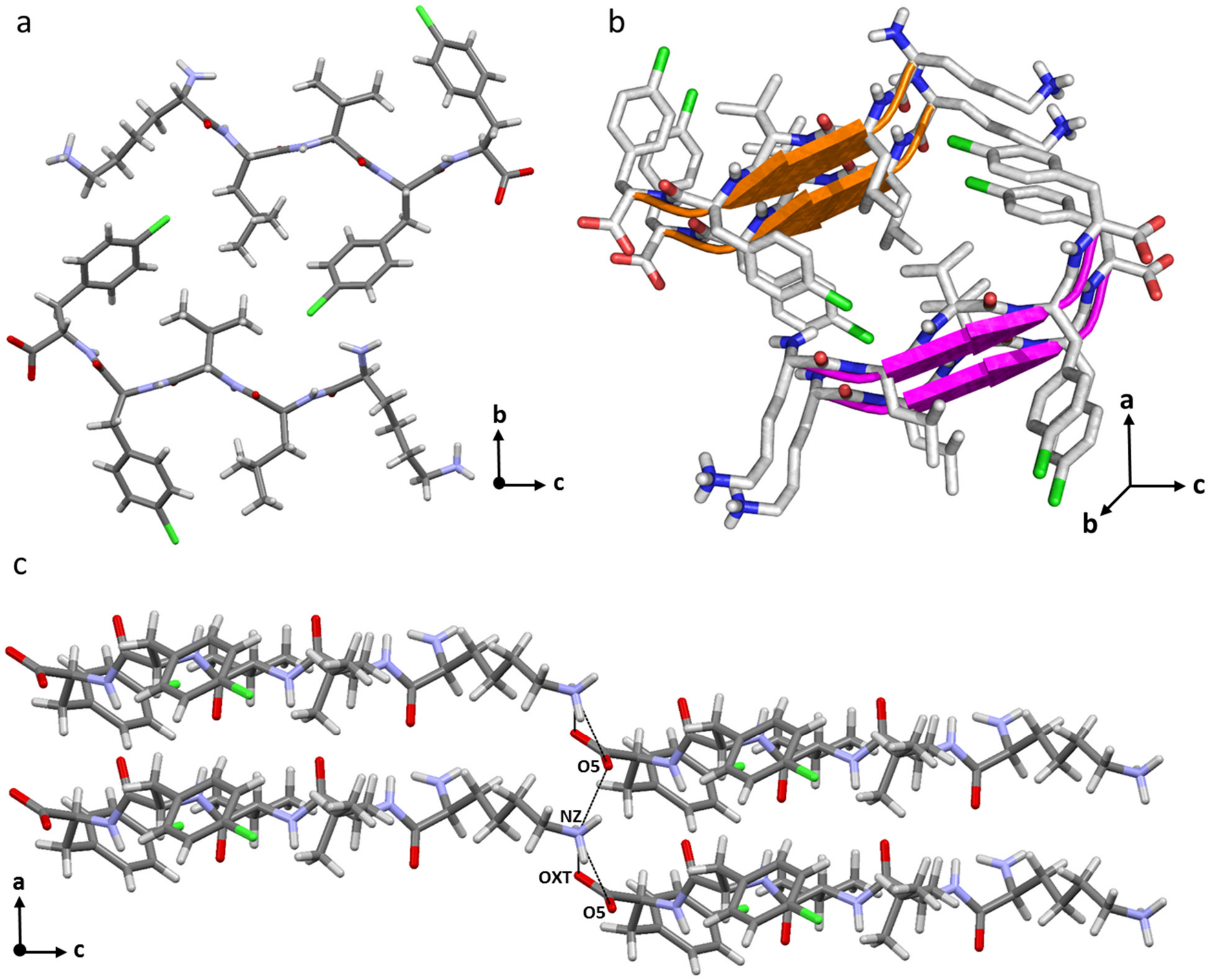


Figure 8: KLVF(Cl)F(Cl) steric zipper and lateral self-assembly. View along the crystallographic *a* axis of two facing β-sheets showing the remarkable shape complementarity characteristic of the 'steric zipper' a). Staggered view of interdigitating β-sheets; β-strands are depicted as cartoons b). Lateral self-assembly of **KLVF(Cl)F(Cl)** driven by electrostatic interactions among the charged groups of the peptide strands (NZ··· O5 distance 2.99 Å, 2.74 Å; NZ··· O5 distance 2.85 Å) c). Color code: C, grey; O, red; N, violet, blue; Cl, green; H, white.

Table 1: Crystallographic data and refinement details for compounds **KLVF(Cl)F(Cl)**, **KLVF(Br)F(Br)** and **KLVF(I)F(I).**

	KLVF(Cl)F(Cl)·$2H_2O$ [$C_{35}H_{50}Cl_2N_6O_6·4H_2O$]	**KLVF(Br)F(Br)**·$C_3H_2F_6O·2H_2O$ [$C_{35}H_{50}Br_2N_6O_6·C_3H_2F_6O·2H_2O$]	**KLVF(I)F(I)**·$C_3H_2F_6O·2H_2O$ [$C_{35}H_{50}I_2N_6O_6·C_3H_2F_6O·2H_2O$]
CCDC Number	1494096	1454959	1454960
Chemical Formula	$C_{38}H_{58}Cl_2N_6O_{10}$	$C_{38}H_{56}F_6Br_2N_6O_9$	$C_{38}H_{56}F_6I_2N_6O_9$
Formula weight	793.77 g/mol	1014.70 g/mol	1108.68 g/mol
Temperature	100(2) K	100(2) K	100(2) K
Wavelength	0.700 Å	0.700 Å	0.850 Å
Crystal system	Monoclinic	Orthorhombic	Orthorhombic
Space Group	$P\ 2_1$	$P\ 2_12_12_1$	$P\ 2_12_12_1$
Unit cell dimensions	a = 4.911(1) Å b = 21.326(4) Å c = 19.682(4) Å α = 90° β = 94.49(3)° γ = 90°	a = 4.909(1) Å b = 20.760(4) Å c = 43.408(9) Å α = 90° β = 90° γ = 90°	a = 4.907(1) Å b = 21.005(4) Å c = 43.152(9) Å α = 90° β = 90° γ = 90°
Volume	2055.0(7) $Å^3$	4423.7(15) $Å^3$	4448.2(15) $Å^3$
Z	2	4	4
Density (calculated)	1.283 g·cm^{-3}	1.524 g·cm^{-3}	1.656 g·cm^{-3}
Absorption coefficient	0.207 mm^{-1}	1.846 mm^{-1}	2.388 mm^{-1}
F(000)	848	2088	2232
Crystal size	0.08 x 0.01 x 0.01 mm^3	0.08 x 0.01 x 0.01 mm^3	0.08 x 0.01 x 0.01 mm^3
Crystal habit	Colorless thin needles	Colorless thin needles	Colorless thin needles
Theta range for data collection	1.02° to 24.32°	1.34° to 16.12°	1.13° to 20.04°
Index ranges	-5 ≤ h 5, -25 ≤ k ≤ 25, -1 ≤ l ≤ 23	-3 ≤ h 3, -15 ≤ k ≤ 16, -34 ≤ l ≤ 29	-3 ≤ h 3, -16 ≤ k ≤ 16, 0 ≤ l ≤ 34
Reflections collected	13325	4168	8932
Independent reflections	3442, 2208 data with I>2σ(I)	2225, 1400 data with I>2σ(I)	2227, 1395 data with I>2σ(I)
Data multiplicity (max resltn)	3.67 (3.55)	2.88 (3.19)	5.87 (6.19)
I/σ(I) (max resltn)	4.56 (1.42)	5.51 (2.90)	5.62 (2.06)
R_{merge} (max resltn)	0.1537 (0.5207)	0.1173 (0.4060)	0.2072 (0.5332)
Data completeness (max resltn)	95.3% (95.1%)	97.7% (98.3%)	96.5% (94.4%)
Refinement method	Full-matrix least-squares on F^2	Full-matrix least-squares on F^2	Full-matrix least-squares on F^2
Data / restraints / parameters	3442 / 13 / 449	2225 / 185 / 255	2227 / 221 / 267
Goodness-of-fit on F^2	0.997	0.998	1.030
Δ/σ_{max}	0.012	0.005	0.009
Final R indices [I>2σ(I)]	R_1 = 0.0810 wR_2 = 0.1981	R_1 = 0.1397 wR_2 = 0.3070	R_1 = 0.0849 wR_2 = 0.1908
R indices (all data)	R_1 = 0.1274 wR_2 = 0.2337	R_1 = 0.1944, wR_2 = 0.3470 R_1(free) = 0.2505 (for 241 reflections)	R_1 = 0.1475, wR_2 = 0.2282 R_1(free) = 0.1577 (for 474 reflections)
Largest diff. peak and hole	0.561 and -0.706 eÅ^{-3}	0.628 and -0.638 eÅ^{-3}	0.860 and -0.714 eÅ^{-3}
R.M.S. deviation from mean	0.081 eÅ^{-3}	0.124 eÅ^{-3}	0.141 eÅ^{-3}

$R_1 = \Sigma ||Fo|–|Fc|| / \Sigma |Fo|$, $wR_2 = \{\Sigma [w(Fo^2 – Fc^2)^2] / \Sigma [w(Fo^2)^2]\}^{½}$

3.4.2 Halogenation and amyloidogenic propensity

Few examples of halogen bonds in biological systems are reported in the literature,[52] suggesting that the potential role of XB in the context of peptide self-assembly is almost unexplored. For example, the effects of halogenation on the self-assembly properties of bio-like systems were studied in the past on the dipeptide **FF**. In particular, the corresponding di-iodinated derivative **F(I)F(I)** is reported to assemble in solution into tubular, β-sheet rich nanostructures,[18] but there is not a detailed crystallographic study on the role of iodine in the folding process of this molecule. Correlating the interaction pattern characterizing a supramolecular assembly in solution to X-ray diffraction data in the solid state is uncommon, since at first glance crystals and fibrillar materials seems to belong to different families. However, their structural similarities have been reported for some amyloidogenic structures.[42] Further evidence of this analogy comes from a recent work by Görbitz, who asserts that diphenylalanine (FF) nanotubes and single crystals are identical.[53] In our case, we studied the effect of halogenation on the self-assembly behavior (both in solution and in the solid state) of more complex systems containing the aromatic moiety FF (KLVFF). As shown in the previous paragraphs, the crystal structures of the di-halogenated peptides, in particular **KLVF(I)F(I)**, showed a not negligible contribution of π-π stacking in stabilizing the overall molecular packing. This indicates that the presence of highly polarizable halogen atoms does not interfere with the most common aromatic interactions, which are considered the main responsible for amyloid formation. Rather, halogen bonding coexists and works in synergy with them. The halogenation effect on the self-assembly of KLVFF derivatives in solution was instead studied through different techniques and the information extracted from the crystal structures were used both as a support and guideline to interpret these results. Halogenated derivatives of KLVFF (Fig. 1) obtained by substitution of phenylalanine with *p*-iodo, *p*-bromo and *p*-chloro phenylalanine, were first evaluated for their capability to assemble into hydrogels,[54] as their formation is highly indicative of fibrillation. Three of the studied peptides – **KLVF(I)F, KLVF(Br)F, KLVF(I)F(I)** – (Figure 9) were found to form gels over a minimum gelation concentration (Table 2). It is important to notice that we obtained hydrogels of the mentioned peptides using much lower concentrations than that of the wild-type fragment KLVFF (60 mM).[8] The best gelator of the series, **KLVF(I)F**, shows a minimum gelation concentration of 7 mM (0.5 % w/w, 8-fold lower than KLVFF). These indicate an active role of the halogen atoms in the fibrillation process, making some of the modified sequences more fibrillogenic than the correspondent wild-type peptide.

Table 2: Minimum gelation concentrations of the studied peptides. For the non-forming gel peptides, the physical aspect of 15 mM samples is reported.

Peptide	Minimum gelation concentration (mM)	Mimimum gelation concentration (% w/w)	Time required for gelation
KLVF(I)F	7	0.5	60 h
KLVF(I)F(I)	15	1.3	15 h
KLVF(Br)F	10	0.7	72 h
KLVFF(I)	-	-	Solution + precipitates
KLVFF(Br)	-	-	Solution + precipitates
KLVF(Br)F(Br)	-	-	Milky solution
KLVF(Cl)F(Cl)	-	-	Slightly milky solution

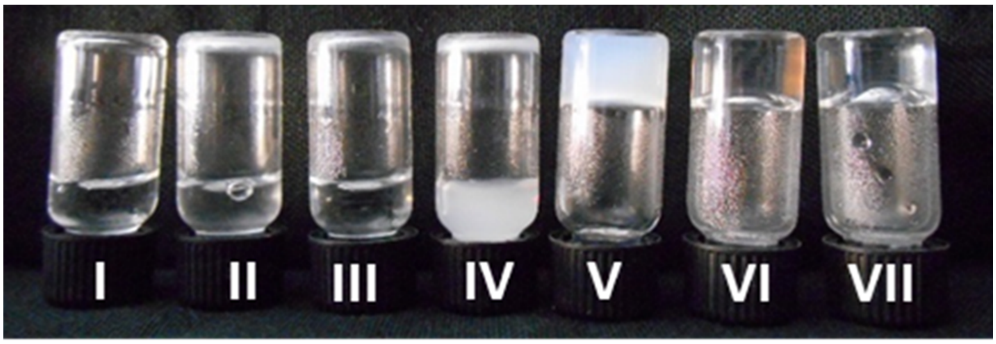


Figure 9: Photograph of vials containing 15 mM peptide solutions upon aging 48 hours at r.t. **KLVFF(I)** (I); **KLVFF(Br)** (II); **KLVF(Cl)F(Cl)** (III); **KLVF(Br)F(Br)** (IV); **KLVF(I)F(I)** (V); **KLVF(I)F** (VI); **KLVF(Br)F** (VII).

In general, halogenation drops the triggering point of the gelation process, as previously showed for the iodination of DFNKF.[4] A working concentration of 15 mM was chosen in order to compare all the halogenated derivatives in the same conditions. At this concentration, the sol-gel transition for the forming gel peptides is complete within a reasonable time (≤ 15 h) (Table 3). The speed of gelation follows the order:

KLVF(I)F > KLVF(Br)F > KLVF(I)F(I)

Table 3: Average gelation time of the forming-gel peptides at 15 mM concentration.

Peptide	Concentration (Mm)	Concentration (%w/w)	Time required for gelation
KLVF(I)F	15	1.1	< 10 min
KLVF(I)F(I)	15	1.3	15 h
KLVF(Br)F	15	1.1	< 30 min

Considering that not all the halogenated derivatives formed hydrogels, it can be inferred that the halogen position in the structure induces different kinds of self-assembly. The pairs **KLVF(I)F – KLVFF(I)** and **KLVF(Br)F – KLVFF(Br)** showed different physical appearances (clear gel *versus* aggregates, respectively) at the same concentration (15 mM). This is a clear example of isomeric peptides, for which the self-assembly properties vary as a function of the position of the halogen atom.[37] In this context, halogenation is a point modification that is able to dictate the type of self-assembly, *i.e.*, a new tool to confer different morphologies to constitutional isomers. The first consideration on halogen position is that introducing a halogen atom on the central phenylalanine (position 4) promotes fibrillation, leading to the formation of a hydrogel. A different behavior characterizes the compounds having a halogen atom on the terminal phenylalanine (position 5), for which the modification induces aggregation and precipitation.

The same concentration of 15 mM was used to characterize the different peptide hydrogels by oscillatory rheology.[55] At this concentration **KLVF(I)F**, **KLVF(Br)F** and **KLVF(I)F(I)** form weak gels. Rheological characterization by frequency and amplitude sweep studies of the peptide hydrogels 48 hours and 2 weeks after preparation was performed. It was not possible to run the experiments for **KLVF(I)F(I)** after 48 h, since it formed a gel not stiff enough to keep the shape of the mold. The rheological analysis of **KLVF(I)F** and **KLVF(Br)F** after 48h (Fig.10) showed a slight difference between the storage and the loss modulus, indicating that the elastic component of the sample is not dominant on the viscous one; thus they behave more as a liquid rather than a solid gel. After 2 weeks an evident effect of aging could be noticed, since the G’ values of **KLVF(I)F** and **KLVF(Br)F** were two-fold compared to those of the first measurement. Furthermore, after 2 weeks also **KLVF(I)F(I)** formed a gel suitable for rheology, similar in strength to **KLVF(I)F** after 48 h. Although all the obtained gels broke at very low strains compared to other small peptide hydrogels reported in the literature,[4,56–58] the rheological experiments confirmed that among the studied fragments the best gelling agents are **KLVF(I)F** and **KLVF(Br)F**, which have the halogen atom at position 4 (inner phenylalanine ring). **KLVF(I)F(I)**, having an additional halogen on the position 5, forms a weaker gel that becomes stiff enough for rheology only after 2 weeks.

Considering the series **KLVF(I)F** – **KLVFF(I)** – **KLVF(I)F(I)**, it is interesting to observe that the properties of **KLVF(I)F(I)** seem to be the combination of those of the two mono-iodinated peptides: **KLVF(I)F** forms clear gels, **KLVFF(I)** tends to form insoluble aggregates, while **KLVF(I)F(I)** forms very weak, milky gels.

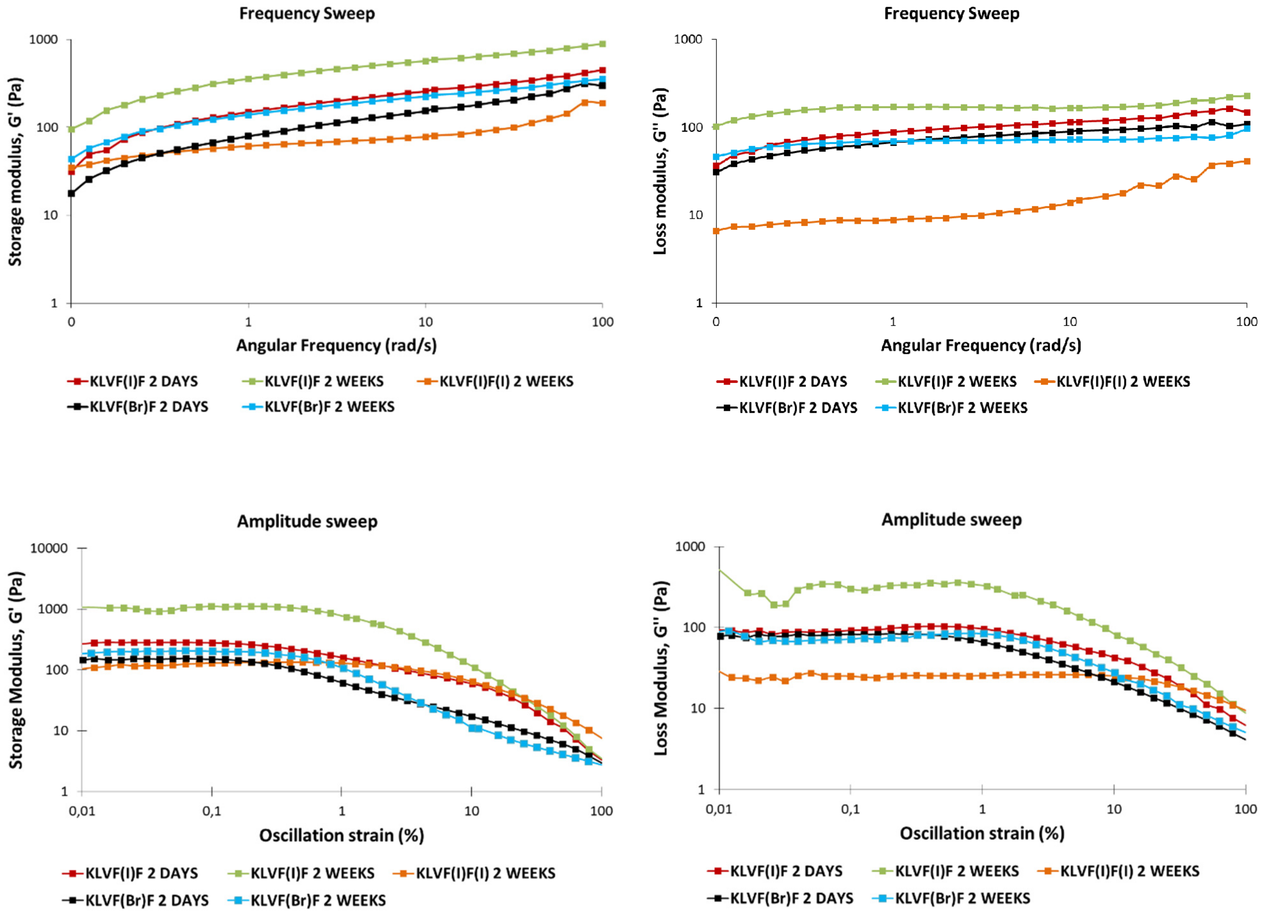


Figure 10: Reological characterization of the gels by frequency sweep studies whereby the storage modulus (G’) was recorded as a function of angular frequency (ω). Amplitude sweep studies of peptide hydrogels showing G’’ as a function of oscillation strain.

The derivatives of the bromine series **KLVF(Br)F** – **KLVFF(Br)** – **KLVF(Br)F(Br)** followed the same trend indicating the influence of the halogen position, except for the di-brominated peptide, which did not form a gel.

Under the hypothesis that halogen bonding contributes to the peptide folding, iodine substituents are expected to give stronger interactions than bromine and chlorine ones, since halogen-bond strength increases with halogen atom polarizability. As a consequence, considering the series **KLVF(I)F(I)** – **KLVF(Br)F(Br)** – **KLVF(Cl)F(Cl)**, it was reasonable that the iodinated peptide formed gels while the brominated and the chlorinated ones did not. The crystal structures of the di-halogenated peptides were consistent with their macroscopic behavior. The iodinated peptide showed additional inter-sheet halogen bonding among the many interactions stabilizing the ‘cross-β’ structure and self-assembled into a solid-like material. The brominated peptide was isostructural to the iodinated one; however, the bromine atom was less prone to act as electrophilic species, hence the additional stability resulting from inter-sheet halogen bonding was less pronounced. Consequently, the self-assemblies of **KLVF(Br)F(Br)** were weaker, consistent with a liquid-like solution. Even the di-chlorinated peptide, similarly to **KLVF(Br)F(Br)**, self-assembles into structures that do not lead to a solid-like material. Another indication that iodine substituents are more efficient halogen-bond donor groups comes from the fact that **KLVF(I)F** forms a gel with higher G' values than **KLVF(Br)F**.

The morphology of the fibrillar network constituting the halogenated hydrogels was assessed by confocal microscopy[59] (Fig. 11).

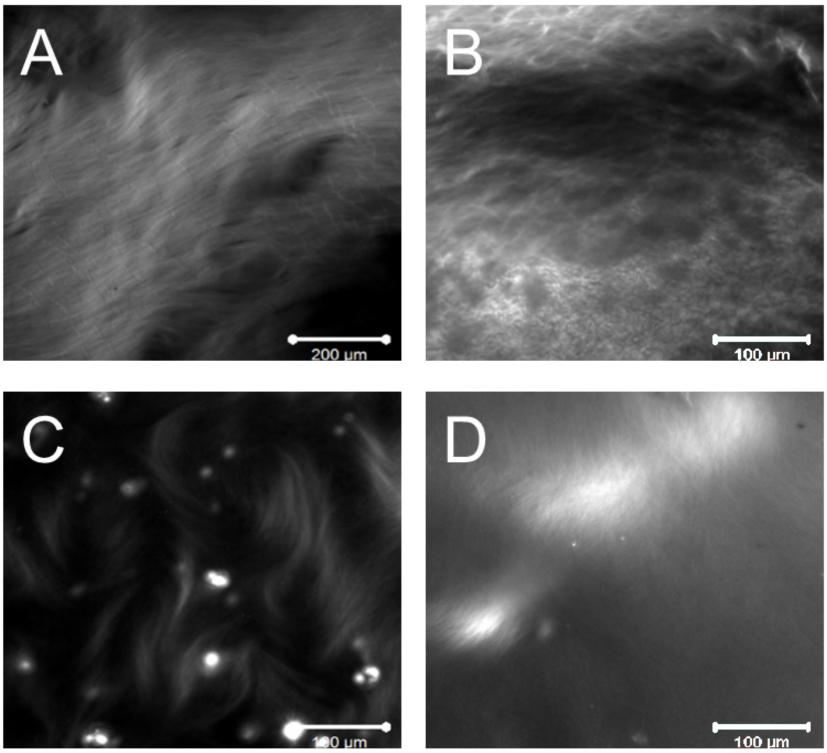


Figure 11: Confocal microscopy images of: **KLVF(I)F** after 48 h (A); **KLVF(I)F** after 2 weeks (B); **KLVF(I)F(I)** after 48 h (C); **KLVF(I)F(I)** after 2 weeks (D).

The network of the hydrogel of **KLVF(I)F** (Figure 11a) appeared to be quite filamentous, composed by thin, micrometer-long fibers propagating along the same direction. These fibrils were less dense, thinner and shorter for **KLVF(I)F(I)** (Fig. 11c). The same samples imaged two weeks after preparation showed more entangled networks (Fig. 11b and 11d). In particular for **KLVF(I)F(I)** the fibrils become longer, therefore able to form the continuous network typical of a hydrogel. It was not possible to obtain clear results from the imaging of **KLVF(Br)F** ever after two weeks, even if it formed a proper hydrogel. The reason may be due to the size of the fibers; very thin assemblies may form a strongly packed network where the dye used for the analysis (Rhodamine B) is not able to intercalate, without giving enough contrast to identify any kind of structure.

There are three different criteria to define the amyloid nature of a protein aggregate: green birefringence upon staining with Congo Red, fibrillar morphology and β-sheet secondary structure.[60] For this reason, Congo Red staining was used as first technique to characterize all the peptide hydrogels (15 mM concentration 48 h after preparation). The peptides containing a halogen atom at position 4 showed a bright green birefringence, as a result of Congo Red binding to amyloids (Fig. 12).

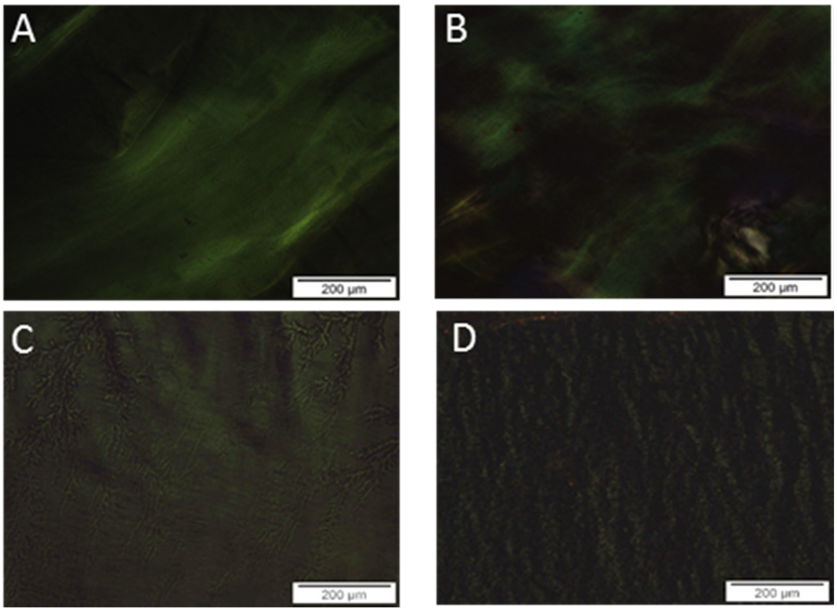


Figure 12: Congo red staining of **KLVF(I)F** (A); **KLVF(I)F(I)** (B); **KLVF(Br)F** (C); **KLVF(Br)F(Br)** (D).

The positive staining of **KLVF(I)F(I)** and **KLVF(Br)F(Br)** confirmed that they formed fibrils although the resulting network was not so dense to give a proper hydrogel. The di-chlorinated peptide **KLVF(Cl)F(Cl)**, unlike its parent di-halogenated derivatives, did not display green birefringence. It is interesting to notice that the propensity to form amyloids decreases with the halogen atom polarizability, following the order:

KLVF(I)F(I) > **KLVF(Br)F(Br)** > **KLVF(Cl)F(Cl)**

In fact, **KLVF(I)F(I)** gave positive staining, forming a gel after 2 weeks; **KLVF(Br)F(Br)** gave positive staining without forming a proper gel, while **KLVF(Cl)F(Cl)** did not display any of these properties. Finally **KLVFF**, **KLVFF(I)**, and **KLVFF(Br)** did not show any green birefringence as expected.

Further evidence that the halogenated peptides self-assembled into amyloid structures was obtained by Fourier Transform Infrared (FT-IR) spectroscopy. The FT-IR spectra of each compound (15 mM concentration in D_2O) were recorded 48 h after preparation in order to have a direct comparison with the results of the other experimental techniques (Fig. 13).

Figure 13: FTIR spectroscopy of 15 mM peptides gels/solutions after standing for 48 hours at r.t. The red lines represent the second derivative of the FTIR spectra.

The FTIR spectra of the gel-forming peptides – **KLVF(I)F**, **KLVF(Br)F**, **KLVF(I)F(I)** – after 48 h were characterized by an intense peak ranging from 1625 to 1630 cm^{-1}. This signal was assigned to the peptide bond carbonyl stretching upon formation of the amyloid β-sheet structure, which is indicative of gelation.[61] All the other peptides showed, after 48 h, a series of peaks of similar intensity that could be associated to both β-sheet (1620-1630 cm^{-1}) and other secondary structures like β-turns and α-helix.[62,63] These results are reasonable considering that for the non-forming gel peptides the β-sheet component does not prevail, while for the forming gel peptides it does.

Circular Dichroism (CD) spectroscopy (Figure 14) of diluted solutions (400 μM) of the KLVFF derivatives further confirms that β-sheet secondary structure is predominant in all the peptides. Amide groups have two electronic transitions of low energy – n→π* and π→π* – giving information about the secondary structure of peptides and proteins.[64,65] In the typical spectrum of a β-sheet protein the transition n→π* is primarily responsible for a negative band in the range 216-218 nm, while the transition π→π* results in a positive band around 195 nm. Compared to helical proteins, where CD allows unambiguously the determination of the secondary structure, in β-sheet rich proteins the peak assignment is less reliable. In fact, β-sheet arrangement presents a wider structural diversity in length, extent, degree of twist and distortion[66] as well as β-strand orientation (parallel or antiparallel). These differences relate to diverse CD signatures, although previous studies reported a poor correlation among β-sheet structural diversity and CD spectra. Recently, a systematic study combining CD and crystallographic features of some β-sheet rich proteins definitely revealed that orientation and twist of the β-sheets strongly affect the observed CD spectral shape.[67] In particular, for proteins having the same β-sheet content but different twists, inversions of maxima and remarkable peak shifts may occur. This finding can explain why the CD signature of the halogenated KLVFF peptides shows considerable deviations from the CD spectra of a classical β-sheet protein. Moreover, also the contribution of aromatic stacking interactions of the phenylalanine residues should be considered in this case.[68-70] In order to give a quantitative estimation of the β-sheet content starting from the CD spectra of the halogenated peptides, we analyzed each spectra with a novel algorithm called β-Structure Selection (BeStSel).[67] The secondary structure was determined also taking into account the twisting angles between β-strands. The algorithm's prediction (Table 4) showed that all the KLVFF derivatives have a main component of β-sheet, with a significant percentage of unordered structure. In particular, considering the iodinated and the brominated peptide series separately (red and blue box in Table 4) it was evident that mono-halogenation on the terminal phenylalanine residue (position 5) resulted in a lower β-sheet contribution compared to mono-halogenation on the inner phenylalanine ring (position 4). Di-halogenation is by far the configuration conferring the highest β-sheet content for both iodinated and brominated peptides series. The secondary structure estimation reported in table 4 is the result of the fit curves presented in figure 15.

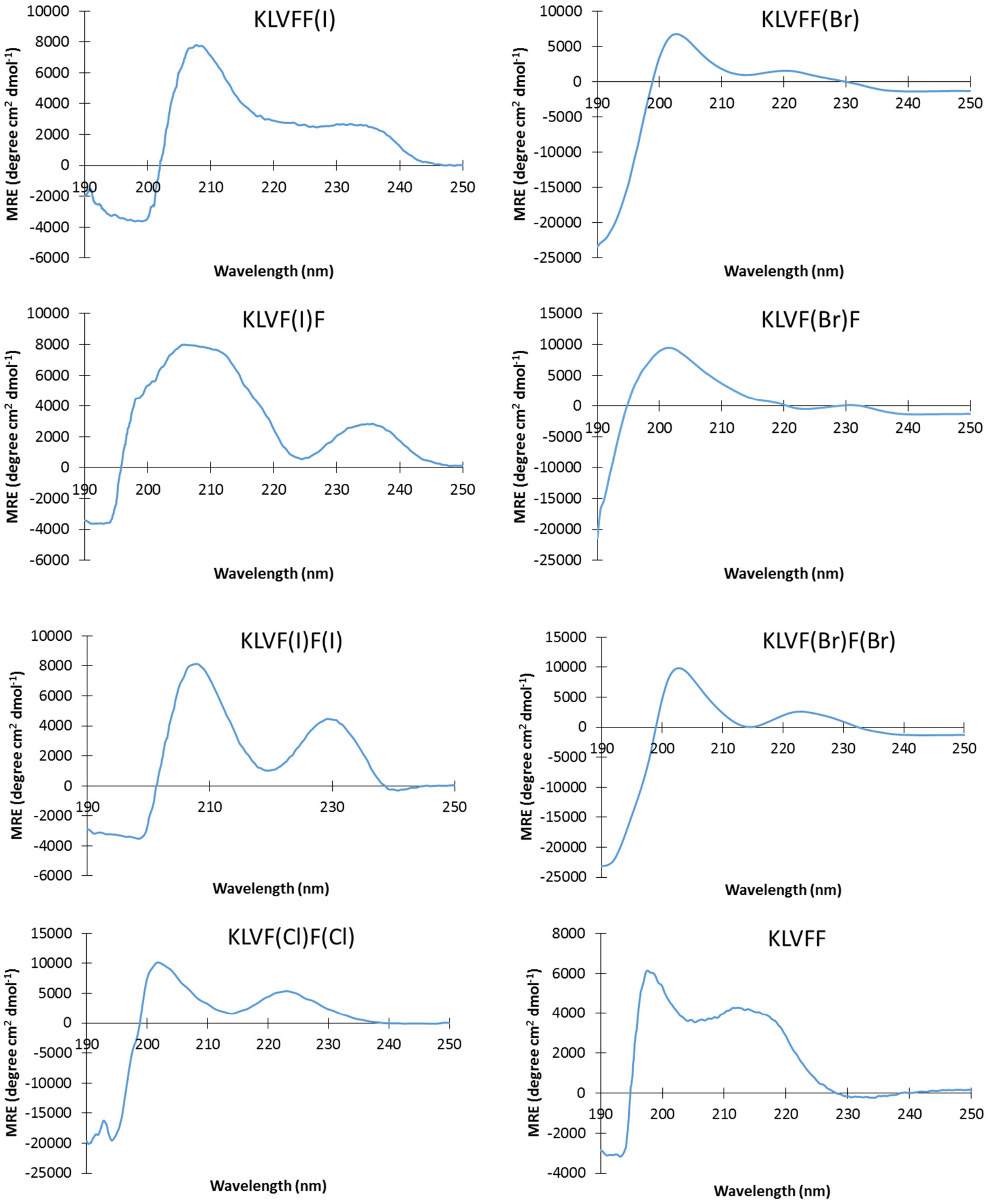


Figure 14: Circular Dichroism spectra of the KLVFF^8 halogenated derivatives. All the spectra were recorded in deionized water at 400 μM concentration.

Table 4: Secondary structure estimation (%) of the algorithm BeStSel for the halogenated KLVFF derivatives. Considering the iodinated peptides series (red box) the β-sheet content increases in the order **KLVFF(I) < KLVF(I)F < KLVF(I)F(I)**. For the brominated peptides series (blue box) the trend is the same.

	β-sheet	Turn	Other	Helix
KLVFF	46	13.6	40.4	-
KLVF(I)F(I)	55.2	10.6	34.2	-
KLVF(I)F	52.5	12	35.5	-
KLVFF(I)	48.8	17.2	34	-
KLVF(Br)F(Br)	58	10.4	29	2.5
KLVF(Br)F	40.2	15.5	44	0.4
KLVFF(Br)	32.5	18.3	49.3	-
KLVF(Cl)F(Cl)	72.2	12.8	15	-

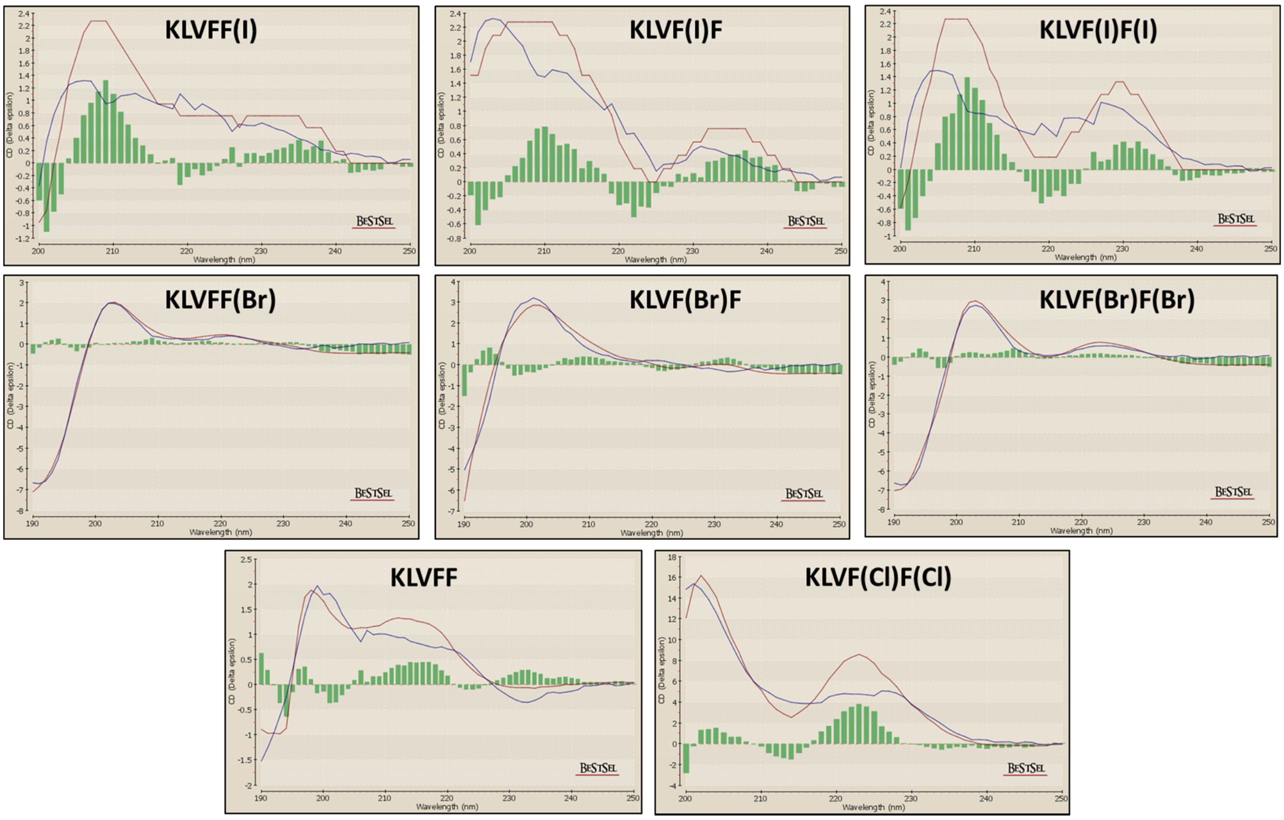


Figure 15: Fit curves of the CD spectra of the KLVFF halogenated derivatives, calculated with BeStSel[67] algorithm. The red line represents the experimental data, the blue line is the fitting curve.

In conclusion, according to the experimental results so far described, the amyloidogenic propensity of the KLVFF sequence is preserved in all the halogenated peptides, although the formation of β-sheet rich material is not clearly detectable with all the techniques. The peptides **KLVF(I)F**, **KLVF(I)F(I)** and **KLVF(Br)F** show macroscopically their fibrillation propensity by forming hydrogels, in addition to positive Congo red staining and clear FTIR signature. For the di-iodinated peptide there is also the direct evidence of the crystal structure showing the 'steric zipper', a peculiar motif of amyloid materials. Although the di-brominated peptide does not form gels, it is clearly amyloidogenic as demonstrated by Congo red staining, FTIR signature and crystal structure showing the characteristic cross-β pattern. The di-chlorinated sequence shows a less marked amyloid behavior, as confirmed by the fact that the unique clear experimental evidence comes from the crystal structure. The peptides having a single halogen atom on the terminal phenylalanine residues, *i.e.* **KLVFF(I)** and **KLVFF(Br)**, have less pronounced amyloidogenic tendency, although their crystal structures (described in chapter 4) show the peculiar β-sheet arrangement. In addition to XRD analysis, also FTIR spectra of **KLVFF(I)** and **KLVFF(Br)** suggest the presence of a β-sheet component, despite the self-assembly of these peptides do not lead to hydrogels. CD spectroscopy further supports β-sheet as favorite structural arrangement for all the halogenated derivatives. However, we hesitate to draw definitive conclusions about the trend observed by CD, since these results refer to diluted peptide solutions. Moreover, the CD signature of the halogenated KLVFF derivatives could be affected by π-stacking among phenylalanine residues. Although halogenation is a punctual modification, it greatly affects the amyloid behavior of the KLVFF peptide. In particular, upon changing position, number and nature of the halogen atoms in the sequence, the overall effect can be both an amplified and a less pronounced fibrillogenic propensity.

3.4.3 Tuning KLVFF nanostructure in solution by halogenation

Transmission electron microscopy (TEM) was used to image the nanostructures (Figure 15) formed by the halogenated peptides (15 mM, 48 h after preparation). Considering that the unique modification on the peptide fragment is halogenation, a huge variety of morphologies could be observed. As expected, the non-forming gel peptides did not show any entangled network of fibers. In particular, **KLVFF(I)** (Fig 16a) dispersions contained quite definite spherical nanoparticles.

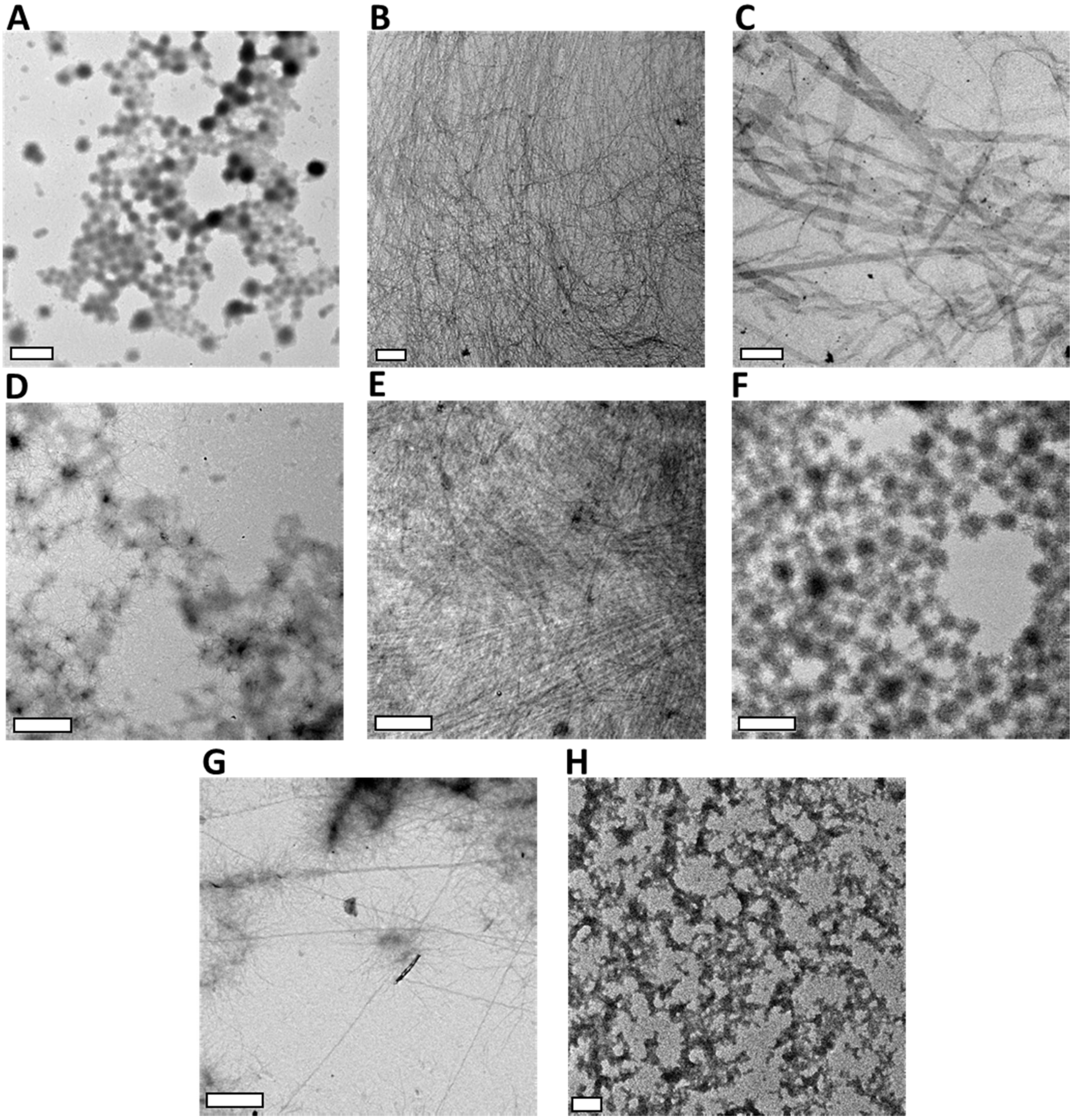


Figure 16: TEM images of 15 mM gels/solutions of the halogenated derivatives of KLVFF showing different self-assemblies varying position, number and nature of the halogen atoms. **KLVFF(I)** (A); **KLVF(I)F** (B); **KLVF(I)F(I)** (C); **KLVFF(Br)** (D); **KLVF(Br)F** (E); **KLVF(Br)F(Br)** (F); **KLVF(Cl)F(Cl)** (G); **KLVFF** (H). The scale bar is 1µm.

Nanoparticles formed by pure peptide material are quite rare. Compared to the wide population of synthetic polymers that are able to self-assemble into nanoparticles, only few examples of peptides showing this behavior have been reported so far.[71] These peptides, defined as "Surfactant-Like Peptides" or "Peptide Amphiphiles" (PA),[72,73] can assemble into a variety of nanostructures.[74] This class of compounds combines the structural features of amphiphilic surfactants with the typical self-assembly properties of peptides. A PA structure can be a simple peptide sequence containing either hydrophilic and hydrophobic domains, or a more complex molecule consisting of an alkyl chain bound to a hydrophilic amino acid sequence. **KLVFF(I)** is an example of pure peptide material behaving as amphiphile, with the lysine residue acting as hydrophilic domain and the remaining part of the sequence characterized by high hydrophobicity.[75] The iodine atom, located in a peripheral position in the hydrophobic domain, contributes to further increase the lipophilic degree of the molecule. A complex balance between different energy contributions represents the driving force directing PA self-assembly, including, among others, hydrophobic interactions, hydrogen bonding between the peptide units, and electrostatic repulsions among the charged amino acids. The size and shape of the final assemblies are the result of the combination of these three components.[76] In general, when a peptide amphiphile includes amino acid sequences with high propensity to form β-sheets, long cylindrical nanofibers are the most common supramolecular arrangements. This is the case of **KLVF(I)F** and **KLVF(Br)F** (Fig. 16b, Fig. 16e), showing a dense network of entangled fibers. Considering that also the segment **KLVFF(I)** possesses an extended β-sheet component – although less pronounced than the forming-gel KLVFF peptides – the spherical shaped arrangement is quite uncommon. In fact, most self-assembling peptides forming nanoparticles adopt helical conformations or even completely lack a defined secondary structure. However, a few studies describing spherical assemblies stabilized by β-sheets have been reported quite recently.[77,78] Considering the aggregation behavior of PAs in water (as in our study), the key factor determining the shape of the self-assembly is related to the interfacial curvature, which is dictated by a network of sheet-like structures formed by multiple peptide molecules and linked together through intermolecular hydrogen bonding. Molecular simulations studying the self-assembly of this class of compounds[76] showed the formation of several kind of nanostructures (Fig. 17), as a result of a delicate balance between hydrophobic interaction (ε_{HH}) and hydrogen bonding involving the peptide units (ε_{β}).

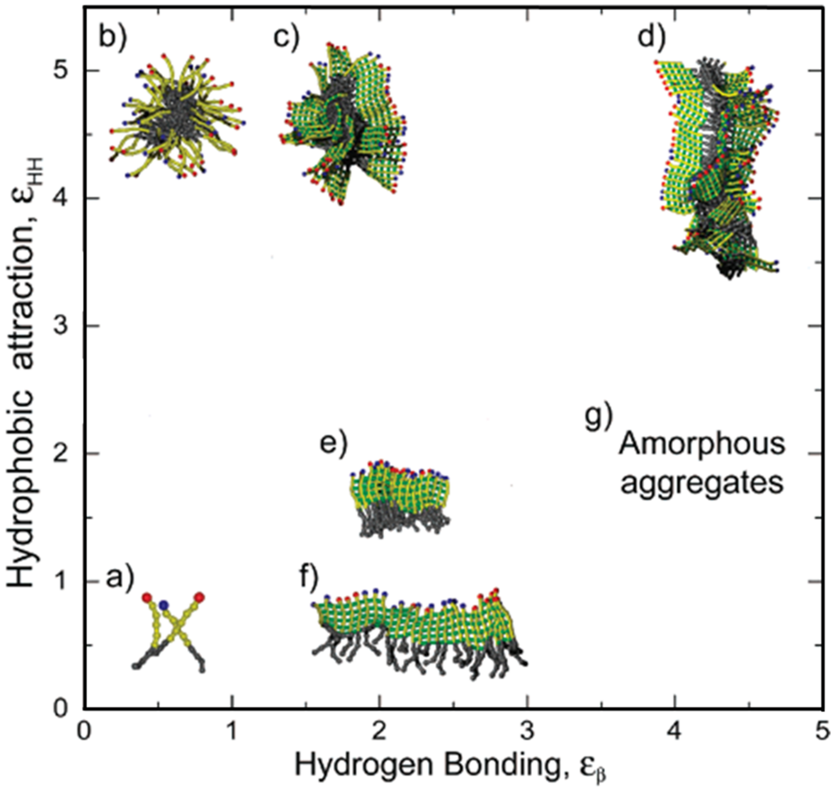


Figure 17: Schematic phase diagram showing different self-assemblies depending on the balance among the energetic contributions of hydrophobic attractions (ε_{HH}) and the energetic component associated to hydrogen bonding (ε_{β}). Free molecules a) Spherical micelles b) Spherical aggregates with β-sheets on the external shell c) long cylindrical fibers d) stacks of parallel β-sheets e) single β-sheets f) amorphous aggregates g).

Considering the ideal situation of total absence of interactive forces, the peptide amphiphiles do not associate and are present in solution as free molecules (Fig 17a). Excluding the energy contribution of hydrogen bonding (ε_{β} ~ 0), pure hydrophobic interactions generate discrete self-assemblies like micelles (Fig. 17b). The formation of micelles follows a complex association pathway, with the initial nucleation process occurring above a certain concentration limit (critical micellar concentration) or temperature. In contrast, a system dominated by pure hydrogen bonding (ε_{HH} ~ 0) results in one-dimensional β-sheets (Fig. 17f). When both energy components coexist in the same system, the resulting assemblies are a combination of those described for the edge cases. When the energy contribution of hydrogen bonding is relatively weak compared to the hydrophobic component, the resulting spherical aggregates may contain randomly oriented β-sheets in the external shell (Fig. 17c). In this case, the peculiar β-sheet rigidity is partially compromised to fit into the spherical interfacial curvature. This model could be a possible description of **KLVFF(I)** aggregation behavior; moreover, the hypothesis of β-sheet distortion to fit the spherical shape can explain why this peptide shows a much less pronounced amyloidogenic propensity (Section 4.2). Increasing contribution of hydrogen bonding combined with high hydrophobic energy disrupt the spherical symmetry and lead to formation of one-dimensional cylindrical fibers, where β-sheets grow along the same direction of the fiber axis (Fig 17d). As already mentioned, a strong hydrogen bonding energy leads to formation of long β-sheets. However, due to the increasing number of stacking peptide units, the local accumulation of hydrophobic domains into restricted regions of the β-sheet occurs. This deeply changes the local net hydrophobicity of the aggregating system, resulting in an energetic imbalance with the surrounding. The main consequence on the self-assembly is the formation of various amorphous aggregates (Fig. 17g). The wild type peptide (Fig. 16h) and **KLVFF(Br)** (Fig. 16d) show this behavior in the reported experimental conditions (15 mM), probably because the overall contribution of hydrophobic attraction (ε_{HH}) is not enough to drive the self-assembly towards ordered structures like cylindrical fibers (Fig 17d) or β-sheet spherical aggregates (Fig. 17c). The structural features of a β-sheet depend on the backbone dihedral angles, which describe the geometrical configuration of the peptide bond. Usually, these angular values lead β-sheets to be planar, which is the most common conformation. However, a minimal change in one of these angles may affect β-sheet linearity and induce helicity. The self-assembly of helical β-sheets results in more complex morphologies compared to those formed by planar β-sheets.[79] This is the case of the di-iodinated peptide **KLVF(I)F(I)** (Fig. 16c), showing a mixture of straight, twisted and helical ribbons (Fig. 18). Because of the large size of these assemblies, the resulting hydrogel network does not appear compact and entangled like the one characterizing **KLVF(I)F** and **KLVF(Br)F** (Fig. 16b, Fig. 16e), showing the classical fibril-like morphology.

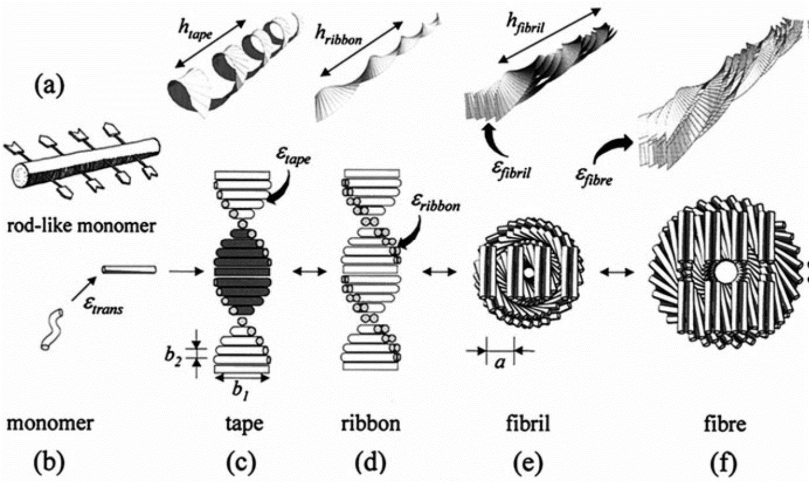


Figure 18: Hierarchical self-assembly model for peptide monomers in a β-strand conformation. Letter ε indicates the stacking attraction energy between structural units constituting each conformation.[80]

This is consistent with the rheology analysis, with the di-iodinated segment forming weaker gels than the peptides halogenated on the inner phenylalanine residue. The di-brominated peptide **KLVF(Br)F(Br)** (Fig. 16f), giving a positive result by Congo red staining without forming a proper gel, shows a series of poorly defined spherical structures larger than the vesicles formed by **KLVFF(I)**. These aggregates are associated to nucleation centers from which undeveloped fibrillary structures branch out. These germinal fibers are responsible for the positive staining, although entanglement among these fibers does not lead to a proper gel formation. The di-chlorinated segment **KLVF(Cl)F(Cl)** shows a mixture of fibrillar aggregates and isolated fibrils, which apparently are not enough to lead to a positive staining. Polymorphism among amyloid structures formed by different peptides or proteins[81–84] is a well-known feature characterizing this kind of materials. The structural differences depend on how the amino acid side-chains interact each other sterically, electrostatically and with the surrounding solvent. This phenomenon can also occur for the same amino acid sequence[81] by varying some experimental conditions like concentration, pH and ionic strength. In our case, the polymorphism of the peptide fragment KLVFF is determined by changing position, number and nature of the halogen atoms in the sequence.

Since TEM requires sample drying before analysis, we considered the possibility that the sample's preparation mode may affect the appearance of the observed objects. In order to exclude this effect, we decided to study some of the halogenated 15 mM peptides through cryogenic Transmission Electron Microscopy (cryo-TEM), which allows visualizing the nanostructure classes present in solution. Among the many possible self-assembled structures showed by the halogenated derivatives of KLVFF, we selected for cryo-TEM study the samples showing the most singular nanostructures, *i.e.* the spherical aggregates formed by **KLVFF(I)** (Fig. 16a)**,** the "cotton ball-like" material formed by **KLVF(Br)F(Br)** (Fig. 16f) and the ribbons formed by **KLVF(I)F(I)** (Fig. 16c).

3.4.4 Characterization of KLVFF(I), KLVF(I)F(I) and KLVF(Br)F(Br) self-assemblies

The ability of **KLVFF(I)** to form spherical assemblies was confirmed also by cryo-TEM. Indeed, 15 mM solution of the peptide (sample aged two months) yielded round-shaped nanoparticles (NPs), as shown in Figure 19.

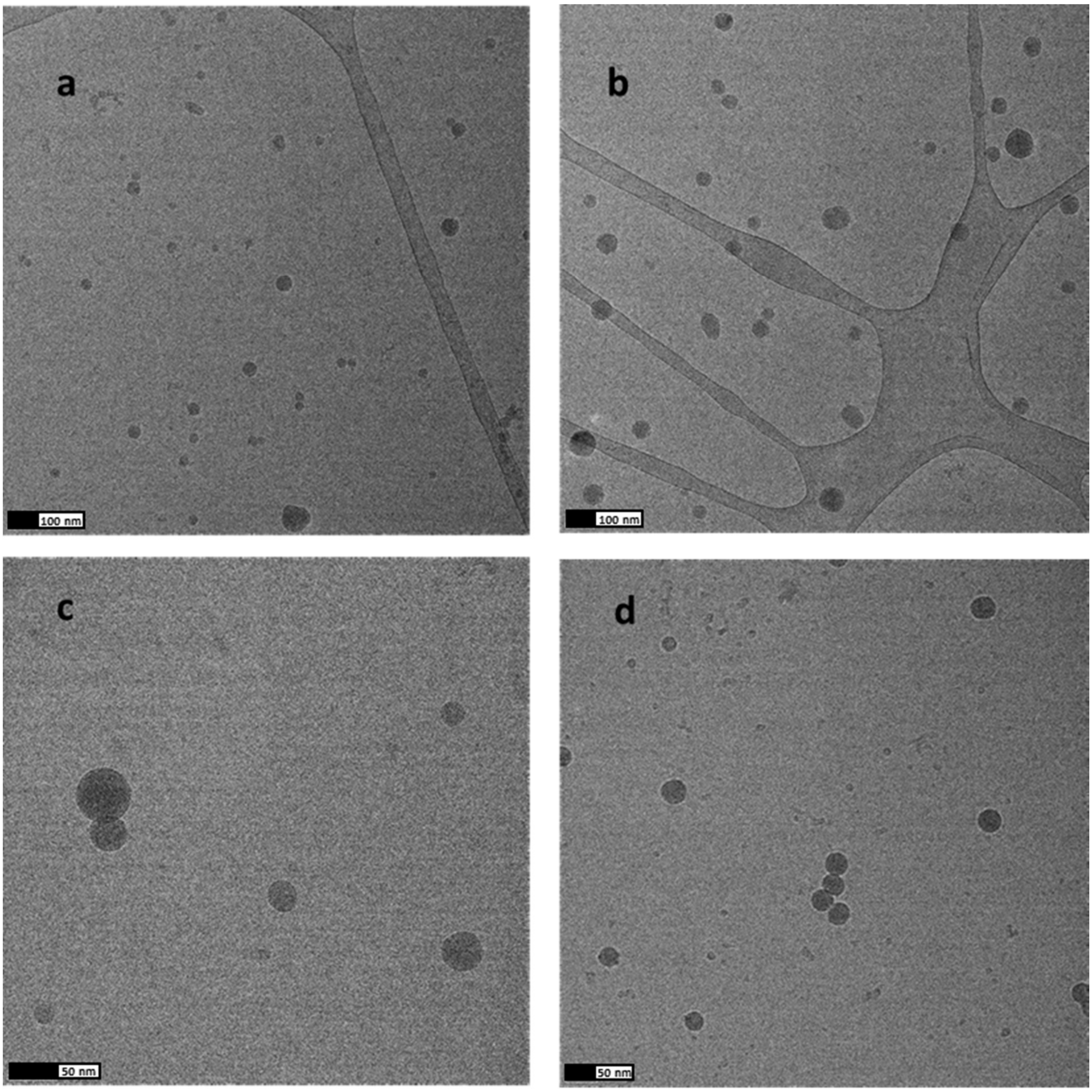


Figure 19: cryo-TEM images of **KLVFF(I)** 15 mM.

Cryo-TEM results are in good agreement with what observed in the dried state. Even TEM (Fig. 15a) showed the presence of spherical aggregates, although these nanostructures appeared collapsed, bigger and less uniform in size because of the drying process. Statistical analysis of cryo-TEM images led to an average size for **KLVFF(I)** 15 mM NPs, which is 42 ± 26 nm (Fig. 20).

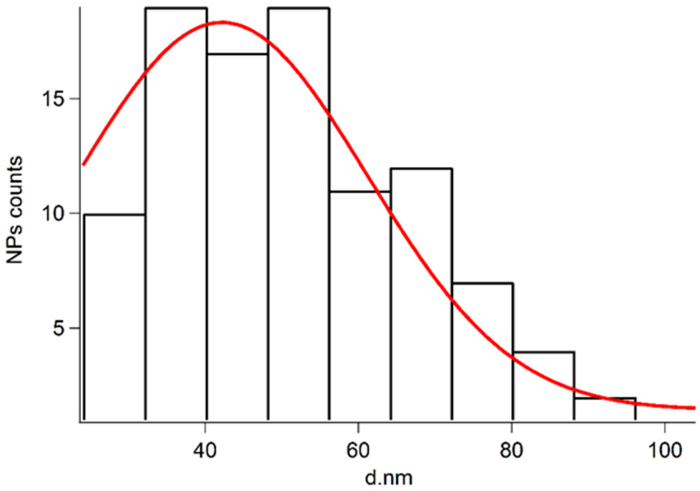


Figure 20: KLVFF(I) 15mM size distribution obtained by Cryo-TEM. The red line represents the Gaussian fitting of the size distribution histogram. The average diameter is 42 ± 26 nm.

Because of the increasing interest in peptide/protein nanoparticles in several fields of biochemistry,[71] we assessed the possibility to get time-stable solutions of **KLVFF(I)**, a fundamental requirement for potential applications being stability. Since 15 mM samples of this peptide form precipitates soon after preparation, more diluted peptide solutions (5 mM) were prepared. As expected, these solutions remained optically transparent for months without forming insoluble aggregates. Cryo-TEM analysis of 5 mM samples (aged two months) confirms that also diluted peptide solutions contain round-shaped NPs (Fig. 22). **KLVFF(I)** NPs size resulted to be concentration-dependent, the average diameter being significantly smaller for the diluted solutions (27 ± 10 nm, Fig. 21). Higher peptide concentration leads to the formation of larger NPs and increased NP polydispersity.

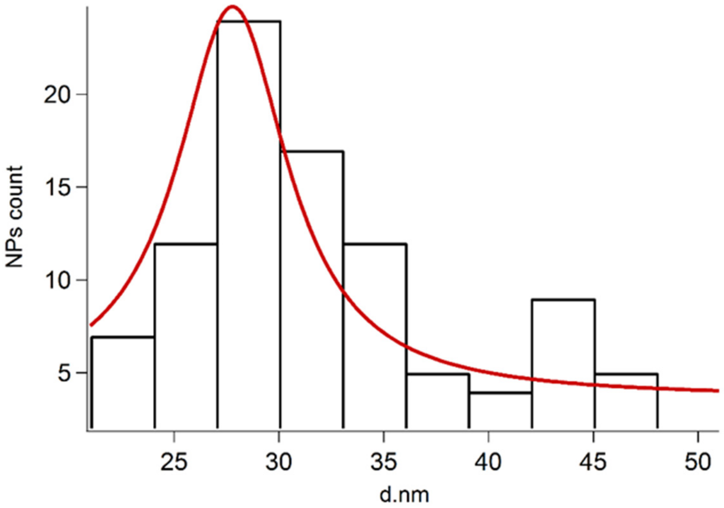


Figure 21: KLVFF(I) 5mM size distribution obtained by Cryo-TEM. The red line represents the Lorentzian fitting of the size distribution histogram. The average diameter is 27 ± 10 nm.

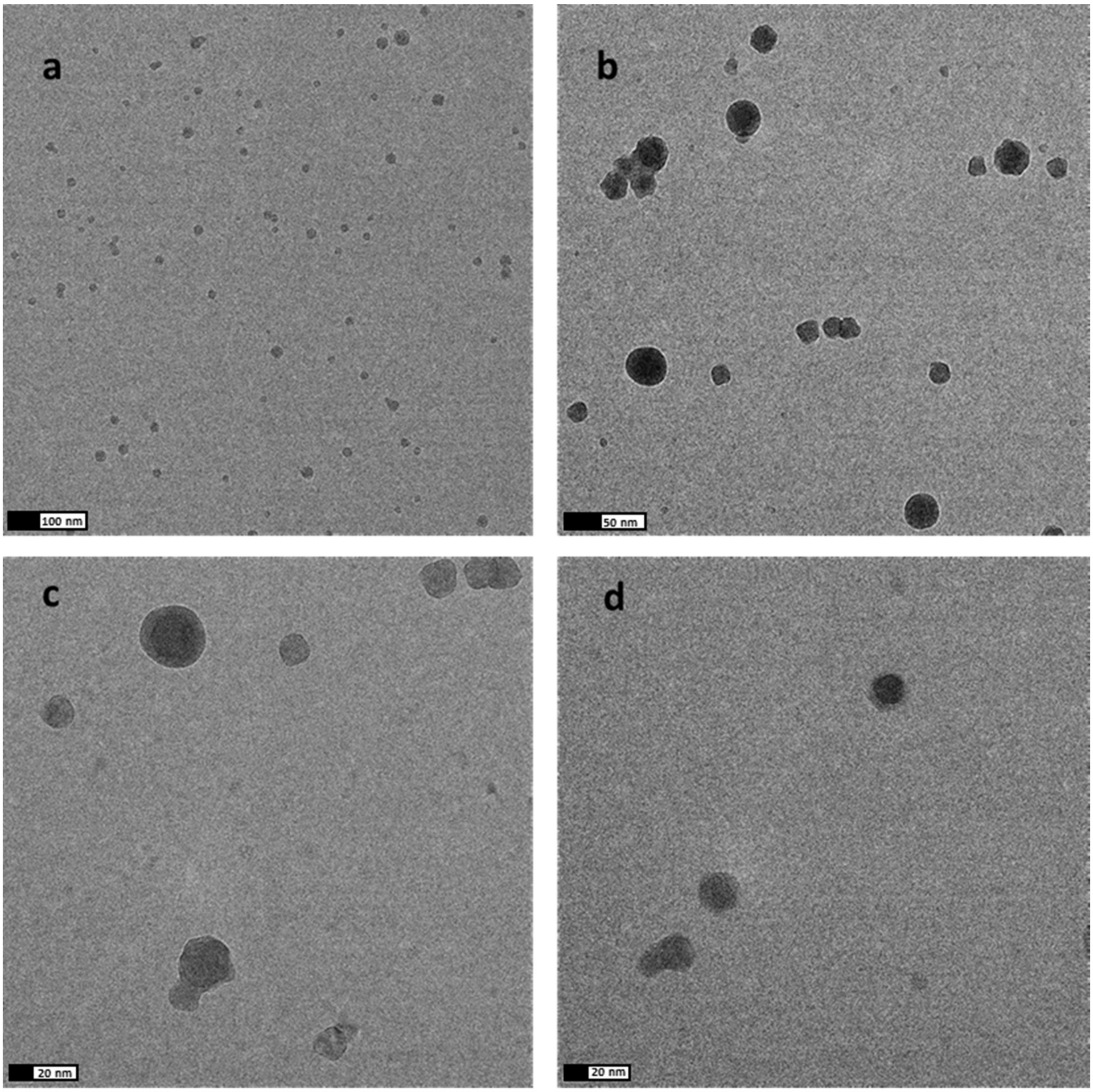
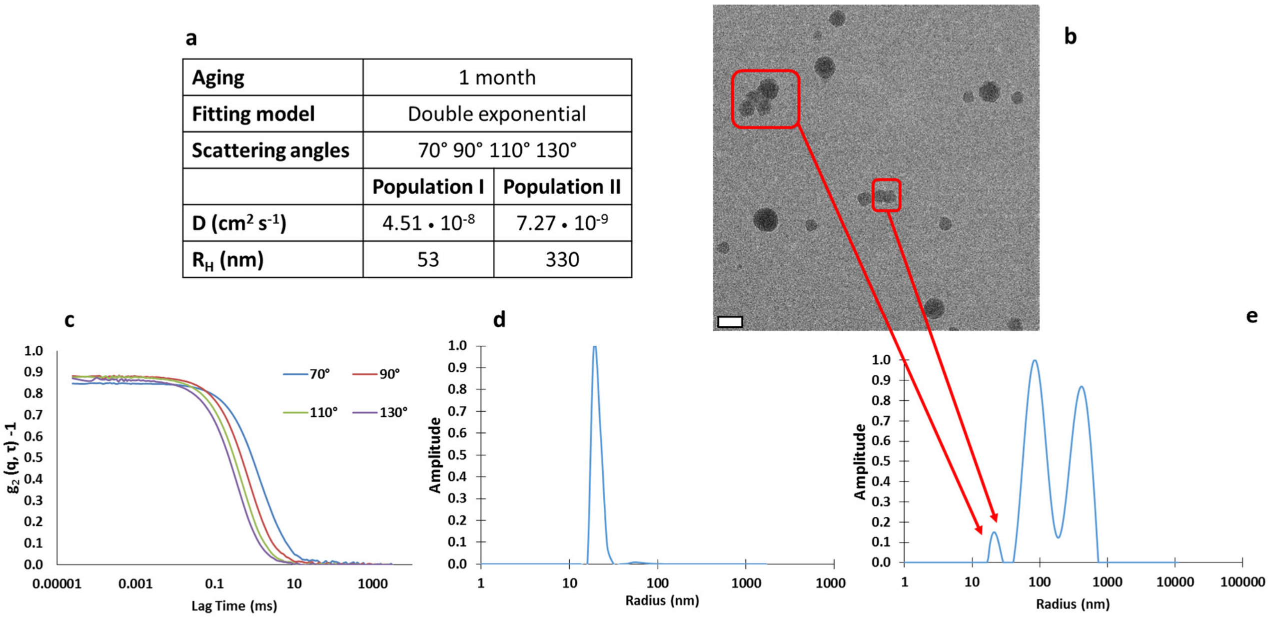


Figure 22: cryo-TEM images of **KLVFF(I)**, 5 mM.

The stability over time of **KLVFF(I)** NPs was assessed through dynamic light scattering[85] (DLS). Peptide solutions at 5 mM concentration were studied at different aging times. The autocorrelation functions associated to four different scattering angles (70°, 90°, 110°, 130°) were analyzed through a double exponential model, indicating the presence of two distinct populations. Multi-angle analysis procedure (see section 3.10) resulted in two different hydrodynamic radii (R_H), correspondent to distinct populations. These populations are remarkably stable over time, as their hydrodynamic radius does not change even after two months. The first one, having an average R_H of 53 ± 6 nm, can be related to the NPs observed with cryo-TEM, although the averaged size is almost two-fold the size obtained by statistical distribution from microscopy (Fig. 21).

The reason of this finding may be two-fold. First, the R_H value calculated through DLS includes the contribution of the solvation shell, thus the calculated radius is reasonably larger than the one obtained with cryo-TEM. Second, the intensity of the scattered light is proportional to the sixth power of the radius for spherical structures;[85] therefore, it is possible that the scattering of small NPs clusters (Fig. 23b) completely overcome the signal associated to isolated NPs. Indeed, reporting the number-weighted size distribution for θ = 90° (Fig. 23 d) the more abundant population (~ 95%) is the smaller one in agreement with cryo-TEM. As the 15 mM dispersion was turbid, it was not suitable for DLS analysis.



Aging	1 month	
Fitting model	Double exponential	
Scattering angles	70° 90° 110° 130°	
	Population I	**Population II**
D (cm^2 s^{-1})	4.51 • 10^{-8}	7.27 • 10^{-9}
R_H (nm)	53	330

Figure 23: DLS analysis for **KLVFF(I)** 5 mM. Table summarizing experimental conditions and fitting results a) Cryo-TEM showing some NPs clusters. Scale bar 50 nm b) Autocorrelation functions associated to the analyzed scattering angles c) Number averaged distribution function calculated with CONTIN at 90° d) Intensity averaged distribution function calculated with CONTIN at 90° e).

The second population, with an average R_H of 330 nm, is not visible with cryo-TEM. Indeed, the maximum thickness required for cryo-TEM samples is 100 nm, and any object beyond this size limit can be confused with the background.[86] In order to confirm the results obtained with the double exponential data fitting, the same autocorrelation functions (Fig. 23c) have been analyzed through Laplace inversion (CONTIN algorithm),[36] that gave comparable results. In fact, two populations were obtained in most cases, although for some angles the second population was very polydisperse and described by two neighboring peaks. These peaks, in the resolution of this technique, can be considered as a single, very broad population (Fig. 23d). Overall, these results show that **KLVFF(I)** solutions form multiple populations, thus further investigation is needed to control the peptide self-assembly and obtain monodisperse spherical particles. Indeed, possible practical applications would require a more homogeneous sample.

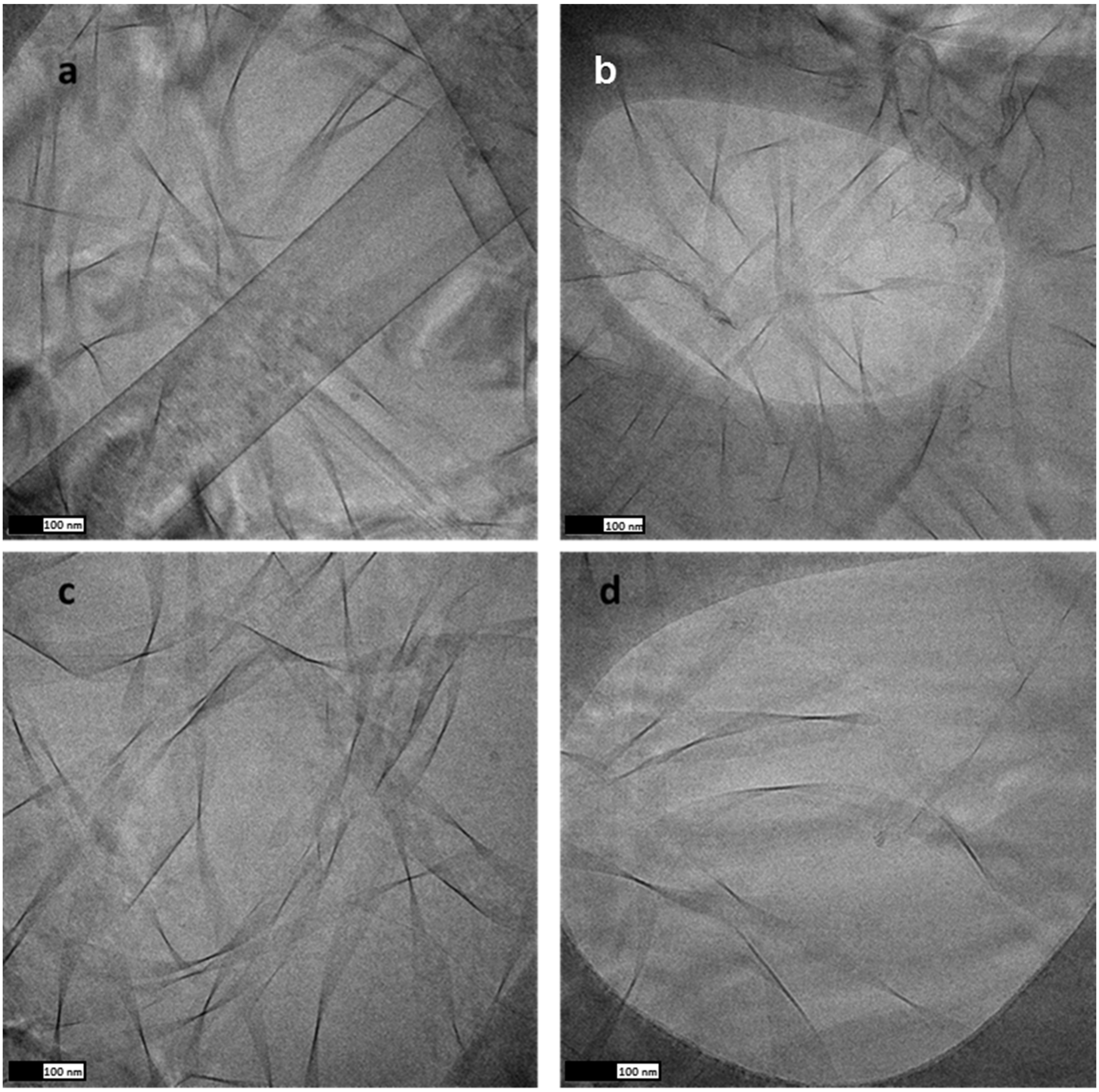


Figure 24: cryo-TEM images of **KLVF(I)F(I)** 15 mM samples aged for two months

Cryo-TEM analysis of **KLVF(I)F(I)** 15 mM samples (aged two months) confirmed that the same nanostructure class observed with TEM in the dried state is present also in solution (Fig. 24). The size of the ribbons visualized by cryo-TEM is reasonably consistent with that observed in the dried state, although the image resolution is lower. As mentioned above, objects exceeding in size the thickness of the sample (100 nm) appear flattened with cryo-TEM,[86] losing three-dimensionality.

The brominated peptide **KLVF(Br)F(Br)** showed an evolving self-assembly over time. Cryo-TEM images of two days-aged 15 mM solutions (the same aging time of TEM samples) showed comparable nanostructures to those visualized in the dried state. As expected, cryo-TEM images are quite confused, since the “cotton balls” formed by this peptide are larger than the maximum sample thickness of 100 nm.[86] Indeed, the visual outcome (Fig. 25a) is a series of shadows resembling the shape of the structures observed through TEM. Cryo-TEM analysis of two months-aged solutions (Fig. 25b, 25c, and 25d) displays a subsequent stage of the self-assembly process. The microscopy images show an entangled network of fibrillary material, which is the kinetic evolution of the “cotton balls”, acting as nucleation centers for fibril growth. Further indication of this morphological change is the physical aspect of the peptide solution that become more viscous upon aging, thus suggesting the formation of a network that is able to trap the solvent.

Overall, these additional studies – aimed to characterize the nanostructures formed by **KLVFF(I)**, **KLVF(I)F(I)** and **KLVF(Br)F(Br)** – further confirm that iodinated derivatives led to more stable assemblies. Indeed, the structures formed by **KLVFF(I)** and **KLVF(I)F(I)** do not change over time, while **KLVF(Br)F(Br)** aggregates display a time-dependent evolution of their morphology. This finding again suggests an active role of the halogen atoms in the self-assembly process. In particular, among the multiplicity of non-covalent interactions driving the aggregation behavior of these peptides, the contribution of halogen bonding cannot be overlooked, especially when the molecule is modified with halogen atoms with a marked electrophilic behavior (as in the case of iodine).

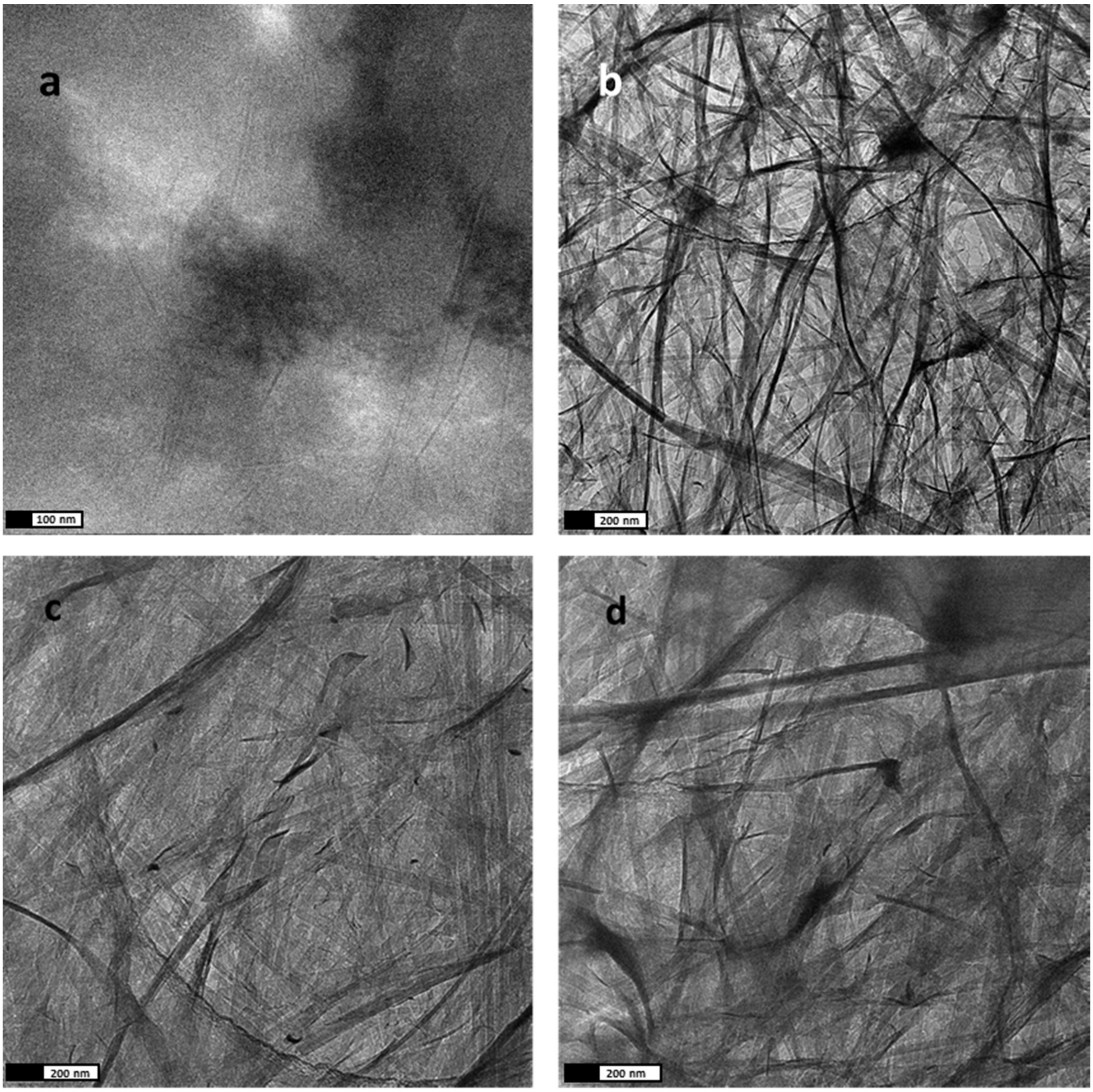


Figure 25: cryo-TEM images of **KLVF(Br)F(Br)** 15 mM. a) 48 h after sample preparation. b) c) d) two months after sample preparation.

3.4.5 The role of halogenation

The self-assembly of peptide amphiphiles is usually controlled by the balance between energies associated to hydrogen bonding, hydrophobic and electrostatic interactions. In the case of halogenated KLVFF derivatives, however, various findings suggest that, in addition to the above mentioned interactions, a direct contribution of interactions involving halogen atoms cannot be excluded. In fact, the particular nature, position and the number of halogen atoms in the peptide sequence clearly affect the self-assembly process, leading to different nanostructures and suggesting that halogenated residues may be directly involved. Rheological experiments suggest that gel strength depends on halogen atom polarizability, as confirmed by the fact that **KLVF(I)F** forms stronger gels than **KLVF(Br)F**. Considering microscopy analysis, di-chlorinated and di-brominated peptides **KLVF(Cl)F(Cl)** and **KLVF(Br)F(Br)** showed less propensity to assemble into well-defined supramolecular architectures than **KLVF(I)F(I)**, suggesting again that halogen atom polarizability affects the complexity of the obtained nanostructures. The polarizability dependence allows reducing the relevance of π-π interactions in the self-assembly of these halogenated derivatives, as they cannot justify the observed trend. Hydrophobic interactions should not be considered as the unique driving force of the self-assembly, since compounds having similar hydrophobicity (Table 5), like **KLVF(I)F** – **KLVFF(I)** and **KLVF(Br)F** – **KLVFF(Br)**, show quite different properties. The obtained crystal structures of the di-halogenated peptides directly demonstrate that, when highly polarizable halogen atoms are present, halogen bonding may occur giving its contribution in coexistence with other more usual interactions. All of the above mentioned findings support the hypothesis that halogen bonding plays a direct role in fine tuning the self-assembly of the studied halogenated peptides.

Table 5: Retention time of studied peptides in reverse-phase HPLC.

Peptide	Retention time (min)
KLVFF	20.14
KLVFF(I)	22.46
KLVF(I)F	22.56
KLVF(I)F(I)	24.38
KLVFF(Br)	22.02
KLVF(Br)F	22.42
KLVF(Br)F(Br)	23.39
KLVF(Cl)F(Cl)	22.73

3.5 Conclusions

We have demonstrated halogenation is a powerful strategy to tune the type of self-assembly of the same amino acidic sequence. Our results demonstrate the importance of halogen position, nature and number in determining the folding mode of the KLVFF peptide. Considering halogenation is a single point mutation, the variety of assemblies that is possible to obtain is impressive. In general placing a halogen on the position 4 (central phenylalanine) increases the propensity to form amyloid fibrils, while a halogen on the position 5 (terminal phenylalanine) causes peptide aggregation (in some cases into spherical NPs) but not fibrillation. Di-halogenation – in particular di-iodination, leads to structures with a higher grade of complexity. This work, thanks to the evidence of the role of halogen atoms in driving the peptide folding through non-covalent interactions, enriches the structural landscape for peptide self-assembly.

3.6 References

(1) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75* (1), 333.

(2) Hauser, C. A. E.; Maurer-Stroh, S.; Martins, I. C. *Chem. Soc. Rev.* **2014**, *43* (15), 5326.

(3) Mankar, S.; Anoop, A.; Sen, S.; Maji, S. K. *Nano Rev.* **2011**, *2*, 1.

(4) Bertolani, A.; Pirrie, L.; Houbenov, N.; Haataja, J.; Stefan, L.; Catalano, L.; Terraneo, G.; Giancane, G.; Valli, L.; Milani, R.; Ikkala, O.; Resnati, G.; Metrangolo, P. *Nat. Commun.* **2015**, *6* (7574), 1.

(5) Tjernberg, L. O.; Näslund, J.; Lindqvist, F.; Johansson, J.; Karlström, a R.; Thyberg, J.; Terenius, L.; Nordstedt, C. *J. Biol. Chem.* **1996**, *271* (15), 8545.

(6) Tjernberg, L. O.; Tjernberg, A.; Bark, N.; Shi, Y.; Ruzsicska, B. P.; R, Z. B. U.; Thyberg, J.; Callaway, D. J. E. *Biochem. J.* **2002**, *366*, 343.

(7) Tjernberg, L. O.; Callaway, D. J. E.; Tjernberg, A.; Hahne, S.; Lilliehook, C.; Terenius, L.; Thyberg, J.; Nordstedt, C. *J. Biol. Chem.* **1999**, *274* (18), 12619.

(8) Krysmann, M. J.; Castelletto, V.; Kelarakis, A.; Hamley, I. W.; Hule, R. A.; Pochan, D. J. *Biochemistry* **2008**, 4597.

(9) Funke, S. A.; Willbold, D. *Curr. Pharm. Des.* **2012**, 755.

(10) Hilbich, C.; Kisters-woike, B.; Reed, J.; Masters, C. L.; Beyreuther, K. *J. Mol. Biol.* **1992**, *228*, 460.

(11) Kallberg, Y.; Gustafsson, M.; Persson, B.; Thyberg, J.; Johansson, J. *J. Biol. Chem.* **2001**, *276* (16), 12945.

(12) Pawar, A. P.; DuBay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. *J. Mol. Biol.* **2005**, *350* (2), 379.

(13) Kirschner, D. A.; Abraham, C.; Selkoe, D. J. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83* (2), 503.

(14) Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (26), 16742.

(15) Cherny, I.; Gazit, E. *Angew. Chem. Int. Ed.* **2008**, *47* (22), 4062.

(16) Gazit, E. *Chem. Soc. Rev.* **2007**, *36* (8), 1263.

(17) Reches, M.; Gazit, E. *Science* **2003**, *300* (5619), 625.

(18) Reches, M.; Gazit, E. *Phys. Biol.* **2006**, *3* (1), S10.

(19) Frederix, P. W. J. M.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V; Tuttle, T. *Nat. Chem.* **2014**, *7* (1), 30.

(20) Jones, O. G.; Mezzenga, R. *Soft Matter* **2012**, *8*, 876.

(21) Burai, R.; Ait-Bouziad, N.; Chiki, A.; Lashuel, H. A. *J. Am. Chem. Soc.* **2015**, *137* (15), 5041.

(22) Lausi, A.; Polentarutti, M.; Onesti, S.; Plaisier, J. R.; Busetto, E.; Bais, G.; Barba, L.; Cassetta, A.; Campi, G.; Lamba, D.; Pifferi, A.; Mande, S. C.; Sarma, D. D.; Sharma, S. M.; Paolucci, G. *Eur. Phys. J. Plus* **2015**, *130* (3), 43.

(23) Kabsch, W. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66* (2), 125.

(24) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2011**, *67* (4), 235.

(25) Evans, P. R.; Murshudov, G. N. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (7), 1204.

(26) Zwart, P. H.; Banumathi, S.; Dauter, M.; Dauter, Z. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (11), 1958.

(27) McGeehan, J. E.; Carpentier, P.; Royant, A.; Bourgeois, D.; Ravelli, R. B. G. *J. Synchrotron Radiat.* **2007**, *14* (1), 99.

(28) Sheldrick, G. M. SADABS, University of Göttingen, Germany 2012,.

(29) Sheldrick, G. M. *Acta Crystallogr. Sect. A Found. Crystallogr.* **2015**, *71* (1), 3.

(30) Sheldrick, G. M. *Acta Crystallogr. Sect. C Struct. Chem.* **2015**, *71* (Md), 3.

(31) Emsley, P.; Cowtan, K. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2004**, *60* (12 I), 2126.

(32) Brunger, A. . T. *Nature* **1992**, *355*, 472.

(33) Parsons, S.; Flack, H. D.; Wagner, T. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2013**, *69* (3), 249.

(34) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; Van De Streek, J.; Wood, P. A. *J. Appl. Crystallogr.* **2008**, *41* (2), 466.

(35) Delano, W. L. "The PyMOL Molecular Graphics System", http://www.pymol.org 2002,.

(36) Provencher, W. *Comput. Phys. Commun.* **1982**, *27*, 229.

(37) Cui, H.; Cheetham, A. G.; Pashuck, E. T.; Stupp, S. I. *J Am Chem Soc* **2014**, *136*, 12461.

(38) Metrangolo, P.; Pilati, T.; Resnati, G. *CrystEngComm* **2006**, *8*, 946.

(39) Desiraju, G. R.; Ho, P. S.; Kloo, L.; Legon, A. C.; Marquardt, R.; Metrangolo, P.; Politzer, P.; Resnati, G.; Rissanen, K. *Pure Appl. Chem.* **2013**, *85* (8), 1711.

(40) Murray, J. S.; Lane, P.; Clark, T.; Riley, K. E.; Politzer, P. *J. Mol. Model.* **2012**, *18* (2), 541.

(41) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435* (7043), 773.

(42) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg, D. *Nature* **2007**, *447* (7143), 453.

(43) Ray, S.; Das, A. K.; Drew, M. G. B.; Banerjee, A. *Chem. Commun.* **2006**, *1* (40), 4230.

(44) Maity, S.; Kumar, P.; Haldar, D. *Org. Biomol. Chem.* **2011**, *9* (10), 3787.

(45) Stroud, J. C. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (4), 540.

(46) Reches, M.; Porat, Y.; Gazit, E. *J. Biol. Chem.* **2002**, *277* (38), 35475.

(47) Cavallo, G.; Metrangolo, P.; Milani, R.; Pilati, T.; Priimägi, A.; Resnati, G.; Terraneo, G. *Chem. Rev.* **2016**, *116* (4), 2478.

(48) Metrangolo, P.; Meyer, F.; Pilati, T.; Resnati, G.; Terraneo, G. *Angew. Chem. Int. Ed.* **2008**, *47* (33), 6114.

(49) Vasylyeva, V.; Nayak, S. K.; Terraneo, G.; Cavallo, G.; Metrangolo, P.; Resnati, G. *Crystengcomm* **2014**, *16* (35), 8102.

(50) Voth, A. R.; Khuu, P.; Oishi, K.; Ho, P. S. *Nat. Chem.* **2009**, *1* (1), 74.

(51) Colletier, J.-P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A. B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16938.

(52) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (48), 16789.

(53) Görbitz, C. H. *Chem. Comm.* **2006**, No. 22, 2332.

(54) Steed, J. W. *Chem. Commun.* **2011**, *47* (5), 1379.

(55) Picout, D. R.; Ross-Murphy, S. B. *Sci. World J.* **2003**, *3*, 105.

(56) Hauser, C. a E.; Deng, R.; Mishra, A.; Loo, Y.; Khoe, U.; Zhuang, F.; Cheong, D. W.; Accardo, A.; Sullivan, M. B.; Riekel, C.; Ying, J. Y.; Hauser, U. a. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (4), 1361.

(57) Mishra, A.; Loo, Y.; Deng, R.; Chuah, Y. J.; Hee, H. T.; Ying, J. Y.; Hauser, C. A. E. *Nano Today* **2011**, *6* (3), 232.

(58) Smith, A. M.; Williams, R. J.; Tang, C.; Coppo, P.; Collins, R. F.; Turner, M. L.; Saiani, A.; Ulijn, R. V. *Adv. Mater.* **2008**, *20* (1), 37.

(59) Webb, R. H. *Reports Prog. Phys.* **1996**, *59* (3), 427.

(60) Nilsson, M. R. *Methods* **2004**, *34* (1), 151.

(61) Shivu, B.; Seshadri, S.; Li, J.; Oberg, K. A.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2013**, *52*, 5176.

(62) Zandomeneghi, G.; Krebs, M. R. H.; McCammon, M. G.; Fändrich, M. *Protein Sci.* **2004**, *13* (12), 3314.

(63) Kong, J.; Yu, S. *Acta Biochim. Biophys. Sin.* **2007**, *39* (8), 549.

(64) Corrêa, D.; Ramos, C. *African J Biochem Res* **2009**, *3* (5), 164.

(65) Chen, Y. H.; Yang, J. T.; Chau, K. H. *Biochemistry* **1974**, *13* (16), 3350.

(66) Ho, B. K.; Curmi, P. M. G. *J. Mol. Biol.* **2002**, *317* (2), 291.

(67) Micsonai, A.; Wien, F.; Kernya, L.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (24), E3095.

(68) Krysmann, M. J.; Castelletto, V.; Hamley I. W. Soft Matter **2007**, *3*, 1401.

(69) Castelletto, V.; Hamley I. W. *Biophys. Chem.* **2009**, *141*, 169.

(70) Gupta, M.; Bagaria, A.; Mishra, A.; Mathur, P.; Basu, A.; Ramakumar, S.; Singh Chauhan, V. *Adv. Mater.* **2007**, *19*, 858.

(71) Doll, T. A. P. F.; Dey, R.; Burkhard, P. *J. Nanobiotechnology* **2015**, *13* (1), 73.

(72) Vauthey, S.; Santoso, S.; Gong, H.; Watson, N.; Zhang, S. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (8), 5355.

(73) Cui, H.; Webber, M. J.; Stupp, S. I. *Biopolymers* **2007**, *94*, 1.

(74) Habibi, N.; Kamaly, N.; Memic, A.; Shafiee, H. *Nano Today* **2016**, *11* (1), 41.

(75) Kumaraswamy, P.; Sethuraman, S.; Krishnan, U. M. *Soft Matter* **2013**, *9* (9), 2684.

(76) Velichko, Y. S.; Stupp, S. I.; Olvera, M.; Cruz, D. *J. Phys. Chem. B* **2008**, *112*, 2326.

(77) Rad-Malekshahi, M.; Visscher, K. M.; Rodrigues, J. P. G. L. M.; De Vries, R.; Hennink, W. E.; Baldus, M.; Bonvin, A. M. J. J.; Mastrobattista, E.; Weingarth, M. *J. Am. Chem. Soc.* **2015**, *137* (24), 7775.

(78) Gudlur, S.; Sukthankar, P.; Gao, J.; Avila, L. A.; Hiromasa, Y.; Chen, J.; Iwamoto, T.; Tomich, J. M. *PLoS One* **2012**, *7* (9), 1.

(79) Berl, V.; Schmutz, M.; Krische, M.; Khoury, R.; Lehn, J.-M. *Chem. Eur. J.* **2002**, *8* (5), 1227.

(80) Aggeli, A.; Nyrkova, I. A.; Bell, M.; Harding, R.; Carrick, L.; McLeish, T. C. B.; Semenov, A. N.; Boden, N. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (21), 11857.

(81) Gosal, W. S.; Morten, I. J.; Hewitt, E. W.; Smith, D. A.; Thomson, N. H.; Radford, S. E. *J. Mol. Biol.* **2005**, *351* (4), 850.

(82) Jiménez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V; Dobson, C. M.; Saibil, H. R. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (14), 9196.

(83) Caughey, B.; Peter T. Lansbury, J. *Annu. Rev. Neurosci.* **2003**, *26* (1), 267.

(84) Chamberlain, A. K.; MacPhee, C. E.; Zurdo, J.; Morozova-Roche, L. A.; Hill, H. A. O.; Dobson, C. M.; Davis, J. J. *Biophys. J.* **2000**, *79* (6), 3282.

(85) Berne, B. J.; Pecora, R. *Dynamic light scattering: with applications to chemistry, biology, and physics*; Courier Corporation, 1976.

(86) Glaeser, R. M. *Nat Methods* **2016**, *13* (1), 28.

Chapter 4

Crystal structures of halogenated derivatives of KLVFF as a model for Aβ self-assembly

4.1 Introduction

The aggregation of soluble peptides or proteins into insoluble nanomaterials is – together with DNA base pairing – the self-assembly process having the bigger impact in life. Amyloid fibrils are the result of this process and their occurrence is associated to several neurodegenerative pathologies including Alzheimer's, Parkinson's and prion diseases.[1] Because of their implication in these neurological disorders, amyloid fibrils are have been studied for a long time in order to enlighten their complex mechanism of formation. The best way to obtain deep knowledge of this process consists in considering the noncovalent interactions leading to the formation of the amyloid structure. As a result, the structural determination of amyloid fibrils through single-crystal X-ray diffraction (XRD) has a central role.[2] Due to their low crystallinity, to get good crystals of amyloidogenic protein samples is a great challenge. For this reason several structural studies focused on shorter amino acidic sequences, derived from full-length amyloid-forming proteins.[3] Often, these peptide fragments represent a reliable model of the full-length protein behavior, since they contain the key residues required for fibrillation.[4] The crystal structure of short amyloidogenic segments highlights the interactions driving the self-assembly of monomers into oligomers and mature fibrils, suggesting new hints for a full understanding of the fibrillation process. The direct relation between crystal structures and fibrillar material is still an open controversy, although it has been already reported that different peptide fragments form microcrystals and fibrils under the same experimental conditions.[2] The pentapeptide KLVFF (H_2N-Lys-Leu-Val-Phe-Phe-COOH) is a small portion of the peptide Aβ40, which spontaneous assembly into amyloid fibrils is associated to Alzheimer's disease.[5] This fragment ($Aβ_{16\text{-}20}$) is a well-studied model for amyloid formation, since both experimental and modelling studies[6–8] suggest it is critical for the fibrillation of the full-length Aβ peptide. However, due to its low crystallinity, high resolution structural details of KLVFF are missing. In fact, only three polymorphs of the fragment $Aβ_{16\text{-}21}$ (PDB_{ID} 2Y2A, 2Y29, 3OW9) were determined at low resolution (the best one being 1.8 Å) by using high energy sources like Synchrotron radiation.[3] Here we report the halogenation in the *para* position of the terminal phenylalanine[9,10] residue of KLVFF (Fig. 1) as a strategy to enhance the propensity of this segment to form crystals.[11] The crystal structure of this fragment has been determined for the first time with excellent resolution (less than 0.9 Å) for both iodinated and brominated derivatives; XRD data were collected in house, using a conventional X-ray source.

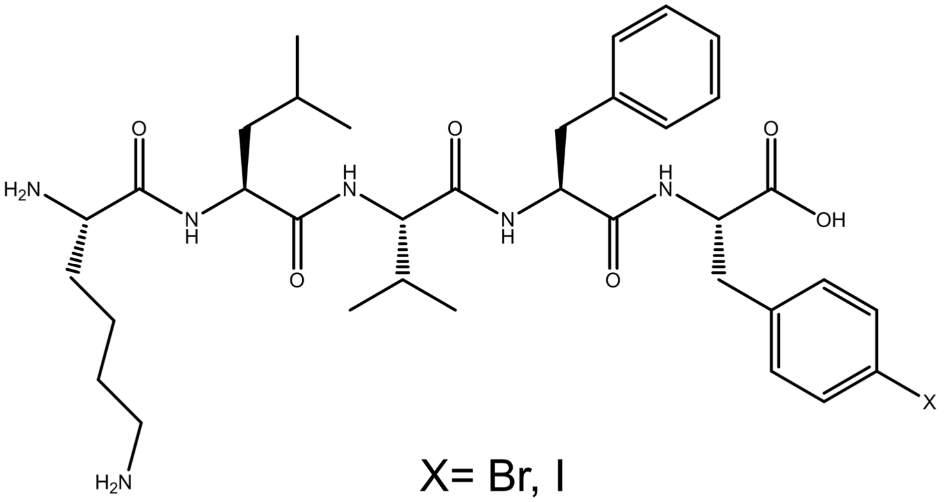


Figure 1: Chemical structure of KLVFF derivatives modified with a halogen on the *p*- position of the terminal phenylalanine residue. The iodinated peptide is referred as **KLVFF(I)**, the brominated one as **KLVFF(Br)**

4.2 Materials

Peptides with confirmed amino acid analysis (purity ≥ 98%), were purchased from Biopeptek (Malvern, USA). The integrity of all peptides was confirmed by ion spray mass spectrometry and the purity was determined by reverse phase high-pressure liquid chromatography (HPLC).

4.3 Methods

4.3.1 Infrared Spectroscopy (FT-IR). Infrared spectra were recorded at room temperature using a Nicolet iS50 FT-IR spectrometer equipped with a DTGS detector. Peptides were analyzed in bulk without any previous treatment. Spectra represent an average of 64 scans recorded in a single beam mode with a 4 cm^{-1} resolution and corrected for the background. The second derivative analyses of the spectra were performed using the Nicolet FTIR software, Omnic 9.0®, with a 13-point and 3rd polynomial order Savitzky and Golay function. Second derivative spectra generated negative bands as compared with the original spectra, thus for comparison all the second-derivative spectra were multiplied by -1.

4.3.2 Peptides crystallization

H_2N-Lys-Leu-Val-(*p*-IodoPhe)-(*p*-IodoPhe)-COOH *i.e.* **KLVFF(I)** was obtained as solvated species by dissolving the peptide (5 mg/ml) in water. Crystals suitable for XRD analysis were obtained after twelve months of slow evaporation.

H_2N-Lys-Leu-Val-(*p*-BromoPhe)-(*p*-BromoPhe)-COOH *i.e.* **KLVFF(Br)** was obtained as solvated species by dissolving the peptide (5 mg/ml) in a mixture of water and 0.2 M sodium tartrate (1:1 ratio). Crystals suitable for XRD analysis were obtained after two months of vapor diffusion in the presence of 0.2 M sodium tartrate as reservoir.

4.3.3 X-ray crystallography data acquisition. X-ray diffraction data were collected at Bruker APEX-II diffractometer equipped with micro-focus and CCD detector, using Mo-$K\alpha$ radiation (λ=0.71073 Å). The crystal was cryo-cooled (103 K) for data collection (Bruker KRYOFLEX). The structure was solved by *SIR2002*[12] and refinements were carried out by full-matrix least-squares on F^2 using the *SHELXL*[13] program.

For **KLVFF(I)** the final set of data was comprised of 13097 unique reflections from a total of 15681 (with $I_o > 2\sigma(I_o)$ and $R_{ave} = 0.042$). All atoms were anisotropically refined, leading to a data/parameters ratio about 14 – 11 considering only the data with $I_o > 2\sigma(I_o)$ – from a total number of 1092 parameters. To decrease the correlations in the refinement some restraints on both geometrical and displacement parameters were imposed, in particular in the case of disordered moieties. All the H atoms of the peptide were placed at calculated positions, while those of the water molecules (partially derived from the difference map and partially calculated based on the shortest O⋯O contacts) were refined with strong geometric restraints. The hydrogen atoms of disordered water molecules were not modeled. The main crystallographic data are reported in Table 1.

For **KLVFF(Br)** the final set of data was comprised of 12794 unique reflections from a total of 9550 (with $I_o > 2\sigma(I_o)$ and $R_{ave} = 0.032$). All atoms were anisotropically refined, leading to a data/parameters ratio about 12 – 9 considering only the data with $I_o > 2\sigma(I_o)$ – from a total number of 1033 parameters. To decrease the correlations in the refinement some restraints on both geometrical and displacement parameters were imposed. All the H atoms ot the peptide were placed at calculated positions, while those of the water molecules (partially derived from the difference map and partially calculated based on the shortest O⋯O contacts) were refined with strong geometric restraints. The main crystallographic data are reported in Table 2.

4.4 Results and discussion

The use of “heavy” atoms in XRD is a well-known strategy to facilitate structure determination;[14,15] introducing atoms having remarkable electron density gives an appreciable contribution to the overall scattered intensity of the molecule. Since the contribution of each atom to the scattered intensity is proportional to the square of the atomic number, halogens like bromine and iodine make a difference in XRD experiments. *p*-iodophenylalanine and *p*-bromophenylalanine are commonly included in biomolecules in order to determine the X-ray crystallographic phases.[16–19] Longer peptide segments derived from Aβ and containing the KLVFF motif have been determined using this strategy; in these studies the self-assembly of the peptide is rationalized by modifications that lead to the formation of macrocyclic structures. Although halogenation is a point mutation, we decided to confine iodine and bromine at the edge of the amino acid sequence, as in the previous chapter we demonstrated this is a minor effect on the self-assembly behavior of the peptide segment.

4.4.1 A reliable model for the wild-type sequence KLVFF

The iodinated and brominated derivatives of KLVFF showed lower solubility in water compared to the native sequence, making their precipitation – *i.e.* their crystallization – easier to control. The wild type fragment KLVFF is reported to form amyloid hydrogels over a minimum gelation concentration,[5] while our halogen-modified peptides tend to precipitate without forming the typical entangled network of a gel. In order to have a preliminary confirmation that halogenation on the last phenylalanine residue (Phe5) does not modify the amyloid nature of the peptide, the FT-IR spectra of **KLVFF(I)** and **KLVFF(Br)** in bulk were recorded first (Fig. 2).

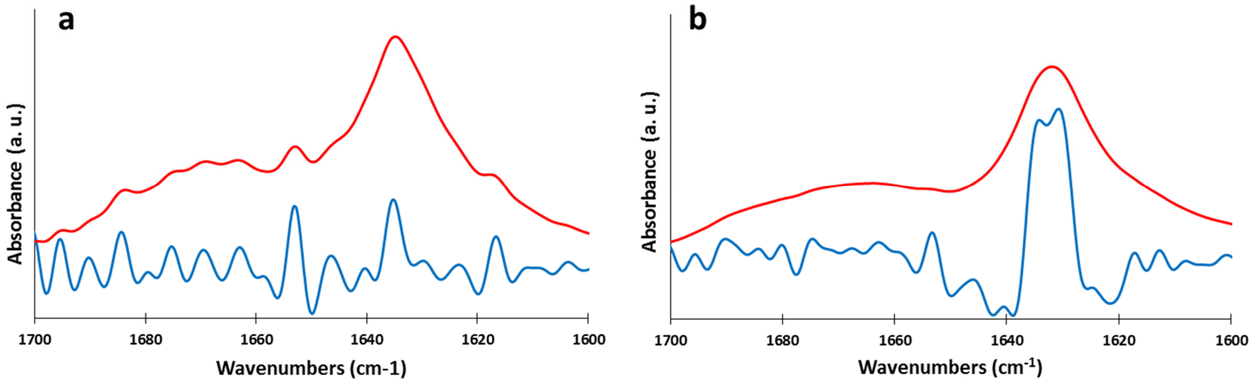


Figure 2: FTIR spectroscopy of **KLVFF(I)** a) and **KLVFF(Br)** b) in bulk. The blue lines represent the second derivative of the FTIR spectra.

We focused our attention mainly on the amide I band (1600-1700 cm^{-1}), representing the stretching vibration of the peptide carbonyl groups.[20] This band is directly related to the peptide backbone conformation, hence it gives information about the self-assembly behavior of the compound. In this interval, both spectra present an intense peak centered in the range 1630-1635 cm^{-1}; this signal can be associated to β-sheet,[21] the prevalent secondary structure constituting amyloid fibrils. This finding confirms that the halogen-modified KLVFF peptides preserve the propensity to assemble in fibrillary material, thus it makes sense to consider **KLVFF(I)** and **KLVFF(Br)** as reliable models of the native sequence. Since the crystal structure of the wild-type pentapeptide has never been determined at high resolution, a direct comparison assessing the similarity between KLVFF and our halogenated derivatives would be possible only through computational studies, like in a recent work focused on the iodination of the core sequence of human calcitonin (DFNKF).[22]

4.4.2 Crystal structure of KLVFF(I) and KLVFF(Br)

KLVFF(I) was successfully crystallized from water by slow evaporation of the solvent. Crystals suitable for single crystals X-ray diffraction formed within twelve months. High-resolution (0.81 Å) diffraction data were collected using Mo-Kα radiation on a Bruker KAPPA APEX II. **KLVFF(I)** crystallizes in the triclinic P1 space group. The asymmetric unit (Fig. 3) consists of two peptide molecules solvated by four water molecules and two trifluoroacetate ions.

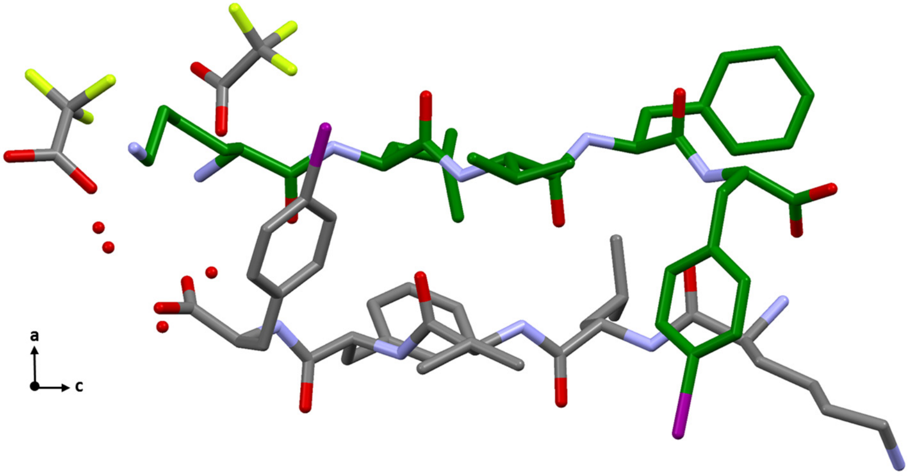


Figure 3: Crystal structure of **KLVFF(I)**. Asymmetric unit consisting of two peptide molecule solvated by four water molecules and two molecules of trifluoroacetate. Color code: C, grey, green; O, red; N, violet; I, purple; F, yellow. Hydrogen atoms have been omitted for clarity.

The peptide monomers form extended strands hydrogen-bonded to each other, resulting in antiparallel β-sheets, which axes run parallel to the crystallographic axis *a*. Each KLVFF(I) molecule forms eight N-H···O hydrogen bonds with neighboring strands belonging to the same β-sheet. The distance between each strand in the β-sheet is 4.43 Å (calculated considering the centroids of two adjacent monomers). The C-terminus gives two additional hydrogen bonds with the charged group of the lysine side-chain and the N-terminus of paired peptide strands, conferring additional strength to the β-sheet.

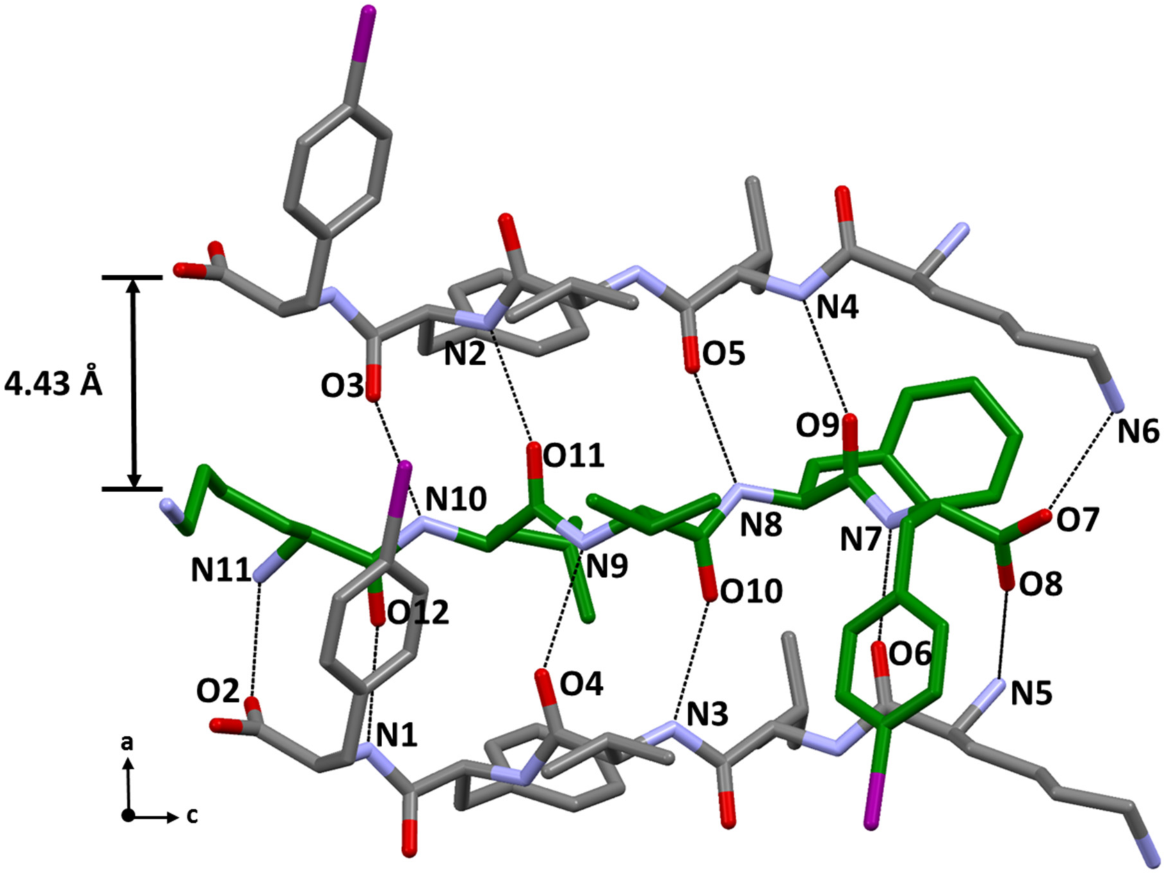


Figure 4: Crystal structure of **KLVFF(I)**. Hydrogen bonding contacts in a parallel β-sheet (Distances: N1H···O12 2.97 Å N2H···O11 2.76 Å N3H···O10 2.90 Å N4H···O9 2.92 Å N5H···O8 2.76 Å N6H···O7 2.74 Å N8H···O5 2.88 Å N9H···O4 2.90 Å N10H···O3 3.03 Å N11H···O2 2.85 Å). Color code: C, grey, green; O, red; N, violet; I, purple. Hydrogen atoms have been omitted for clarity.

The lateral self-assembly of the peptide (Fig. 5) is driven by electrostatic interactions; along the crystallographic axis *c,* the charged termini of a peptide monomer form salt bridges with the charged groups of adjacent peptide monomers. Indeed the C-terminus interacts with the lysine side chain of a neighboring fragment (N6H···O1: 2.84 Å, N12H···O7: 2.73 Å). Along the same direction, the crystal packing is further stabilized by interaction between the N-terminus and the carboxylic oxygen of a trifluoroacetate ion, bridging two different peptide fragments (N5H···O16: 2.79 Å; N11H···O15: 2.91 Å).

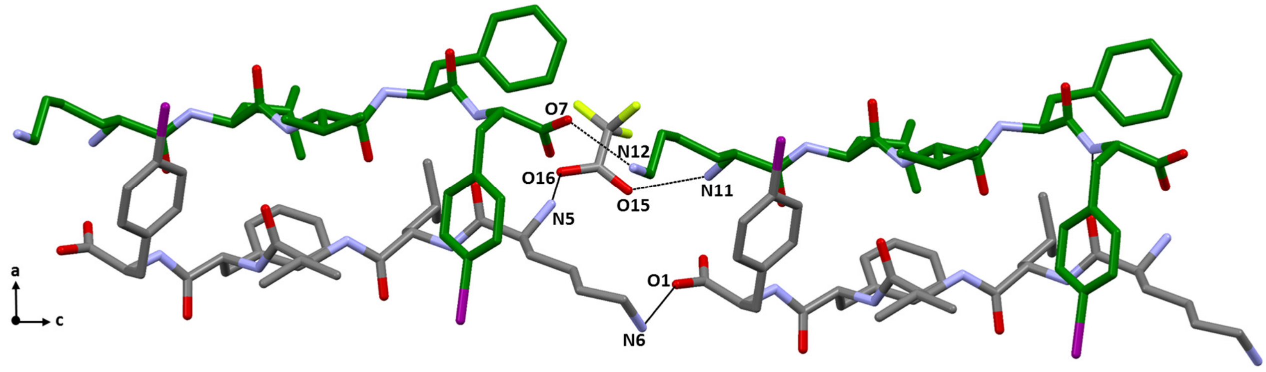


Figure 5: Crystal structure of **KLVFF(I)**. Electrostatic interactions driving the lateral self-assembly of the peptide. (Distances: N5H···O16 2.79 Å N6H···O1 2.84 Å N11H···O15 2.91 Å N12H···O7 2.73 Å). Color code: C, grey, green; O, red; N, violet; I, purple; F, yellow. Hydrogen atoms have been omitted for clarity.

Along the crystallographic axis *b*, paired β-sheets form the typical cross-β spine structure, which is a common peculiarity of different amyloidogenic sequences.[23] The 'spine' consists of a pair of β-sheets running parallel to the fibril axis (here the fibril axis coincides with the crystallographic axis *a*). The facing side-chains of paired β-sheets form a self-complementary dry interface, commonly defined as 'steric zipper'.[24] This motif recurs in all amyloid structures and gives information that can elucidate the mechanism of aggregation of monomers into oligomers and fibrils. Each amyloidogenic segment has a peculiar 'steric zipper',[25] resulting from the non-covalent interactions (typically weak hydrophobic interactions) involving the interdigitated amino acid side chains of facing β-sheets. The 'steric zippers' can be distinguished in 'homosteric' (consisting of β-strands of a single sequence, as in our case) and 'heterosteric' (the β-strands have more than one sequence). Sawaya *et al.*[2] classified the known 'homosteric zippers' considering three different features: (I) orientation of the β-strands in the same β-sheet (parallel or antiparallel); (II) orientation of the paired β-sheet (parallel or antiparallel) with respect to one another; (III) packing of the paired β-sheet (identical or opposite surfaces of the facing β-sheets creating the zipper interface). Considering these criteria, the steric zipper associated to **KLVFF(I)** can be classified as UP-UP—FACE=BACK—CLASS 7. Indeed a single β-sheet has antiparallel orientation, the β-sheets forming the spine run parallel to each other and opposite faces of the paired β-sheets generate the zipper interface. The steric zipper of the iodinated peptide is the same of the one reported for KLVFFA.[3]

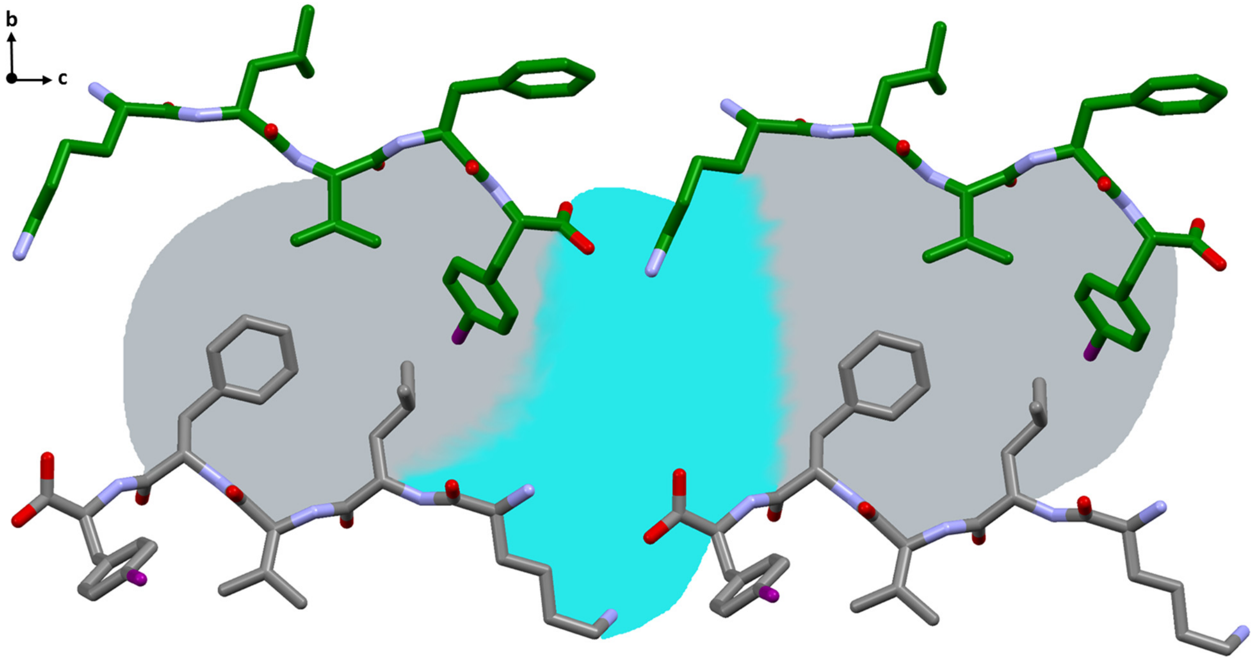


Figure 6: Steric zipper of **KLVFF(I)**. Pairing of the β-sheets into the steric zippers creates the dry (grey) and wet (sky blue) interfaces. Color code: C, grey, green; O, red; N, violet; I, purple. Hydrogen atoms have been omitted for clarity.

The separation of facing **KLVFF(I)** β-sheets is 10.49 Å (distance between centroids of monomers belonging to different β-sheets); this distance coincides with the crystallographic axis *b*. It is possible to distinguish two different interfaces between sheets, named dry and wet interfaces (Fig. 6). The wet interface of **KLVFF(I)** contains water molecules and trifluoroacetate ions, which are stabilized through a network of hydrogen bonds involving the charged termini and the lysine side chain of the peptide strands. In contrast, the dry interface does not contain water molecules and the amino acidic side chains of one β-sheet are tightly interdigitated with the residues of its interacting partner sheet. In **KLVFF(I)** leucine and phenylalanine residues face valine and *p*-iodophenylalanine residues of a complementary β-sheet. Although these residues are in close proximity to one another, there is not a specific non-covalent interaction contributing to the packing stability, but only shape complementarity. The crystal structure of **KLVFF(I)** confirms the overall features of the cross-β spine motif that have been previously reported in the literature for other peptide fragments.[22] The native peptide and the iodinated derivative differ only by the iodine atom in the *para*-position of the terminal phenylalanine residue. Since the crystal structure shows that the iodine atom is not involved in any particular interaction, the modified sequence can be considered a reliable model of the native one. Further confirmation that the presence of iodine does not change the conformation comes from the fact that **KLVFF(I)** and KLVFFA generate the same class of steric zipper.

KLVFF(Br) was successfully crystallized by vapor diffusion using a 0.2 M solution of sodium tartrate as reservoir. Crystals suitable for single crystal X-ray diffraction formed within two month. High-resolution (0.83 Å) diffraction data were collected using Mo-Kα radiation on a Bruker KAPPA APEX II.

KLVFF(Br) crystallizes in the triclinic P1 space group. The asymmetric unit consists of two peptide molecules solvated by several disordered water molecules and one tartrate molecule (Fig. 7).

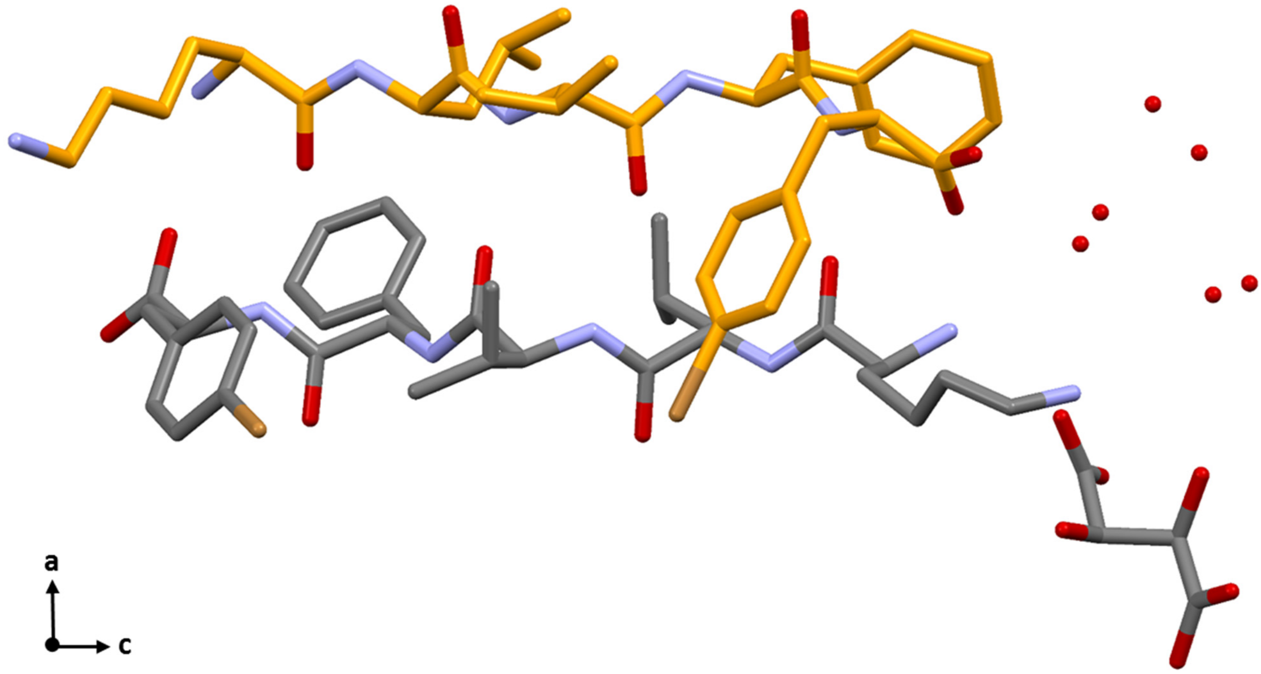


Figure 7: Crystal structure of **KLVFF(Br)**. Asymmetric unit consisting of two peptide molecule solvated by six water molecules and one molecule of tartrate. Color code: C, grey, yellow; O, red; N, violet; Br, orange; Hydrogen atoms have been omitted for clarity.

The crystal structure of **KLVFF(Br)** shares very similar structural parameters with that of the iodinated peptide. The interaction pattern of **KLVFF(Br)** is in agreement with an antiparallel β-sheet conformation, that is stabilized by hydrogen bonds involving the backbone. The packing of peptide strands is more compact in the brominated derivative. Indeed, the distance between each strand in the β-sheet is 4.34 Å (calculated considering the centroids of two adjacent monomers). Similarly to **KLVFF(I)**, the lateral self-assembly is driven by electrostatic interactions involving the charged groups of the peptide; a network of hydrogen bonding involving the crystallization solvent gives further stability along the same direction (crystallographic axis *c*). The steric zipper of the brominated peptide (Fig. 8) is the same as **KLVFF(I)** (UP-UP—FACE=BACK—CLASS 7), with distinct wet and dry interfaces. The most remarkable difference with the iodinated peptide regards the dry interface. While for **KLVFF(I)** there is only shape complementarity between the interdigitated side chains of the facing β-sheets, for the brominated peptide it is possible to notice a specific non-covalent interaction. It is a short contact C···C involving the phenylalanine and the *p*-bromophenylalanine rings of adjacent strands belonging to facing β-sheets. (C14B···C21B: 3.33 Å).

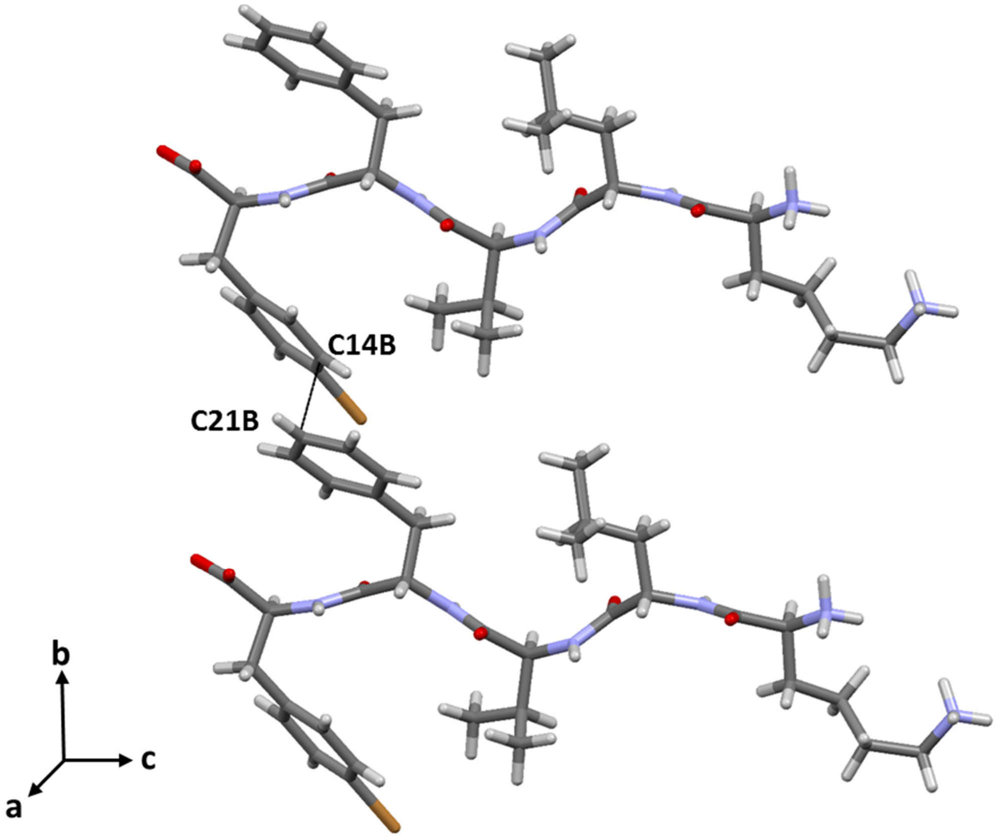


Figure 8: Steric zipper of **KLVFF(Br)**. Pairing of facing β-sheets forms the steric zipper. Two neighboring aromatic rings give a short contact stabilizing the crystal packing (C14B···C21B 3.33 Å). Color code: C, grey; O, red; N, violet; Br, orange. Hydrogen atoms have been omitted for clarity.

Despite this interaction, the resulting cross β-spine is less packed compared to **KLVFF(I)** (10.85 Å *vs* 10.49 Å, referred as the distance between centroids of facing peptide strands). This aromatic interaction is made possible by two complementary factors. First, the peculiar orientation of the *p*-bromophenylalanine ring, pointing to the opposite direction respect to its corresponding iodinated residue. Second, the electron withdrawing effect of bromine reduces the electron density on the adjacent carbon atom, enhancing the dipole attraction between the two aromatic moieties.

Table 1: Crystallographic data and refinement details for **KLVFF(I)**.

KLVFF(I)	
Crystal data	
Chemical formula	$2(C_{35}H_{52}N_6O_6^+)\cdot 2(C_2F_3O_2^-)\cdot 3(H_2O)$
M_r	1839.54
Crystal system, space group	Triclinic, *P*1
Temperature (K)	103
***a*, *b*, *c* (Å)**	9.4878 (8), 11.4256 (9), 20.593 (2)
α, β, γ (°)	82.599 (10), 80.068 (10), 76.102 (8)
V **(Å^3)**	2125.6 (3)
Z	1
Radiation type	Mo *K*α
μ (mm^{-1})	0.83
Crystal size (mm)	0.24 × 0.06 × 0.02
Data collection	
Diffractometer	Bruker *APEX*-II CCD diffractometer
Absorption correction	Multi-scan Bruker (2001). *SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
T_{min}, T_{max}	0.438, 0.577
No. of measured, independent and observed [$I > 2\sigma(I)$] reflections	33221, 15681, 13097
R_{int}	0.042
$(\sin \theta/\lambda)_{max}$ (Å^{-1})	0.617
Refinement	
$R[F^2 > 2\sigma(F^2)]$, $wR(F^2)$, S	0.048, 0.124, 1.04
No. of reflections	15681
No. of parameters	1092
No. of restraints	94
H-atom treatment	H atoms treated by a mixture of independent and constrained refinement
$\Delta\rho_{max}$, $\Delta\rho_{min}$ **(e Å^{-3})**	1.78, −0.77
Absolute structure	Flack H D (1983), Acta Cryst. A39, 876-881
Absolute structure parameter	−0.009 (12)

Table 2: Crystallographic data and refinement details for **KLVFF(Br)**.

KLVFF(Br)	
Crystal data	
Chemical formula	$2(C_{35}H_{52}BrN_6O_6)\cdot C_4H_4O_6\cdot 6(H_2O)$
M_r	860.82
Crystal system, space group	Triclinic, *P*1
Temperature (K)	103
***a*, *b*, *c* (Å)**	9.516 (3), 10.850 (3), 22.709 (6)
α, β, γ (°)	78.780 (12), 80.110 (14), 74.368 (15)
V **(Å^3)**	2197.1 (11)
Z	2
Radiation type	Mo *K*α
μ (mm^{-1})	1.00
Crystal size (mm)	0.24 × 0.12 × 0.02
Data collection	
Diffractometer	Bruker *APEX*-II CCD diffractometer
Absorption correction	Multi-scan Bruker (2001). *SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
T_{min}, T_{max}	0.809, 0.888
No. of measured, independent and observed [$I > 2\sigma(I)$] reflections	21603, 12794, 9550
R_{int}	0.032
$(\sin \theta/\lambda)_{max}$ (Å^{-1})	0.602
Refinement	
$R[F^2 > 2\sigma(F^2)]$, $wR(F^2)$, S	0.051, 0.119, 0.96
No. of reflections	12794
No. of parameters	1033
No. of restraints	245
H-atom treatment	H-atom parameters constrained
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e Å^{-3})	1.20, −0.56
Absolute structure	Flack x determined using 3289 quotients [(I+)-(I-)]/[(I+)+(I-)] (Parsons and Flack (2004), Acta Cryst. A60, s61).
Absolute structure parameter	0.021 (5)

4.5 Conclusions

Using halogenated derivatives of the natural peptide KLVFF we were able to obtain suitable crystals for X-ray diffraction analysis. The structures, obtained at high resolution, confirm experimentally for the first time the real interaction pattern of this amyloidogenic peptide. Halogenation can be used as a general tool to promote crystallization and facilitate phase determination, helping to solve the structure of molecules that are not prone to crystallization. The generality of this strategy is confirmed by the fact that both iodination and bromination lead to solutions for the peptide crystal structure with high resolution. Moreover, the two molecules are almost isostructural, further confirming that using bromine or iodine does not affect the result in these structures. Although these crystal structures do not show the halogen atoms involved in any kind of interactions, it is reasonable to think that halogens share their electron density with neighboring peptide regions, promoting the stabilization of the molecular packing and making easier the crystallization. This effect has been demonstrated for the crystal structure of an iodinated derivative of the peptide DFNKF through non-covalent interactions (NCI) analysis.[22] The same effect may be potentially expected to occur also in the structures of **KLVFF(I)** and **KLVFF(Br)**.

4.6 References

(1) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75* (1), 333.

(2) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg, D. *Nature* **2007**, *447* (7143), 453.

(3) Colletier, J.-P.; Laganowsky, A.; Landau, M.; Zhao, M.; Soriaga, A. B.; Goldschmidt, L.; Flot, D.; Cascio, D.; Sawaya, M. R.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16938.

(4) Frederix, P. W. J. M.; Scott, G. G.; Abul-Haija, Y. M.; Kalafatovic, D.; Pappas, C. G.; Javid, N.; Hunt, N. T.; Ulijn, R. V; Tuttle, T. *Nat. Chem.* **2014**, *7* (1), 30.

(5) Krysmann, M. J.; Castelletto, V.; Kelarakis, A.; Hamley, I. W.; Hule, R. A.; Pochan, D. J. *Biochemistry* **2008**, 4597.

(6) Hilbich, C.; Kisters-woike, B.; Reed, J.; Masters, C. L.; Beyreuther, K. *J. Mol. Biol.* **1992**, *228*, 460.

(7) Pawar, A. P.; DuBay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. *J. Mol. Biol.* **2005**, *350* (2), 379.

(8) Kallberg, Y.; Gustafsson, M.; Persson, B.; Thyberg, J.; Johansson, J. *J. Biol. Chem.* **2001**, *276* (16), 12945.

(9) Xie, J.; Wang, L.; Wu, N.; Brock, A.; Spraggon, G.; Schultz, P. G. *Nat. Biotechnol.* **2004**, *22* (10), 1297.

(10) Chen, L.; Rose, J. P.; Breslowt, E.; Yang, D.; Chang, W.; Furey, W. F.; Saxf, M.; Wang, B. *Biochemistry* **1991**, *88* (May), 4240.

(11) Obrecht, D.; Altorfer, M.; Bohdal, U.; Daly, J.; Huber, W.; Labhardt, A.; Lehmann, C.; Muller, K.; Ruffieux, R.; Schonholzer, P.; Spiegler, C.; Zumbrunn, C. *Biopolymers* **1997**, *42* (5), 575.

(12) Burla, M. C.; Camalli, M.; Cascarano, G. L. *J. Appl. Crystallogr.* **2003**, *36*, 1103.

(13) Sheldrick, G. M. *Acta Crystallogr. Sect. C Struct. Chem.* **2015**, *71* (Md), 3.

(14) Perutz, M. F. *Acta Crystallogr.* **1956**, *9* (11), 867.

(15) Blow, D. M.; Rossmann, M. G. *Acta Crystallogr.* **1962**, *15* (10), 1060.

(16) Kreutzer, A. G.; Hamza, I. L.; Spencer, R. K.; Nowick, J. S. *J. Am. Chem. Soc.* **2016**, *138* (13), 4634.

(17) Pham, J. D.; Spencer, R. K.; Chen, K. H.; Nowick, J. S. *J. Am. Chem. Soc.* **2014**, *136* (36), 12682.

(18) Liu, C.; Sawaya, M. R.; Cheng, P. N.; Zheng, J.; Nowick, J. S.; Eisenberg, D. *J. Am. Chem. Soc.* **2011**, *133* (17), 6736.

(19) Pham, J. D.; Chim, N.; Goulding, C. W.; Nowick, J. S. *J Am Chem Soc* **2013**, *135* (33), 12460.

(20) Shivu, B.; Seshadri, S.; Li, J.; Oberg, K. A.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2013**, *52*, 5176.

(21) Zandomeneghi, G.; Krebs, M. R. H.; McCammon, M. G.; Fändrich, M. *Protein Sci.* **2004**, *13* (12), 3314.

(22) Bertolani, A.; Pizzi, A.; Pirrie, L.; Gazzera, L.; Morra, G.; Meli, M.; Colombo, G.; Genoni, A.; Cavallo, G.; Terraneo, G.; Metrangolo, P. *Chem. Eur. J.* **2017**, *23*, 2051.

(23) Nelson, R.; Eisenberg, D. *Curr. Opin. Struct. Biol.* **2006**, *16* (2), 260.

(24) Stroud, J. C. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2013**, *69* (4), 540.

(25) Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nat. Cell Biol.* **2005**, *435* (7043), 773.

General conclusions

Among the wide range of noncovalent interactions involving halogen atoms, the potential role of halogen bonding (XB) in biomolecular self-assembly processes has seldom been explored. This thesis demonstrated the potential role of XB in the fine tuning of self-assembly involving biomolecules containing halogenated amino acids.

Since it has been demonstrated both theoretically and experimentally that aromatic residues are of great importance in the aggregation process leading to amyloid formation, the aromatic amino acid phenylalanine (Phe) was selected as a scaffold for studying the effect of halogenation both on single aminoacids and more complex peptides.

In the first part of the thesis, halogenation has been exploited as a new tool to improve the self-assembly properties of simple organic systems based on a single amino acid. N-Fmoc-phenylalanine, a well-known hydrogelator, was modified by placing halogen atoms in the *p-* position of the phenyl ring. The macroscopic properties of the resulting Fmoc-4-X-Phe hydrogels (X = F, Cl, Br, I) have been related for the first time with halogen atom polarizability, *i.e.* the propensity to act as halogen bond donor. In particular, iodinated and brominated derivatives (containing the most polarizable halogens) were found to have the highest fibrillation propensity (assessed by kinetic experiments). Moreover, their fully entangled fibril network resulted in the strongest hydrogels of the series (assessed by rheology). The crystal structure of Fmoc-4-I-Phe and Fmoc-4-Br-Phe confirms that the main non-covalent interaction driving the self-assembly is halogen bond (XB) between the halogen atom and the electron density of Fmoc aromatic moiety.

In the second part of the thesis, halogenation was applied with the same strategy on a more complex system like an amyloidogenic peptide. The segment KLVFF, core sequence of the amyloid beta (Aβ) protein, was modified to obtain seven different halogenated derivatives. Some of the halogenated peptides were able to form hydrogels over a minimum gelation concentration (MGC), lower compared to the MGC of the wild-type sequence. In this work, we reported for the first time the crystal structure of some of the halogenated KLVFF peptides, demonstrating that the halogen interaction pattern nicely correlates with the aggregation properties of the peptides. In particular, iodinated derivatives were found to assemble into more stable nanostructures, consistent with the crystal structure of **KLVF(I)F(I)** showing the peculiar cross-β pattern motif further stabilized by halogen bonding between iodine and the carbonyl oxygen of nearby β-sheets. Brominated and chlorinated derivatives showed less compact aggregates, corroborating the results of the crystal structures of the di-brominated and di-chlorinated peptides that do not show the contribution of halogen bonding in stabilizing the structures. In addition to the dependence on halogen atom polarizability, the self-assembly of the halogenated KLVFF peptides was strongly affected by the number and position of halogen atoms in the amino acidic sequence. A single point modification like halogenation resulted in a new, powerful tool to tune the self-assembly process of the same peptide segment. In fact, TEM images showed that upon changing the halogenation scheme several kinds of nanostructures were observed.

Finally, in the third part of the thesis, halogenation was shown to be a general strategy to make easier the structural determination of amyloidogenic molecules. It was possible to solve, for the first time, the crystal structure of iodinated and brominated derivatives of the KLVFF segment with excellent resolution by using a conventional X-ray source. In addition to anabling phase determination because of their remarkable electron density, halogen atoms promote the stabilization of the molecular packing, making the amyloidogenic peptides more prone to form crystals.

In conclusion, we demonstrated the potential of halogenation to become a routine method to tune the self-assembly of amyloid-like compounds. On one hand halogenation is a single-point mutation which, however, deeply changes the supramolecular properties of a molecule. On the other hand this strategy is effective both with single amino acids and peptide systems. Single crystal XRD experiments gave, for the first time, direct evidence that halogen bonding may contribute, in addition to the other non-covalent interactions, in driving the self-assembly process of halogenated biomolecules.

Acknowledgments

Firstly, I would like to express my sincere gratitude to my supervisor Prof. Pierangelo Metrangolo for the support of my PhD work and for the opportunity to join his team. Special thanks goes to Prof. Giuseppe Resnati, Prof. Giancarlo Terraneo and Prof. Francesca Baldelli Bombelli for their help and suggestions.

Besides my supervisor, I would like to thank the rest of my thesis committee – Prof. Ian W. Hamley and Prof. Alessandro Moretto – for spending their precious time to evaluate my thesis work and for their comments and suggestions that have been useful to improve my research.

My sincere thanks also goes to Dr. Nicola Demitri and Dr. Tullio Pilati for their huge work in crystal structure refinement. Without they precious support it would not be possible to conduct this research.

Many thanks to VTT – technical research centre of Finland – which hosted me in Espoo for three months giving access to the laboratory and research facilities.

A big thank goes to Elettra synchrotron Trieste for allocating the beamtime that was used to collect most of the XRD data presented in this work.

Heartfelt thanks to my labmates for each single day spent together and to Dr. Maurizio Ursini for his patience in helping me every time in designing strange experimental set up.

Last but not the least, I would like to thank my family and Roberto for their continuous and sincere support that helped me to reach this goal.