Structural studies of the in vitro assembly of tau and α-synuclein amyloids

Sofia Karin Anna Lövestam

MRC Laboratory of Molecular Biology
University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Newnham College July 2023
This thesis is dedicated to all those who have lost their lives or are currently suffering from neurodegenerative diseases.
Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Sofia Karin Anna Lövestam
July 2023
The MRC-LMB is a place where the pursuit of excellence and scientific inquiry is at the heart of everything they do. It has been wonderful to be part of this special community, and I am deeply grateful to everyone at the LMB for creating such an inspiring and intellectually stimulating environment. I would especially like to thank past and present members of the Scheres and Goedert groups. A big thank you to Max Wilkinson who has fed and stimulated my ideas. I would like to thank Diana Arseni for her continuous support. None of this would be possible without Kiarash Jamali, who compiled this document. I pledge my allegiance to Dari of House Kimanius, first of his name. I would also like to thank my cat, Max Perutz, for our many stimulating conversations. Lastly, a super big thanks to Sjors and Michel for giving me the freedom to explore science and for their continuous support.

Tau is like a never-ending kaleidoscope, with new colours and shapes emerging each time you look at it. I have to thank tau for its dynamic and ever-changing nature that has kept me on my toes and inspired me to always look deeper.

Jag tackar också min familj för er konstanta support. Tack för att ni väckte mitt intresse för forskning, annars hade jag varit helt uttråkad i livet.

I have loved doing my PhD, and I cannot wait to continue doing more research.
Abstract

Neurodegenerative diseases are characterised by the accumulation of filamentous protein aggregates, which are composed of amyloids. Cryo-EM studies of amyloid filaments isolated from human brains have revealed that specific conformers of tau and α-synuclein are associated with different diseases. These findings suggest that specific molecular mechanisms underlie the formation of filaments in the different diseases. However, studying the molecular mechanisms of amyloid assembly in post mortem brains is difficult to study. In my PhD research, I studied the molecular mechanisms of amyloid formation for tau and α-synuclein by developing in vitro amyloid assembly reactions that replicate the same structures as observed in diseased brains. My results are divided into three parts.

First, I describe the seeded assembly of recombinant α-synuclein filaments with seeds from brains with Multiple System Atrophy (MSA), and show that seeded assembly does not necessarily replicate the structures of the seeds. The results in this section have important implications when interpreting seeded assembly assays. In the future, it will be important to identify the factors that determine which structures are formed in seeded aggregation experiments.

Second, I focused on the in vitro assembly of tau. I identified truncated tau constructs, lacking the N- and C-termini, and the in vitro assembly conditions, which can accurately replicate disease-relevant folds observed in Alzheimer’s Disease (AD) and Chronic Traumatic Encephalopathy (CTE). These findings are the first to describe the formation of disease-specific structures of any amyloid, using recombinant protein in vitro. The conditions identified in this study can be used for the development of high affinity binders which are specific to the AD and CTE folds.

Finally, I studied the time-resolved filament formation of tau into AD and CTE filament folds. I show that tau filament formation is a step-wise and dynamic process, characterised by the formation of initial intermediate filaments, which I call First Intermediate Amyloids (FIAs), that subsequently mature into AD and CTE folds through a variety of later intermedi-
ate amyloid structures. This study is the first to provide a tangible structural characterisation of intermediates of amyloid filament formation.
# Table of contents

## List of figures

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Neurodegeneration</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Neuropathological characteristics</td>
<td>1</td>
</tr>
<tr>
<td>1.3</td>
<td>The case for filament formation causing neurodegeneration</td>
<td>2</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Similarities with prion diseases</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>Tauopathies</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>Synucleinopathies</td>
<td>9</td>
</tr>
<tr>
<td>1.6</td>
<td>Amyloid Structure</td>
<td>10</td>
</tr>
<tr>
<td>1.6.1</td>
<td>Cryo-EM structures of tau filaments from human brains</td>
<td>13</td>
</tr>
<tr>
<td>1.6.2</td>
<td>Cryo-EM structures of α-synuclein filaments from human brain</td>
<td>18</td>
</tr>
<tr>
<td>1.7</td>
<td>Studying amyloid formation</td>
<td>20</td>
</tr>
<tr>
<td>1.7.1</td>
<td>Solution state Nuclear Magnetic Resonance</td>
<td>20</td>
</tr>
<tr>
<td>1.7.2</td>
<td>Monitoring the kinetics of amyloid formation with Thioflavin T</td>
<td>21</td>
</tr>
<tr>
<td>1.7.3</td>
<td>Molecular mechanisms of amyloid filament formation</td>
<td>21</td>
</tr>
<tr>
<td>1.7.4</td>
<td><em>In vitro</em> assembly</td>
<td>23</td>
</tr>
<tr>
<td>1.7.5</td>
<td>Seeded aggregation</td>
<td>24</td>
</tr>
<tr>
<td>1.8</td>
<td>Overview of this dissertation</td>
<td>26</td>
</tr>
<tr>
<td>1.8.1</td>
<td>Chapter 2: Seeded aggregation of recombinant α-synuclein</td>
<td>26</td>
</tr>
<tr>
<td>1.8.2</td>
<td>Chapter 3: Spontaneous <em>in vitro</em> assembly with recombinant tau</td>
<td>27</td>
</tr>
<tr>
<td>1.8.3</td>
<td>Chapter 4: Time-resolved studies of tau assembly</td>
<td>27</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 Neurodegeneration

1.2 Neuropathological characteristics

1.3 The case for filament formation causing neurodegeneration

1.3.1 Similarities with prion diseases

1.4 Tauopathies

1.5 Synucleinopathies

1.6 Amyloid Structure

1.6.1 Cryo-EM structures of tau filaments from human brains

1.6.2 Cryo-EM structures of α-synuclein filaments from human brain

1.7 Studying amyloid formation

1.7.1 Solution state Nuclear Magnetic Resonance

1.7.2 Monitoring the kinetics of amyloid formation with Thioflavin T

1.7.3 Molecular mechanisms of amyloid filament formation

1.7.4 *In vitro* assembly

1.7.5 Seeded aggregation

1.8 Overview of this dissertation

1.8.1 Chapter 2: Seeded aggregation of recombinant α-synuclein

1.8.2 Chapter 3: Spontaneous *in vitro* assembly with recombinant tau

1.8.3 Chapter 4: Time-resolved studies of tau assembly

## 2 Seeded assembly of recombinant α-Synuclein

2.1 Introduction

2.2 Results

2.2.1 Seeding recombinant wild-type α-synuclein with seeds from three cases of MSA
2.2.2 Cryo-EM imaging of seeded $\alpha$-synuclein filaments .............. 31
2.2.3 Cryo-EM structures of type 1 and type 2 $\alpha$-synuclein filaments ... 36
2.2.4 Cryo-EM structure of type 3 $\alpha$-synuclein filaments .............. 37
2.2.5 Cryo-EM structures of $\alpha$-synuclein filaments from second-generation seeded aggregation .......................... 43
2.3 Discussion ........................................................................... 46
2.4 Conclusions ....................................................................... 49

3 Assembly of recombinant tau .................................................. 51
3.1 Introduction ........................................................................ 51
3.2 Results ............................................................................. 52
  3.2.1 In vitro assembly of tau (297–391) into PHFs ................. 52
  3.2.2 The effects of salts on tau filament assembly .............. 62
  3.2.3 The effects of protein length on tau filament assembly .... 68
  3.2.4 The effect of pseudo-phosphorylation on tau filament assembly . 71
  3.2.5 The effects of anionic cofactors on tau filament assembly . 74
3.3 Discussion ........................................................................... 75

4 Time-resolved studies of tau filaments .................................... 79
4.1 Introduction ........................................................................ 79
4.2 Results ............................................................................. 80
  4.2.1 Monomeric tau ............................................................. 80
  4.2.2 Cryo-EM structures of early tau intermediate filaments .... 82
  4.2.3 Time-resolved filament formation in the presence of MgCl$_2$ ... 86
  4.2.4 Time-resolved filament formation in the presence of NaCl .... 96
  4.2.5 Fold conversion ............................................................ 104
4.3 Discussion ........................................................................... 105

5 Discussion ............................................................................. 109
5.1 Overview ........................................................................... 109
5.2 Outlook ............................................................................ 112

6 Methods .................................................................................. 115
6.1 Molecular cloning ............................................................... 115
6.2 Protein purification for $\alpha$-synuclein ............................... 116
6.3 Protein purification for tau ................................................. 117
  6.3.1 Considerations for the purification of amyloid proteins .... 118
# List of figures

1.1 Tauopathies and Synucleinopathies ............................................. 3
1.2 Abnormal inclusions in brain ..................................................... 6
1.3 X-ray fibre diffraction of amyloid filaments ................................. 11
1.4 Cryo-EM structures of amyloid filaments .................................. 12
1.5 Tau folds .................................................................................. 17
1.6 α-Synuclein folds ....................................................................... 19
1.7 Schematic of the kinetic profile for amyloid filament filament in the presence of ThT ......................................................... 22

2.1 Cryo-EM structures of type I and type II of α-synuclein filaments from MSA brains ................................................................. 30
2.2 α-synuclein with filament preparations from MSA brains ............... 32
2.3 Purified filaments from MSA case 5 ............................................. 33
2.4 Biochemical characterisation of seeded material ............................. 34
2.5 Cryo-EM structures of type 1 and type 2 α-synuclein filaments with protofilament fold A assembled using seeds from MSA case 2 .... 35
2.6 Cryo-EM structures of type 1 and type 2 filaments with protofilament fold B assembled using seeds from MSA case 1 ............................. 38
2.7 Cryo-EM structures of type 1 and type 2 filaments with protofilament fold A assembled using seeds from MSA case 1 ............................. 39
2.8 Comparison of protofilament A with PDB-entry 6SSX of recombinant wild-type α-synuclein ......................................................... 40
2.9 Comparison of protofilament fold A with PDB-entry 6UFR of assembled recombinant E46K α-synuclein ......................................................... 41
2.10 Comparison of protofilament folds A and B .................................. 42
2.11 Cryo-EM structure of type 3 filaments assembled using seeds from MSA case 5 ................................................................. 43
2.12 Comparison of type 3 α-synuclein filaments with protofilament IIB from MSA case 5 ................................................................. 44
List of figures

2.13 Second-generation type 3 filaments .............................................. 45
2.14 Comparison of type 3 filament with PDB entry 6PEO of assembled recombi-
nant H50Q $\alpha$-synuclein .............................................................. 45
2.15 Summary of MSA seeded aggregation experiments .......................... 47

3.1 New electron cryo-microscopy (cryo-EM) structures ......................... 54
3.2 Cross-sections of electron cryo-microscopy (cryo-EM) reconstructions ... 55
3.3 Electron cryo-microscopy (cryo-EM) maps and models of new structures
   (part 1) ....................................................................................... 56
3.4 Electron cryo-microscopy (cryo-EM) maps and models of new structures
   (part 2) ....................................................................................... 57
3.5 Schematics of the tau folds (part 1) .................................................. 58
3.6 Schematics of the tau folds (part 2) .................................................. 59
3.7 Assembly of recombinant tau into filaments like Alzheimer’s disease paired
   helical filaments (AD PHFs) .......................................................... 60
3.8 Electron cryo-microscopy (cryo-EM) micrographs and density maps com-
   paring in vitro assembled paired helical filaments (PHFs) and Alzheimer’s
   disease paired helical filaments (AD PHFs) ..................................... 61
3.9 Assembly of recombinant tau into filaments like chronic traumatic encephalopa-
   thy (CTE) ....................................................................................... 64
3.10 Extended and C-shaped NaCl tau protofilaments. .............................. 65
3.11 Extended and C-shaped LiCl tau protofilaments. ............................... 66
3.12 Cryo-EM densities inside the cavities of KCl and NaCl tau filaments ... 67
3.13 The effects of protein length on tau filament assembly ...................... 70
3.14 Comparison of in vitro assembled tau filaments with the globular glial tauopa-
   thy (GGT) fold ............................................................................ 72
3.15 Coomassie stained sodium dodecyl-sulfate polyacrylamide gel electrophore-
   sis (SDS-PAGE) gel for constructs 297–391, 297–408, and 296–441 before
   and after filament assembly ......................................................... 73

4.1 Analytical ultracentrifugation of monomeric tau ............................... 80
4.2 ThT and cryo-EM micrographs shown for a time-course .................... 81
4.3 Cryo-EM micrographs of tau intermediate filaments ........................ 83
4.4 Cryo-EM micrograph and the 2D reconstruction of early tau filaments: FIA 83
4.5 Structure of FIAs ......................................................................... 84
4.6 Structures of tau filaments formed in the presence of MgCl$_2$ ........... 86
4.7 Structures of tau filaments formed after 180 min in the presence of MgCl$_2$ . 87
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.8</td>
<td>Structures of tau filaments formed after 240 min in the presence of MgCl&lt;sub&gt;2&lt;/sub&gt; . 90</td>
</tr>
<tr>
<td>4.9</td>
<td>Comparison of extended and C-shaped protofilaments formed at time-intervals 180-240 min in the presence of MgCl&lt;sub&gt;2&lt;/sub&gt; . 91</td>
</tr>
<tr>
<td>4.10</td>
<td>Comparison of PHF crossover distances formed after 300 min in the presence of MgCl&lt;sub&gt;2&lt;/sub&gt; . 93</td>
</tr>
<tr>
<td>4.11</td>
<td>Structures of C-shaped tau protofilaments formed after 300 min in the presence of MgCl&lt;sub&gt;2&lt;/sub&gt; . 94</td>
</tr>
<tr>
<td>4.12</td>
<td>Comparison of extended protofilament and C-shaped protofilament structures formed in the presence of MgCl&lt;sub&gt;2&lt;/sub&gt; . 95</td>
</tr>
<tr>
<td>4.13</td>
<td>Structures of tau filaments formed in the presence of NaCl . 96</td>
</tr>
<tr>
<td>4.14</td>
<td>Structures of tau filaments formed after 180 min in the presence of NaCl . 98</td>
</tr>
<tr>
<td>4.15</td>
<td>Structures of tau filaments formed after 240 min in the presence of NaCl . 99</td>
</tr>
<tr>
<td>4.16</td>
<td>Structures of tau filaments formed after 300 min in the presence of NaCl . 101</td>
</tr>
<tr>
<td>4.17</td>
<td>Structures of tau filaments formed after 720 in the presence of NaCl . 102</td>
</tr>
<tr>
<td>4.18</td>
<td>Comparison of extended and C-shaped protofilament structures formed in the presence of NaCl . 103</td>
</tr>
<tr>
<td>4.19</td>
<td>Comparison of extended and C-shaped protofilaments at residues S320, G355 and G356 . 105</td>
</tr>
<tr>
<td>6.1</td>
<td>Concepts of In Vivo Assembly . 116</td>
</tr>
<tr>
<td>6.2</td>
<td>Cryo-EM micrograph inspection of amyloid filaments . 122</td>
</tr>
<tr>
<td>A.1</td>
<td>Additional cryo-EM data on type 1 and type 2 filaments with protofilament fold A . 130</td>
</tr>
<tr>
<td>A.2</td>
<td>Additional cryo-EM data on type 1 and type 2 filaments with protofilament fold B . 131</td>
</tr>
<tr>
<td>A.3</td>
<td>Additional cryo-EM data on type 3 filaments . 132</td>
</tr>
<tr>
<td>A.4</td>
<td>Statistics on structures solved for the seeded aggregation of recombinant α-synuclein using seeds from MSA . 133</td>
</tr>
<tr>
<td>A.5</td>
<td>Fourier shell correlation curves (part 1) . 134</td>
</tr>
<tr>
<td>A.6</td>
<td>Fourier shell correlation curves (part 2) . 135</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Neurodegeneration

Neurodegeneration refers to the progressive loss of structure and function of neurons in the human brain. The majority of neurodegenerative diseases are caused by the accumulation of abnormal proteins within the brain, which can interfere with neural function and cause nerve cell death. The most common neurodegenerative diseases are Alzheimer’s disease (AD) and Parkinson’s disease (PD) (Przedborski et al., 2003).

The impact of neurodegenerative diseases on society is profound. The importance of understanding the molecular mechanisms of neurodegeneration, particularly the role of abnormal protein aggregation, is essential for developing effective treatments.

1.2 Neuropathological characteristics

The major features of most neurodegenerative diseases are the accumulation of abnormal inclusions within the brain (Dugger & Dickson, 2017). In 1907, Alois Alzheimer used Bielschowsky silver to identify the combined presence of plaques and tangles in the brain of Auguste Deter who suffered from dementia (Alzheimer, 1907). Four years later, Alzheimer also described what we now call Pick bodies in cases of frontotemporal dementia (Alzheimer, 1911). In 1912, Friedrich Lewy described the eosinophilic intraneuronal inclusion bodies in patients with PD (Lewy, 1912). In the 1960s, electron microscopy (EM) showed that these inclusions contain abnormal filaments (Duffy & Tennyson, 1965; Kidd, 1963; Rewcastle & Ball, 1968). Other protein deposits were subsequently described, which led to the generalisation that inclusions comprising abnormal filaments are associated with several
neuropathologies. The presence and morphology of these lesions have helped to classify neurodegenerative diseases (Dugger & Dickson, 2017).

Between 1984 and 1997, the proteins amyloid-β (Aβ), tau and α-synuclein were identified as the major components of many filamentous inclusions (Brion et al., 1985; Glenner & Wong, 1984; Goedert, Wischik, et al., 1988; V. M. Lee et al., 1991; Masters et al., 1985; Pollock et al., 1986; Rasool & Selkoe, 1985; Spillantini, Schmidt, et al., 1997). Aβ is the major component of plaques and blood vessel deposits in AD. Tau is the major component of tangles in AD and Pick bodies in Pick’s disease. Tau is also the major component of the filamentous inclusions of a number of additional neurodegenerative diseases, including chronic traumatic encephalopathy (CTE), progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD). α-Synuclein is the major component of the filamentous inclusions of Lewy bodies and Lewy neurites in PD, PD dementia (PDD) and dementia with Lewy bodies (DLB), as well as of the glial cytoplasmic inclusions (GCIs) and neuronal deposits in multiple system atrophy (MSA).

1.3 The case for filament formation causing neurodegeneration

The discovery that some mutations in the genes that encode for the proteins that make up the amyloid filaments, which define neurodegenerative diseases, cause rare cases of dominantly inherited disease, has made it clear that filament formation is of importance for disease.

Mutations (missense, deletion and gene dosage) in APP, the gene encoding the amyloid precursor protein, cause familial forms of AD, with abundant plaques, tangles and nerve cell loss (Goate et al., 1991; Rovelet-Lecrux et al., 2006). The existence of these mutations is a major argument underpinning the amyloid cascade hypothesis, which states that APP dysfunction leads to all downstream effects of AD, including tangle formation and dementia (Hardy and Higgins, 1992). APP is a type I transmembrane protein whose gene maps to chromosome 21 (Kang et al., 1987) and which gets cleaved physiologically by β-secretase and γ-secretase to give rise to Aβ with lengths of 38-43 amino acids (Haass et al., 2012). The α-secretase cleaves within the Aβ peptide in APP, preventing the generation of Aβ. The most abundant form of soluble Aβ is 40 amino acid residues long, whilst Aβ42 is the predominant species in the assembled peptide in plaques. The γ-secretase complex generates the C-terminal end of Aβ from APP. Mutations in the presenilin genes PSEN1 and PSEN2,
1.3 The case for filament formation causing neurodegeneration

![Image]

**Tauopathies**

<table>
<thead>
<tr>
<th>3R + 4R tauopathies</th>
<th>Synucleinopathies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer's Disease (AD)</td>
<td>Parkinson’s Disease (PD)</td>
</tr>
<tr>
<td>Chronic Traumatic Encephalopathy (CTE)</td>
<td>Parkinson’s Disease Dementia (PDD)</td>
</tr>
<tr>
<td>Primary age-related tauopathy (PART)</td>
<td>Dementia with Lewy bodies (DLB)</td>
</tr>
<tr>
<td>Familial British Dementia (FBD)</td>
<td>Multiple System Atrophy (MSA)</td>
</tr>
<tr>
<td>Familial Danish Dementia (FDD)</td>
<td>Juvenile Onset Synucleinopathy (JOS)</td>
</tr>
<tr>
<td>Gerstmann–Sträussler–Scheinker disease (GSS)</td>
<td>(7 amino acid insertion)</td>
</tr>
<tr>
<td>Niemann–Pick disease, type C (N-PiD)</td>
<td>Familial Parkinson’s Disease (P-PD)</td>
</tr>
<tr>
<td>Tangle only dementia</td>
<td>(SNCA mutations e.g. E46K, A30P, A53T and multiplications of the gene)</td>
</tr>
<tr>
<td>Familial frontotemporal dementia and parkinsonism - FTLD-17T (MAPT mutations, e.g. V337M R406W)</td>
<td></td>
</tr>
<tr>
<td>Diffuse neurofibrillary tangles with calcification Down’s syndrome</td>
<td></td>
</tr>
<tr>
<td>Non-Guamanian motor neuron disease with neurofibrillarytangles Postencephalitic parkinsonism</td>
<td></td>
</tr>
<tr>
<td>Progressive ataxia and palatal tremor ALS/parkinsonism–dementia complex Anti-IgLON5-related tauopathy</td>
<td></td>
</tr>
</tbody>
</table>

**3R tauopathies**

<table>
<thead>
<tr>
<th>Pick’s Disease (PiD)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Familial frontotemporal dementia and parkinsonism - FTLD-17T (MAPT mutations, e.g. G272V Q336R)</td>
<td></td>
</tr>
</tbody>
</table>

**4R tauopathies**

<table>
<thead>
<tr>
<th>Argyrophilic grain Disease (AGD)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aging-relate tau astroglialopathy (ARTAG)</td>
<td></td>
</tr>
<tr>
<td>Guadeloupean parkinsonism</td>
<td></td>
</tr>
<tr>
<td>Tau astroglialopathy</td>
<td></td>
</tr>
<tr>
<td>limbic-predominant neuronal inclusion body</td>
<td></td>
</tr>
<tr>
<td>4R tauopathy’ (LNT)</td>
<td></td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td></td>
</tr>
<tr>
<td>SLC9a–related parkinsonism</td>
<td></td>
</tr>
<tr>
<td>Corticobasal degeneration (CBD)</td>
<td></td>
</tr>
<tr>
<td>Globular glial Tauopathy (GGT)</td>
<td></td>
</tr>
<tr>
<td>Progressive supranuclear palsy (PSP)</td>
<td></td>
</tr>
<tr>
<td>Familial frontotemporal dementia and parkinsonism - FTLD-17T (MAPT mutations, e.g. P301S/P301L/P301T and all intronic mutations flanking exon 10 e.g. +3 and +16)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1.1 Tauopathies and Synucleinopathies
are the most common cause of familial AD (De Strooper et al., 1998; Levy-Lahad et al., 1995; Rogaev et al., 1995; Sherrington et al., 1995). Currently, 160 diseases-associated mutations in the \textit{PSEN} genes have been identified. Presenilins are part of the \(\gamma\)-secretase complex. Mutations associated with familial Alzheimer’s disease increase the ratios of A\(\beta\)42 to A\(\beta\)40 (Scheuner et al., 1996; Suzuki et al., 1994), the concentration of A\(\beta\)42 (Citron et al., 1994) and/or the assembly of A\(\beta\)42 into filaments (Pagnon de la Vega et al., 2021).

The first mutations in \textit{MAPT}, the tau gene, were identified in 1998 in a familial form of frontotemporal dementia and parkinsonisms linked to chromosome 17, (FTDP-17T) (Hutton et al., 1998; Poorkaj et al., 1998; Spillantini, Murrell, et al., 1998). \textit{MAPT} mutations have their primary effects at the mRNA or the protein level (this is detailed in the tau section below). All cases of FTDP-17T have abundant filamentous tau inclusions in the brain, which indicates that the formation of tau filaments is sufficient to cause neurodegeneration and dementia. A\(\beta\) deposits are not observed.

The first mutation in \textit{SNCA}, the \(\alpha\)-synuclein gene, was identified in 1997 as the cause of familial PD (Polymeropoulos et al., 1997). Some missense mutations in \textit{SNCA} also cause DLB (Zarranz et al., 2004). Additional missense mutations, as well as gene dosage mutations (Singleton et al., 2003) and an insertion mutation in \textit{SNCA} (Yang, Garringer, et al., 2022) have been described. Additionally, sequence variations in the regulatory region of \textit{SNCA}, which increase the expression of \(\alpha\)-synuclein, have been identified as a risk factor for idiopathic PD (Nalls et al., 2014). In all cases, abundant \(\alpha\)-synuclein filaments are present.

1.3.1 Similarities with prion diseases

The discovery of a self-replicating protein component devoid of DNA or RNA transmission was first described for the prion protein (Prusiner, 1982). Prion diseases are sporadic, inherited or acquired (Imran & Mahmood, 2011). The prion protein (PrP) is able to convert its cellular, soluble form (PrPc) into an abnormal aggregated form (PrP\(\delta\)). PrP\(\delta\) was shown to be transmissible across individuals and across species, as PrP\(\delta\) could induce the conversion of PrPc to PrP\(\delta\). The transmission of PrP\(\delta\) across species has resulted in a few disease outbreaks. In the 1980s, bovine spongiform encephalopathy (BSE) caused by PrP\(\delta\) in cattle was transmitted to humans after ingestion of contaminated meat, resulting in variant Creutzfeldt–Jakob disease (vCJD) (Prusiner, 1997). Injecting PrP\(\delta\) aggregates from sheep (scrapie) into mice or hamsters expressing PrPc resulted in neurodegeneration (Scott et al., 1989). Distinct prion
isolates or strains showed different incubation times and neuropathological characteristics.

The observation that aggregates can be transmitted between individuals suggests a molecular mechanism of filament spreading through templated seeding. In templated seeding, a small amount of filaments can induce the conversion of the soluble form of the protein into filaments beyond what would have happened spontaneously. Recent studies have suggested that a “prion-like” mechanism is also implicated in the most common neurodegenerative diseases, including AD and PD (Goedert, Falcon, et al., 2014). Studies have shown that iatrogenic transmission of Aβ seeds can occur. In a study involving eight patients who died from CJD and received injections of cadaver-derived human growth hormone, four of them exhibited a high level of Aβ plaques (Jaunmuktane et al., 2015). Additionally, the transmission of iatrogenic CJD through cadaveric human dura mater grafts resulted in the development of abundant Aβ plaques (Preusser et al., 2006; Frontzek et al., 2016). Indeed, the propagation and distribution of inclusions in AD and PD follows a stereotypical progression from single sites in the brain (H. Braak & Del Tredici, 2016). However, unlike prion diseases, there is no evidence of transmissibility between individuals for the most common neurodegenerative diseases (Goedert, 2015; Scialò et al., 2019). Nevertheless, the stereotypical spreading of filamentous inclusions in disease, the observation that specific filament structures characterise different diseases (which I will describe below), and that the structures of the filaments are shared in distinct brain regions within a brain (Arseni et al., 2021; Y. Shi, Zhang, et al., 2021; Yang, Arseni, et al., 2022), support the hypothesis that templated seeding also underlies the propagation of filamentous inclusions in neurodegenerative diseases.

In this thesis, I will focus on neurodegenerative diseases that are characterised by the inclusions of tau and α-synuclein
Fig. 1.2 Abnormal inclusions in brain Abnormal inclusions shown for a amyloid-β, b, d-f tau and c, g α-synuclein. Schematic representations of the brain is shown in black outline. Colours denote the occurrence of protein inclusions from an early (left) to late (right) onset of disease. e-f pathology’s show end-stage only, where colours show sites with abnormal inclusions. a and b is AD, c PD d PSP, e CTE, f PiD, g MSA.
1.4 Tauopathies

The misfolding and accumulation of tau into amyloid filaments are characteristic of more than 20 neurodegenerative diseases, which are collectively called tauopathies (Goedert, 2018). Six tau isoforms are expressed in adult human brains; they are produced from a single gene by alternative mRNA splicing and range from 352 to 441 amino acids in length (Goedert, Spillantini, Jakes, et al., 1989). The isoforms differ by the presence or absence of 29 and 58 amino acids in the N-terminal domain (0N, 1N and 2N), and the inclusion or exclusion of the 31 amino acids of the second microtubule-binding repeat, resulting in three repeat (3R) and four repeat (4R) tau. Tau can be divided into an N-terminal domain (residues 1-150), a proline-rich region (residues 151-243), four microtubule binding repeats (residues 244-368) and the C-terminal domain (residues 369-441) (Figure 1.5). The proline-rich region contains several PxxP or xPPx motifs in which “x” amino acids are often serines or threonines. This region mediates tau’s interaction with SH3 domain containing proteins, including kinases. The C-terminus of tau forms an α-helical structure and is thought to inhibit filament assembly (G. Lee et al., 1998). The microtubule-binding repeats promote microtubule assembly and stability of microtubules. Similar levels of 3R and 4R tau are expressed in normal adult human brain (Goedert & Jakes, 1990). 4R tau isoforms are better at promoting microtubule assembly than 3R isoforms. During the development of the human brain, only the shortest 3R tau isoform is expressed.

The relevance of tau for neurodegeneration and dementia became apparent with the discovery of mutations in MAPT that cause FTDP-17T (Hutton et al., 1998; Poorkaj et al., 1998; Spillantini, Murrell, et al., 1998). Currently 65 disease mutations have been identified. They are concentrated in exons 9-12, which encode the microtubule-binding repeats (R1-R4) and in the introns flanking exon 10. Mutations have their primary effects at the mRNA and/or the protein levels. Mutations in intron 10 and many mutations in exon 10 increase the relative expression of 4R tau, resulting in the assembly of 4R tau filaments in nerve cells and glial cells. The intronic mutations destabilise a tau exon 10 splicing regulatory element (Hutton et al., 1998; Spillantini, Murrell, et al., 1998; Varani et al., 1999). Only mutation ΔK280 in exon 10 causes a relative overproduction of 3R tau compared to 4R tau and results in the assembly of 3R tau in nerve cells and glial cells, reminiscent of PiD. It reduces exon 10 inclusion by abolishing a splicing enhancer element (D’Souza et al., 1999). Most MAPT mutations are missense mutations that change single amino acids. They reduce the ability of tau to interact with microtubules, with some mutations also increasing heparin-induced assembly of recombinant tau into filaments (Goedert, Eisenberg, et al., 2017).
Immunohistochemistry has shown distinct regional distributions and morphological differences of tau inclusions for different tauopathies. Tau inclusions are commonly observed in neurons and in glial cells. In AD, extracellular senile plaques composed of Aβ are accompanied by intracellular tau neurofibrillary tangles, neuropil threads and extracellular ghost tangles (Wilcock & Esiri, 1982). In chronic traumatic encephalopathy (CTE), tau inclusions are present in neurons, astroglia and around blood vessels (Geddes et al., 1999; McKee et al., 2016). PiD is characterised by Pick bodies which are spherical inclusions in some neurons (Delacourte et al., 1996; Pollock et al., 1986; Probst et al., 1996; Rasool & Selkoe, 1985). Progressive supranuclear palsy (PSP) is characterised by neurofibrillary tangles and neuropil threads, tufted astrocytes and oligodendroglial coiled bodies (Höglinger et al., 2017; Kovacs, Lukic, et al., 2020). Corticobasal degeneration (CBD) contains astrocytic and neuronal tau inclusions (Gibb et al., 1989; Rebeiz et al., 1968). Astrocytic plaques are pathognomonic. Globular glial tauopathy (GGT) contains globular glial inclusions (Ahmed, Bigio, et al., 2013; Ahmed, Doherty, et al., 2011; Molina et al., 1998). Argyrophilic grain disease (AGD) is defined by argyrophilic grains in the neuropil (H. Braak & E. Braak, 1987, 1989; Ferrer et al., 2008; Tolnay & Clavaguera, 2004). Aging-related tau astrogliopathy (ARTAG) contains mostly neuronal and glial inclusions (Kovacs, Ferrer, et al., 2016). AGD and ARTAG typically co-exist with neurofibrillary tangles. The locations of the inclusions and staging patterns in the brain is different for the different diseases (H. Braak & Del Tredici, 2016).

Tau filaments from human brain can be extracted using tissue solubilisation with detergents such as sarkosyl, followed by differential centrifugation. Water extraction has also been used (Masters et al., 1985). Tau filaments are sarkosyl insoluble. Immunoblot analysis of sarkosyl insoluble tau shows different isoform patterns depending on the tauopathy. AD Tau shows pathological bands at 60, 64, 68 and 72 kDa and comprises all six isoforms. PiD tau shows bands for 3R tau at 60 and 64 kDa. CBD tau, PSP tau, AGD tau and ARTAG tau show bands for 4R tau at 64 and 68 kDa. Truncations of tau resulting in C-terminal fragments of tau are characteristic for each disease and results in distinct band patterns. In AD, fragments consisting of 19, 22, 25, 30, 36 and 40 kDa are observed. In PiD bands corresponding to 21, 34 and 39 kDa are present. In PSP bands migrate at 22 and 33 kDa whereas in CBD bands of 37 and 43 kDa are observed. The 33 kDa and 37 kDa bands are characteristic to PSP and CBD, respectively (Hasegawa et al., 1992). The 33 kDa band is also observed in GGT, whilst the 37 kDa bands are characteristic of AGD, ARTAG and tauopathies caused by splicing defects in the MAPT intron 10(+3/+16) (Forrest et al., 2018; Hutton et al., 1998; Spillantini, R. A. Crowther, et al., 1998; Spillantini, Goedert, R. A. Crowther, et al., 1997). Immunoblot analysis of proteolytic digestions of tau filaments has identified that the structured core of tau
1.5 Synucleinopathies

Synucleinopathies, which comprise PD, PDD, DLB and MSA, are defined by the presence of abundant α-synuclein filaments in nerve cells and glial cells (Goedert, 2001). α-Synuclein belongs to a protein family that also comprises β-synuclein and γ-synuclein (George, 2002; Maroteaux et al., 1988). Only α-synuclein is present in disease filaments. α-Synuclein consists of 140 amino acids and can be divided into an N-terminal region that comprises 7 imperfect repeats (residues 7-87) with the consensus sequence KTKEGV that are lipid-binding domains (Davidson et al., 1998), a non-αβ component (NAC, residues 61-95) and a negatively charged C-terminus (residues 96-140) (Goedert, 2001) (Figure 1.6). α-Synuclein is localised to pre-synaptic membranes in neurons and is believed to be involved in neural plasticity (Murphy et al., 2000).

In Lewy pathologies, filamentous cytoplasmic inclusion bodies occur mostly in neurons in the form of Lewy bodies and Lewy neurites. In MSA, in addition to neurons, abundant α-synuclein inclusions are present in glial cells, in particular oligodendrocytes, as glial cytoplasmic inclusions (GCIs) also known as Papp-Lantos bodies (Papp et al., 1989; Spillantini,
α-Synuclein filaments from diseased brains with Lewy bodies and MSA can be extracted using sarkosyl solubilisation and differential ultra-centrifugation. Immunoblotting sarkosyl insoluble α-synuclein results in a predominant 15 kDa species (Spillantini, Goedert, Farlow, et al., 1998; Spillantini, Schmidt, et al., 1997). Tryptic digestion of α-synuclein filaments, followed by purification and mass spectrometry of the proteolytically stable core, showed a 7 kDa fragment. The 7 kDa fragment comprises amino acid residues 31-109 of α-synuclein (Miake et al., 2002). Mass spectrometry of α-synuclein filaments has shown that C-terminal truncations and phosphorylation at S129 are common in Lewy pathologies and in MSA (Turtani et al., 2018). Negative stain EM showed that the morphologies of filaments extracted from Lewy pathologies, including PD, PDD and DLB are similar, but different for MSA filaments (R. A. Crowther, Daniel, et al., 2000; Spillantini, R. A. Crowther, et al., 1998; Spillantini, Schmidt, et al., 1997).

1.6 Amyloid Structure

The term amyloid was coined by Virchow in 1854 to describe abnormal inclusions observed in the corpora amylacea of the brain, as they stained blue with iodine sulfuric acid and were thus mistaken for starch. In 1859, Friedrich and Kekule demonstrated that amyloids contained high amounts of nitrogen and no carbohydrate, suggesting that proteins were the major components of the abnormal inclusions (Friedreich, 1859). Subsequent studies using Congo red and polarised light showed the apple green birefringence inherent to amyloids (Divry & Florkin, 1927).

X-ray fibre diffraction pattern of post mortem deposits showed that amyloids contain a repeated pattern of cross β-sheet structure (Figure 1.3) (Eanes & Glenner, 1968). Negative stain EM of AD extracted filaments showed unbranched paired helical filaments (PHFs) and straight filaments (SFs) consisting of identical C-shaped subunits (R. A. Crowther, 1991; Kidd, 1963). Because amyloids from human brains cannot be crystallised, structure determination was not possible using crystallography. X-ray crystallography of chemically synthesised shorter peptides encoding regions of the amyloid proteins that were grown into filaments in vitro revealed the first atomic structures of amyloid filaments (Nelson et al., 2005). These structures confirmed that the cross-β pattern, which was observed in previous diffraction studies, came from repeated β-strands which arrange perpendicular to the helical...
axis and consist of multiple copies of the protein molecule separated by 4.75 Å. β-sheets align parallel to the helical axis and pack against each other with inter-sheet distances of 10-12 Å. The extended H-bonding network along the filament axis explains the insoluble nature of amyloids.

Solid-state nuclear magnetic resonance (ssNMR) resulted in the first atomic structure of filaments assembled from full-length recombinant proteins, including Aβ 40 and α-synuclein. The structures of Aβ 40 filaments were polymorphic and consisted of two straight protofilaments which were parallel or antiparallel (Petkova et al., 2002). The structure of α-synuclein filaments comprised several in-register parallel β-sheets that adopted an L-shaped motif folded into a greek key-like conformation (Tuttle et al., 2016). Because solid-state NMR requires isotopically labelled proteins, it cannot be used for structure determination of samples from post mortem human brains.

---

**Fig. 1.3 X-ray fibre diffraction of amyloid filaments.** An incoming beam of x-rays is shot onto the fibrous amyloids. The scattered beam diffracts onto a plate resulting in a diffraction pattern characteristic to amyloids. This shows the 4.70-4.90 Å spacing of the β-strands and the 10 Å spacing of the β-sheets.

Since the availability of direct electron detectors and imaging processing software around 2013, electron cryo-microscopy (cryo-EM) has become a powerful tool for elucidating the structures of many biological structures. But it was not until new image processing software for helical filaments was developed, that *de novo* atomic modelling in cryo-EM maps also
became possible for amyloids (He & S. H. W. Scheres, 2017). RELION performs helical processing in a Bayesian framework, where experimental data is combined with prior knowledge about the structure (Sh, 2012). Cryo-EM is the most widely used tool for elucidating the structures of amyloids, including those extracted from human brains (Figure 1.4).

Fig. 1.4 Cryo-EM structures of amyloid filaments a Cryo-EM micrograph shows amyloid filaments in vitrified ice, scale bar 50 nm. A filament is highlighted in blue. b Cryo-EM density map of an amyloid filament showing the β-strand separation, 4.70 - 4.90 Å, in the cryo-EM map. c X-Y Cross-section of the cryo-EM density map. The example shows the structure of the Paired Helical Filament from AD (EMDB: 3741, PDB: 6HRE)
1.6 Amyloid Structure

1.6.1 Cryo-EM structures of tau filaments from human brains

Cryo-EM studies of filaments extracted from post mortem tauopathy brains have revealed the atomic structures of tau filaments in disease (Figure 1.5). The first structures were of AD PHFs and SFs, which are each made of two protofilaments with a common C-shaped core (Fitzpatrick et al., 2017). The core is made of amino acids V306-F378, which adopt a cross-β structure. Residues 306-378 comprise R3 and R4 of tau and explains the incorporation of all isoforms of tau into the filaments, since these repeats are shared between 3R and 4R tau. The core of the structure also contains 10-13 amino acids after R4 (368-378/380). Each protofilament consists of eight β-strands (β1-β8). Five β-strands give rise to two pairs of antiparallel β-sheets. The other three β-strands form a β-helix. PHFs and SFs differ in their inter-protofilament packing. In PHFs, the two C-shaped protofilaments are packed symmetrically against each other at residues 332-338. In SFs, the protofilaments pack asymmetrically and are stabilised by an additional, non-proteinaceous density located between the side chains of K317 and K321 in both protofilaments. As such, PHFs and SFs are ultrastructural polymorphs. The structures of PHFs and SFs have been solved from filaments extracted from the brains of multiple individuals with AD. This study showed that tau filament structures of different individuals with AD are identical (Falcon, Zhang, Schweighauser, et al., 2018).

Tau filaments extracted from the brains of individuals with CTE revealed a novel fold (Falcon, Zivanov, et al., 2019). Similar to the AD fold, the CTE fold shares the protofilament C-shaped arrangement and consists of two types of filaments, CTE type I and CTE type II (Falcon, Zivanov, et al., 2019). The core comprises residues 305-379 of the third and fourth repeats of tau, plus 12 residues after R4. However, the CTE fold adopts a different and more open conformation of the β-helix when compared to the AD fold. This open conformation results in a cavity within the β-helix that encloses an additional density which is not connected to the main chain of tau. CTE type I and type II filaments differ from each other by their inter-protofilament packing. In CTE type I filaments, protofilaments form an antiparallel steric zipper at residues 324-329. In CTE type II filaments, the protofilament interface comprises residues 331-338 and is identical to the interface adopted by AD PHFs.

Tau filaments extracted from the brains of individuals with PiD revealed a fold, which is different from those of AD and CTE (Falcon, Zhang, Murzin, et al., 2018). The Pick fold adopts an elongated J-shaped conformation and consists of residues K254-F378 of 3R tau; unlike Alzheimer and CTE folds, the Pick fold also contains the C-terminal and two thirds of R1 in its core (Falcon, Zhang, Schweighauser, et al., 2018). The filament core consists
of 9 $\beta$-strands ($\beta1-\beta9$), which arrange into four cross-$\beta$ sheets, resulting in a two-layered fold. The PiD filaments consists of two filaments, the narrow Pick filament (NPF) and the wide Pick filament (WPF). The NPF consists of a single protofilament and the WPF of two identical protofilaments, which pack against each other symmetrically at the tip of the J comprising residues 322-324.

Tau filaments extracted from the brains of individuals with CBD revealed yet another different fold (Zhang, Tarutani, et al., 2020). The CBD fold comprises residues K274-E380 and adopts a four-layered arrangement consisting of 11 $\beta$-strands (Zhang, Tarutani, et al., 2020). Residues K274-E380 contains one residue of R1, the whole of R2, R3 and R4, as well as 12 amino acids after R4. Similar to what was observed for the CTE fold, the CBD fold encloses a non-proteinaceous molecule which is disconnected from the main chain of tau. This density is surrounded by the side chains of lysines at residues K290, K294 from R2 and K379 in the C-terminus after R4. The CBD filaments, like in PiD filaments, can arrange into single protofilaments, CBD type I, or two protofilaments, CBD type II, which pack symmetrically against each other through an anti-parallel steric zipper at residues 342-347 in R4.

Tau filaments extracted from the brains of individuals with AGD, ARTAG and $MAPT$ intron 10 mutations +16/+3 are identical. The AGD fold differs from all previously observed tau folds. Nevertheless, it comprises residues G273-D387 and adopts the same four-layered arrangement as the CBD fold. The AGD fold is identical to the CBD fold at residues 293-357 in R2 and R3. It differs from the CBD fold at residues 273-293 in R2 and 372-387 in the C-terminal segment. In the AGD fold, residues 273-293 in R2 and 372-387 in the C-terminal segment stack against each other, whereas in the CBD fold, only residues 274-279 and 374-380 stack against each other. The AGD fold is seven amino acid residues longer than the CBD fold. The AGD fold also contains a cavity between R2 and the C-terminal segment at residues 368-386, but this cavity is smaller than in the CBD fold. It encloses an additional unknown density and is oriented by lysines K294 from R2 and K370 from the region after R4. The AGD filaments can arrange into two types: type I filaments consist of a single protofilament, whereas AGD type II filaments consist of two protofilaments, that pack symmetrically against each other through an anti-parallel steric zipper at residues 342-347. The latter is identical to the CBD type II interface.

The PSP fold adopts a different conformation to the folds described so far but it comprises the same residues as the CBD fold and consists of 13 $\beta$-strands (Y. Shi, Zhang, et al., 2021).
The PSP fold comprises repeats R2, R3 and R4, which arrange into a three-layered structure. The R3–R4 interface contains a cavity with an unknown density that is oriented by lysines K317 and K321 from R3 and K340 from R4. PSP filaments consist of a single protofilament.

The GGT fold has a similar three-layered arrangement as the PSP fold and spans residues G272-R379, with 12 β-strands (Y. Shi, Zhang, et al., 2021). The GGT fold is similar to the PSP fold, but it differs in the chain turns at the PGGG motifs. Moreover, the C-terminal segment at residues 369-379 in the GGT fold points in the opposite direction when compared to the PSP fold. Like the PSP fold, an additional non-proteinaceous density was observed at the interface of R3 and R4 orienting lysines K317, K321 and K340. Filaments with the GGT fold come in three different types that differ in their inter-protofilament arrangements. Type I is made of a single protofilament, whereas types II and III consist of two identical protofilaments that are arranged asymmetrically. In type II, the interface is formed between the hairpin loop connecting R4 and the C-terminal segment of one protofilament and residues 299-301 from R2 of the second protofilament. In type III, the interface is formed by the hairpin loop connecting R4 and the C-terminal domain of one protofilament and residues 283-286 from R2 of the other protofilament.

Cryo-EM also identified a new type of filament in a case that had been diagnosed as PSP (Y. Shi, Zhang, et al., 2021). The discovery of a new tau fold together with the observation that specific folds defined distinct diseases thus far, suggested that this was a distinct disease. Neuropathological characterisation revealed basophilic neuronal inclusions, primarily in limbic brain regions. The new disease was named Limbic-predominant neuronal inclusion body 4R tauopathy, LNT. The tau filament structures extracted from LNT brain were named the GPT fold. The GPT fold adopts similar structures as the GGT and PSP folds (GGT-PSP-tauopathy) (Y. Shi, Zhang, et al., 2021). The GPT fold adopts a three-layered arrangement comprising residues G272-R379 which form 13 β-strands and contains the repeats R2, R3 and R4, as well as 13 amino acids after R4. The GPT fold has similar conformations as the GGT fold at residues 273-294, 312-346 and the hairpin loop at residues 356-378. Like the PSP and GGT folds, the R3-R4 interface contains a cavity which contains an unknown density. The GPT fold comes in two different types of filaments. GPT type I filaments consist of a single protofilament and GPT type II filaments of two symmetric protofilaments, with each protofilament adopting the same interface as in GGT type III filaments.

Although specific tau protofilament folds characterise different diseases, some diseases share the same fold. Thus, familial British and Danish dementia’s (FBD and FDD), pri-
mary age related tauopathy (PART), mutations in the APP gene and some individuals with Gerstmann-Sträussler-Scheinker (GSS) - which is caused by mutations in the PRNP gene, share the Alzheimer fold (Hallinan et al., 2021; Y. Shi, Zhang, et al., 2021). The CTE fold is also found in subacute sclerosing panencephalitis (SSPE), a largely inflammatory condition (Qi et al., 2023). The AGD fold is also found in ARTAG and MAPT intron 10 mutations +16/+3; the latter indicates that the relative overproduction of 4R tau is sufficient to give rise to the AGD fold (Y. Shi, Zhang, et al., 2021).

In most brain-derived tau filament structures, extra densities that are not attributed to tau, are observed. Most of the extra densities are on the outside of the ordered cores, facing the solvent; whereas others are buried inside the core of the filaments. The extra densities facing the solvent are often near lysines. In AD PHFs and in CTE type 1 and type 2 filaments, densities are observed near lysines 317 and 321. As mentioned above, a cavity at the β-helix of the CTE fold also comprises an unknown density. The amino acid composition of this cavity suggests that the molecule may be hydrophobic. In the CBD fold, a large internal cavity contains an extra density, which is oriented by three lysines and is therefore likely anionic. A density in the AGD fold is smaller and oriented by two lysines. In PSP and in GGT, the interface of R3–R4 contains a cavity with an extra density between lysines K317, K321 and K340. In GGT, this density is slightly larger because of a salt bridge between E338 with K331 (Y. Shi, Zhang, et al., 2021). It is unknown what molecules cause these extra densities and if they are required for assembling tau into filaments. It also remains to be shown whether these densities are comprised of specific molecules or merely need to meet certain biochemical criteria to support the filament fold. Additional methods, including mass spectrometry and high-resolution cryo-EM may help to establish the molecular identities of these these densities.

The cryo-EM structures of brain-derived tau filaments have revealed that specific tau folds or conformers are associated with different diseases. This has led to a structure-based classification of tauopathies (Y. Shi, Zhang, et al., 2021). The first level of classification is based on the isoform composition of the tau filaments. All known tau folds from diseased brain contain R3 and R4 as well as an additional 10-13 amino acids after R4. The folds differ in their incorporation of R2 and their N-terminal extensions. The AD and CTE folds consist of a mixture of 3R and 4R tau isoforms. The PiD fold consists of 3R tau and comprises the C-terminal and two thirds of R1. The CBD, GGT, PSP, GPT and AGD folds contain R2 in the core of the structures and are therefore 4R tauopathies. At the second level, tauopathies can be separated based on their distinct folds. In the 3R/4R mixed isoform folds, the CTE fold is
distinct from the AD fold. In the 3R isoform folds, the PiD fold is the only 3R tauopathy with a known filament structure. The 4R tauopathies can be divided into two classes. One class consists of the folds which arrange into three layers. The three-layered arrangement is adopted by the PSP, GGT and GPT folds. The second class consists of the folds which arrange into four layers. The four-layered arrangement is adopted by the CBD and the AGD folds. At the third level of classification, the folds can be distinguished at the amino acid residue level (Figure 1.5).

---

**Fig. 1.5 Tau folds**  
**a** Schematic representation of MAPT. Protein coding exons are indicated. The microtubule-binding repeats corresponding to amino acid residues 244-368 are shown in colour. R1 is shown in purple (244-274), R2 in blue (275-305), R3 in green (306-336) and R4 in yellow (337-368). The C-terminal domain is indicated in orange (369-441).  
**b** Schematic representation of the secondary structures of the tau folds. β-sheets are shown as arrows. Non-proteinaceous densities are shown in black.
1.6.2 Cryo-EM structures of α-synuclein filaments from human brain

The first cryo-EM structures of α-synuclein amyloid filaments from post mortem tissues were described for MSA (Schweighauser et al., 2020). They revealed two types of α-synuclein filaments, each consisting of two different protofilaments with 9-12 β-strands. The ratios of the two types of filaments differed amongst individuals. MSA filaments are asymmetric and comprise an N-terminal extended arm and a compact C-terminal body. The MSA type I filament consists of residues 14-94 in protofilament Ia (PFIa) with 12 β-strands and 21-99 in PFIb consisting of 10 β-strands. Type II filaments comprise residues 14-94 in PFIIa with 12 β-strands and 36-99 in PFIIb consisting of 9 β-strands. The four protofilaments differ from each other, but they share an L-shaped fold with a C-terminal body. The L-shaped topology has also been observed for in vitro generated α-synuclein filaments (Guerrero-Ferreira, Taylor, Mona, et al., 2018; Tuttle et al., 2016). At the interface of the protofilaments lies a non-proteinaceous density that is disconnected from the main chains. This density is oriented by two lysines from one protofilament at residues K43, K45 and lysine residue K58 from the opposing protofilament and is therefore probably anionic.

The structures of α-synuclein filaments extracted from PD, PDD and DLB brains share the same fold and comprise a single protofilament - the Lewy fold (Yang, Arseni, et al., 2022). The Lewy fold comprises residues 31-100 of α-synuclein and adopts a three-layered structure consisting of 9 β-sheets. Two layers are disrupted consisting of β-sheets 1-5 (residues 31-60) and β-sheets 6-8 (residues 61-86). The third layer consists of β9 comprising residues 87-101. A non-proteinaceous density is present in front of K32, K34, Y39, K43 and K45. It has a similar size to the density observed in MSA type I and type II filaments. The generality of the Lewy fold across PD, PDD and DLB indicates that these diseases lie on a continuum. It is possible that similar mechanisms underlie the misfolding of α-synuclein in these diseases. The structures of MSA type I and type II filaments are different from the Lewy fold, establishing the existence of distinct molecular conformers of assembled α-synuclein.

An individual with a heterozygous 21-nucleotide duplication in SNCA, was diagnosed with dementia at the age of 13 and died 2 years later. This disease was called juvenile-onset synucleinopathy (JOS) (Yang, Garringer, et al., 2022). The mutation results in a 7 amino acid insertion (MAAAEKT) after amino acid residue 22 of α-synuclein. Immunoblotting for sarkosyl insoluble α-synuclein from this brain resulted in two distinct bands, one corresponding to wildtype α-synuclein and the other to α-synuclein with the 7 amino acid insertion, travelling at 15 and 16 kDa respectively. This was confirmed by trypsin-digest
mass-spectrometry. As such, JOS filaments comprise both WT and mutant α-synuclein. The cryo-EM structure of filaments extracted from the individual with JOS, revealed the JOS fold. The JOS fold is distinct from the MSA and Lewy body folds, suggesting that this disease is a different synucleinopathy. The JOS fold comprises residues 36-100 of the α-synuclein sequence. The majority of the filaments are singlets, but in some doublets, two identical protofilaments pack against each other. Two protein islands with unidentified sequences were also observed. The structure does not explain how the mutant may affect the formation of the JOS fold. It is possible that the mutant is more aggregation-prone and has the ability to incorporate both wild-type and mutant α-synucleins into the filaments through templated seeding.

As observed for tau, cryo-EM structures of brain-derived α-synuclein filaments have revealed that specific α-synuclein folds define distinct diseases (Figure 1.6).

Fig. 1.6 α-Synuclein folds

a Schematic representation of the α-synuclein sequence. The N-terminal domain is shown in yellow (1-60) the non-amyloid component in green (61-95) and the C-terminus in blue (96-140).
b Schematic representation of the secondary structures of α-synuclein folds. The Lewy fold is found in PD, PDD and DLB. β-sheets are shown as arrows. Non-proteinaceous molecules are shown in black. Unidentified proteinaceous species are shown in light grey.
1.7 Studying amyloid formation

Since amyloid pathology is characteristic of most neurodegenerative diseases, a better understanding of how it forms will be essential for the development of mechanism-based therapies. Post mortem tissue samples represent the end-stage of a long disease process and cannot be perturbed experimentally. For studying the molecular mechanisms of amyloid assembly, in vitro assembly with recombinant protein provides unique opportunities.

In vitro assembly of recombinant proteins into amyloid filaments requires the purification of the proteins of interest and subsequent incubation to promote self-assembly. Purification protocols for tau and α-synuclein have been described (see Methods). A high purity of the protein is essential for reproducible and quantitative studies (Linse, 2020). The addition of tags may interfere with experiments, as amino acid composition and sequence are important for filament formation (Anderson & Webb, 2011; Bousset et al., 2020). Several techniques have been used to study in vitro amyloid assembly of tau and α-synuclein.

1.7.1 Solution state Nuclear Magnetic Resonance

The monomeric state of a protein can be studied by solution-state nuclear magnetic resonance (NMR). This can provide information of the overall dynamics of the protein and its conformational landscape. The dynamics of full-length monomeric tau, as well as α-synuclein, has been studied using solution-state NMR (Croke et al., 2011; Lippens, Landrieu, et al., 2016; Lippens, Sillen, et al., 2006; Mukrasch et al., 2005; Schwalbe et al., 2014).

A complete assignment of the backbone resonances of tau is challenging due to its intrinsically disordered protein (IDP) characteristics and relatively large molecular weight. Nevertheless, combining studies of tau fragments comprising the repeats and full-length tau has shown that residues 275-280 (PHF6*) and 306-311 (PHF6) in R2 and R3, respectively, are in equilibrium of monomer-dimer transitions and adopt an extended conformation (Peterson et al., 2008). In full-length tau, P1 and P2 domains (151-243) adopt polyproline II helices and the C-terminal tail forms α-helices and a coiled coil (Landrieu et al., 2010; Lippens, Sillen, et al., 2006; Smet et al., 2004).

α-Synuclein shows three regions with distinct dynamical behaviour. The N-terminal region adopts an amphipathic α-helix at residues 1-25, which provides an anchor for its interaction with membranes, a central region at residues 26-97 has intermediate dynamic properties and the C-terminal domain at residues 98-140 is flexible (Bodner et al., 2009).
1.7 Studying amyloid formation

For solution-state NMR experiments, the buffer conditions have been shown to be crucial (Karamanos et al., 2015). The pH of the buffer has the ability to modify the charged states of the proteins. A pH of 4-6 has been shown to promote the assembly of both tau and α-synuclein (Hoyer et al., 2002; Lundvig et al., 2005; Zhang, Falcon, et al., 2019) however, their structures are different to those observed in disease.

1.7.2 Monitoring the kinetics of amyloid formation with Thioflavin T

The ability of a protein to assemble into an amyloid is dependent on an environment that favours a conformational switch in the monomeric soluble protein, which can then recruit and convert other soluble molecules into the elongating filament. The growth of the filament can be monitored using fluorescent dyes which bind to amyloid proteins (Biancalana & Koide, 2010). Thioflavin T (ThT) is the most widely used compound for monitoring amyloid kinetics.

The kinetic profile measured by ThT fluorescence during amyloid assembly is sigmoidal and characterised by a lag phase, followed by a growth phase and a plateau. The two flat regions of the curve are the lag and the plateau phases and the steep transition is the growth phase (Figure 1.7) (Arosio et al., 2015). Based on the observation of non-linear growth kinetics, researchers have proposed a secondary mechanism of filament assembly. This assumes a direct correlation between ThT intensity and filament concentration. After primary nucleation of a filament, existing soluble monomers can nucleate from the surfaces of the filament in a self-sustaining manner, which leads to a positive and cooperative feedback loop of filament growth (Cohen et al., 2013; Knowles et al., 2009; Törnquist et al., 2018). This is known as secondary nucleation. The plateau is reached when all monomers have assembled into filaments. However, because the molecular mechanisms of ThT binding to the filaments are not well understood, the theoretical understanding of filament formation is at the disposal of the experiment itself. A molecular mechanism of how ThT binds to filaments will be required for us to model ThT amyloid kinetics accurately.

1.7.3 Molecular mechanisms of amyloid filament formation

Different types of aggregates are observed during amyloid assembly. Observations made by circular dichromism (CD), negative stain EM and atomic force microscopy (AFM) have reported small oligomeric species prior to the formation of end-stage filaments (Bode et
Fig. 1.7 Schematic of the kinetic profile for amyloid filament in the presence of ThT. The lag phase (light blue) has no fluorescence and is constituted of monomers in the solution, shown as small blue spheres. The growth phase (dark blue) is characterised by a rapid increase in fluorescence and is constituted by a mixture of monomers and fibrils in the solution, shown as red sticks. The plateau phase (red) is when all the monomers have formed amyloid filaments. The Y axis shows the fluorescence and the X axis shows the time course of the experiment.

The oligomers are described as small pre-aggregates, consisting of a few protein molecules. The morphology of the oligomers is unclear, though it has been proposed that they may exist as globular-like or small soluble fibrils. It has also been reported that small filaments or protofibrils may serve as transitional stages in the formation of mature filaments (Harper et al., 1997; Walsh et al., 1999). Protofibrils are on-pathway structures to become mature filaments and are not to be confused with the protofilaments of mature filaments, which are end-stage. Preparations using synthetic peptides of PrP have shown to form small bead-like structures by AFM and TEM. The PrP peptide and SH3 amyloid have been shown to form protofibrillar intermediates, which had the ability to detach and reanneal β-sheets within the filament. This mechanism was termed strand-recycling (Carulla et al., 2005; Petty & Decatur, 2005). It showed that protofibrils are structurally dynamic and have the ability to rearrange within a fibril, before maturing into end-stage filaments. Short-lived oligomeric and protofibril intermediates are hypothesised to be the main toxic species in amyloid pathologies. Recently, the FDA approved a humanised monoclonal antibody, Lecanemab, which was raised against protofibrils of synthetic Aβ 40 E22G filaments (van Dyck et al., 2023). These were shown to bind to mature Aβ 42 filaments from AD brain. It remains to be shown whether Lecanemab selectively binds to distinct Aβ filaments. Further investigation is necessary to elucidate the molecular mechanisms and
1.7 Studying amyloid formation

structures of preliminary aggregates that precede the formation of mature amyloid filaments.

1.7.4 *In vitro* assembly

The self-assembly of proteins into amyloids requires conditions which promote protein-protein interactions. Chemical and physical parameters can be modified for this purpose. Chemical parameters such as the buffer composition, pH and the concentration of the protein will affect assembly. The pH of the buffer can alter the charged state of the protein. The composition of the buffer, i.e., the presence of molecules or crowding reagents, may affect interactions with the protein. Physical parameters, such as temperature and shaking conditions, are used to accelerate self-assembly. Surface conditions and the air-water interface have been shown to affect assembly (Moores *et al*., 2011). It is also important to consider that some compounds have expiration times and can alter their chemical composition during a reaction. All components have the ability to alter the proteins conformation and its ability to assemble.

Self-assembly of tau and α-synuclein has been studied since their discovery as the main components of inclusions in disease. Full-length tau is highly soluble and its *in vitro* assembly into filaments requires the addition of anionic cofactors such as RNA, sulphated glycosaminoglycans and certain fatty acids (Friedhoff *et al*., 1998; Goedert, Jakes, *et al*., 1996; Kampers *et al*., 1996; Pérez *et al*., 1996; Wilson & Binder, 1997). The cryo-EM structures of full-length 3R and 4R recombinant tau assembled into filaments by heparin, led to the formation of polymorphic filaments with structures that were different from those extracted from human brains (Zhang, Falcon, *et al*., 2019). Tau constructs consisting only of the microtubule-binding repeat region assemble into filaments in the absence of cofactors (R. Crowther *et al*., 1992; Wille *et al*., 1992). Fragments comprising the ordered cores of tau filaments from AD/CTE (residues 306-378), PiD (254-378, excluding R2), CBD and PSP (residues 274-380) also assemble into filaments (Carlomagno *et al*., 2021). Whether the fragments comprising the ordered cores of the disease filaments resemble those from human brains remains to be determined. Two commonly used constructs are the K18 and K19 fragments, which comprise the the repeats and four additional amino acids (244-372) in K18; and the same residues, but excluding R2 in K19). The K18 and K19 constructs spontaneously assemble into filaments and are seeding-active (Gustke *et al*., 1994; W. Li, Sperry, *et al*., 2009; Mukrasch *et al*., 2005; Shammas *et al*., 2015; von Bergen *et al*., 2000; Yu *et al*., 2012). However, all tau filament structures from human brain include residues beyond 372 in the core and as such the filament structures of K18 and K19 cannot be the same as those in disease. The 297-391 fragment (dGAE) was identified as the proteolytically
Introduction

stable core of PHFs from AD and assembles spontaneously into filaments with morphologies like PHFs by negative stain EM (Al-Hilaly, Pollack, et al., 2017; Wischik et al., 1988).

Full length α-synuclein spontaneously assembles into filaments. Familial mutations E46K, H50Q and A53T and truncations of the C-terminus (1-120) in the α-synuclein sequence have been found to accelerate filament assembly (Conway, Harper, et al., 1998; R. A. Crowther, Jakes, et al., 1998; Ghosh et al., 2013). But, the familial mutation A30P decelerates filament formation (Conway, S. J. Lee, et al., 2000; Krüger et al., 1998). Cryo-EM structures of α-synuclein filaments assembled from recombinant proteins in vitro are polymorphic and different from those extracted from human brains, although certain secondary structure elements, such as the three-layered L-shaped motif, are present in both (Guerrero-Ferreira, Taylor, Mona, et al., 2018; Schweighauser et al., 2020; Tuttle et al., 2016; Yang, Garringer, et al., 2022).

1.7.5 Seeded aggregation

The concept of templated seeding, which was first established as a mechanism for prion diseases (Frost & Diamond, 2010; Jucker & Walker, 2013), is commonly used to generate filament structures in vitro. Here, small amounts of amyloids (the seeds) are introduced into a solution consisting of monomeric protein. Seeds will nucleate assembly of the monomeric protein into amyloids, leading to the amplification of amyloid formation beyond levels that would have occurred spontaneously.

Protein Misfolding Cyclic Amplification (PMCA) and Real-time quaking-induced conversion (RT-QuIC) are procedures that are commonly used for templated seeding (Orru et al., 2012; Saborio et al., 2001; Soto et al., 2002). ThT fluorescence is used to monitor the reaction. PMCA performs cyclic rounds of sonication and re-seeding. Filaments are quantified by ThT fluorescence after each round of amplification at individual time-points. RT-QuIC employs shake-rest cycles of a seeded reaction in the presence of ThT and its fluorescence is monitored in real-time. Both methods use physical parameters (sonication or shaking) to fragment fibrils. RT-QuIC uses ThT in the assembly reaction, which may affect the aggregation of amyloids. As such, several replicates without ThT are used for downstream biochemical analysis and applications. RT-QuIC is used in the clinic for the detection of vCJD caused by the PrPs and has been reported to have a sensitivity and accuracy of 91% and 98% respectively (McGuire et al., 2012). PMCA and RT-QuIC of α-synuclein have been reported to discriminate between MSA and PD (Shahnawaz, Mukherjee, et al., 2012).
The K18 and K19 fragments of tau showed different RT-QuIC profiles for the different tauopathies (Metrick et al., 2020). Whether the structures amplified using PMCA or RT-QuIC replicate the folds from human brains is unknown. It is often assumed that the newly formed amyloids adopt the same structures as the seeds. However, because the tau constructs K18 and K19 are RT-QuIC seeding competent but do not have the sequence requirements for adopting the correct structures, it is questionable whether if the filaments have any disease relevance.
1.8 Overview of this dissertation

I summarised above what is known about the biochemical and structural characteristics of tau and α-synuclein. There are many outstanding questions. Cryo-EM studies have revealed the existence of specific tau and α-synuclein conformers that are associated with distinct tauopathies and synucleinopathies, respectively. What is the relationship between conformers and the development of disease? The mechanisms by which filaments spread within the human brain through prion-like mechanisms remain unclear. Several in vitro amyloid assembly systems are used to study these diseases. Do the current in vitro assembly assays reproduce the mechanisms that occur in disease; and do they replicate the conformers that are associated with disease? Ultimately, to develop effective mechanism-based treatments for these diseases, it is essential to understand the molecular mechanisms underlying amyloid filament formation.

In this dissertation, I describe the development of in vitro amyloid assembly reactions that aim to replicate disease-relevant structures, in order to understand the mechanisms of amyloid filament formation. Below, I provide a brief overview of the main results chapters.

1.8.1 Chapter 2: Seeded aggregation of recombinant α-synuclein

In 2020, it was reported that PMCA assembly assays can discriminate between MSA and PD when using cerebrospinal Fluid (CSF) and brain seeds extracted post mortem (Shahnawaz, Mukherjee, et al., 2020). This paper also reported differences in amplified filaments by their seeding ability, fluorescence intensity and filament morphology as observed by negative stain EM. In this chapter, I use cryo-EM to assess whether the seeded assembly of recombinant α-synuclein with seeds from MSA brains yields the same amyloid structures as those of the initial seeds. Using seeds extracted from the brains of three individuals with MSA, I find that the filament structures formed were different from those of the seeds. These results suggest that additional co-factors or substrate modifications may be required for the faithful amplification of disease-relevant structures. The results of this chapter were published as Lövestam, et al., 2021.
1.8.2 Chapter 3: Spontaneous in vitro assembly with recombinant tau

When I started this project in 2021, there was no known in vitro system that could form any of the amyloid structures observed in diseased brains. I focused on developing in vitro assembly reactions with recombinant tau. The assembly of wild-type full-length tau requires the addition of anionic compounds, but the structures of heparin-induced tau filaments were shown to be different from those in disease (Zhang, Falcon, et al., 2019). I show that N- and C-terminally truncated tau constructs, under specific conditions, can form filaments that are identical to those in disease, including Alzheimer PHFs, and type 1 and type 2 filaments from CTE. Additional molecules or post-translational modifications and identification of molecules may be required for making the structures from some of the other diseases. The results in this chapter were published as Lövestam, et al., 2022. This work also led to the development of high-throughput cryo-EM structure determination methods for amyloids. These methodological developments are described in the Methods section and were published as Lövestam & S. H. Scheres, 2022.

1.8.3 Chapter 4: Time-resolved studies of tau assembly

I started this project in 2022, when I was developing in vitro assembly systems for tau. I observed that filaments that were on-pathway to mature PHFs had different structures at during assembly. As such, I started to investigate the process of tau filament assembly into PHFs and CTE filaments using cryo-EM. I show that residues $^{302}$GGSVQIVYK$^{315}$PVDLS are essential for initial dimerisation, and use cryo-EM to solve intermediate structures that are on-pathway to forming PHFs and CTE filaments. The results described in this chapter will be published after submission of this thesis.
Chapter 2

Seeded assembly of recombinant α-Synuclein

2.1 Introduction

At the time of this study, in 2019, Manuel Schweighauser and Yang Shi were preparing a manuscript on the structures of α-synuclein filaments that were extracted from the post mortem brains of five individuals with Multiple System Atrophy (MSA) Schweighauser et al., 2020. MSA is a sporadic disease defined by regional nerve cell loss and the accumulation of α-synuclein inclusions in oligodendrocytes and some nerve cells. The structures of α-synuclein from five individuals with MSA were conserved, consisting of two types of filaments, each of which comprises two different protofilaments. At the interface for both types of filaments lies a non-proteinaceous molecule (Figure 2.1). This study also showed that filaments extracted from individuals with DLB are different to those of MSA, indicating that a similar trend exists for synucleinopathies as was observed previously for tauopathies: specific folds characterise different diseases. Furthermore, this study showed that many structures of in vitro assembled α-synuclein structures, which had been determined using solid-state NMR or cryo-EM, were different to those extracted from post-mortem brains. Different methods or conditions are thus required to reproduce the structures of α-synuclein filaments from MSA, in vitro. Some researchers have used in vitro seeded assembly, in which a small amount of filaments from post mortem tissue is used to seed the assembly of recombinant protein. Often, these studies implicitly assume that the seeded structures have the same structure as the initial seeds. In this chapter, I tested this hypothesis. I used the same extractions of α-synuclein filaments from the brains of three of the same individuals as were used in the study by Manuel Schweighauser and Yang Shi to seed the assembly of
recombinantly purified WT α-synuclein, and then solved the resulting structures by cryo-EM. The study was published in Lövestam et al. FEBS Open Bio (2021).

Figures and text of this publication are used in this chapter. Post mortem brain samples were originally received from our collaborator in Japan; Masato Hasegawa.

Fig. 2.1 Cryo-EM structures of type I and type II of α-synuclein filaments from MSA brains

a Schematic representation of the α-synuclein protein sequence. The N-terminal domain is shown in yellow (1-59) the non-amyloid component is shown in green (60-95) and the C-terminus is shown in blue (96-140).
b Cryo-EM of maps are shown in transparent grey and the atomic models are coloured with the same colours as in a for MSA filament of type I and type II. PF; protofilament
2.2 Results

2.2.1 Seeding recombinant wild-type α-synuclein with seeds from three cases of MSA

I seeded the in vitro assembly of recombinant wild-type (WT) human α-synuclein with filament preparations from the putamen of three cases of MSA (see Methods). The cryo-EM structures of the filaments from these cases are known (cases 1, 2 and 5 in ref Schweighauser et al., 2020). They contain variable proportions of type I and type II MSA filaments, with I:II ratios of 80:20 for case 1; 20:80 for case 2; and 0:100 for case 5.

I monitored the kinetics of aggregation using thioflavin T (ThT). The assembly conditions were as described (Shahnawaz, Mukherjee, et al., 2020), using 100 mM piperazine-N,N-bis(2-ethanesulfonic acid) (PIPES) and 500 mM NaCl at 37 °C, pH 6.5. Upon addition of seeds, I observed a clear lag phase of 20–40 h, before fluorescence increased rapidly and plateaued after 30–60 h (Figure 2.2). Case 5 seeds were faster at seeding recombinant α-synuclein and resulted in higher fluorescence intensities than seeds from cases 1 and 2. No increase in fluorescence was observed in the absence of seeds. Negative-stain EM confirmed the presence of abundant filaments after incubation with MSA seeds. Filaments were stable in extraction buffers which was used for purifying the filaments from the individuals with MSA. Some filaments were stable in 2% sarkosyl whereas all filaments were solubilised upon addition of in 1% SDS (Figure 2.4).

2.2.2 Cryo-EM imaging of seeded α-synuclein filaments

I subsequently used cryo-EM to image the filaments formed following incubation of recombinant α-synuclein with seeds from each MSA case. Visual inspection of micrographs of filaments from experiments that used seeds from MSA cases 1 and 2 indicated the presence of two main filament types, which I called type 1 and type 2. Type 1 filaments have an average cross-over distance of 800 Å and widths of 60–130 Å; type 2 filaments have a cross-over distance of 900 Å and widths of 80–130 Å. Furthermore, I also observed straight filaments with no observable twist. It is unclear if these filaments correspond to types 1 or 2 that untwisted because of sample preparation artefacts, such as interactions with the air-water interface, or if they represent additional filament types. Due to the lack of twist, I was unable to solve the structures of these filaments.
Fig. 2.2 Seeded assembly of recombinant α-synuclein with filament preparations from MSA brains. 

a Recombinant wild-type human α-synuclein was mixed with sonicated MSA seeds in 100 mM PIPES, 500 mM NaCl, 0.05% NaN3, pH 6.5. Seeds had variable ratios of type I and type II filaments. 

b Assembly was monitored by thioflavin T fluorescence of recombinant α-synuclein in the presence of MSA filament seeds from case 1 (green), case 2 (blue) and case 5 (red). Controls (grey) were without seeds. Curves represent the mean and dots correspond to the values in each experiment, (n = 5).

c Negative stain micrographs of α-synuclein filaments after seeded assembly (scale bar = 200 nm).

d Cryo-EM 2D class averages in boxes spanning 825 Å of the types of filaments observed. Assembly with seeds from MSA cases 1 and 2 gave rise to type 1 and type 2 filaments. Type 3 filaments formed when the seeds were from MSA case 5.
Fig. 2.3 **Purified filaments from MSA case 5.**  

- a Sarkosyl insoluble pellet before sonication; and  
- b after sonication (scale bar = 200 nm).  
- c Cryo-EM 2D class averages of MSA case 5 purified filaments before sonication and  
- d after sonication.
Seeded assembly of recombinant α-Synuclein

Fig. 2.4 Biochemical characterisation of seeded material Seeded filaments were treated with 2% sarkosyl (Sar), extraction buffer (20 mM Tris pH 7.4 15% sucrose 800 mM NaCl 5 mM EDTA) (Extr) or 1% SDS. After treatment, the pellets (P) and supernatants (S) were analysed. a 4-20% SDS-PAGE Tris-Glycine gel of the seeded filaments. The cases used for seeding are shown above the gel. b Negative stain EM of the treated pellets. Cases are indicated on the left. (scale bar = 200 nm).
Fig. 2.5 Cryo-EM structures of type 1 and type 2 α-synuclein filaments with protofilament fold A assembled using seeds from MSA case 2.

a Primary sequence of α-synuclein with β-strands and loop regions shown from dark blue (N-terminal) to light blue (C-terminal).

b Central slice of the 3D map for type 1 filaments with protofilament fold A.

c Cryo-EM density (transparent grey) and fitted atomic model (with the same colour scheme as in a for type 1 filaments).

d Cartoon view of three successive rungs of the type 1 filament.

e-g As b-d, but for type 2 filaments.
Seeded assembly of recombinant α-Synuclein

Two-dimensional classification readily separated type 1 and type 2 filaments for further processing and indicated that both types are 2-fold symmetric along their helical axis (Figure 2.2d). Further 3D classification revealed that type 1 and type 2 filaments occurred in two variants in the data set of filaments that formed with seeds from MSA case 1. They are characterised by small differences in the protofilament folds. I called the predominant protofilament ‘fold A’ and the minor protofilament ‘fold B’. I could not identify protofilaments with fold B when seeds from MSA case 2 were used. Using helical reconstruction in RELION, I determined cryo-EM structures of type 1 and type 2 filaments with only protofilament fold A to 3.4 Å resolution (Figure 2.5; Figure appendix A.1 and A.2). Reconstructions of type 1 and type 2 filaments with two protofilaments of fold B, or with one protofilament of fold A and another protofilament of fold B, were solved to resolutions of 3.4–4.1 Å (Figure 2.6; Figure appendix A.2). Reconstructions of filaments containing protofilaments of fold B were less well defined than those of filaments with two protofilaments of fold A. Assembly with seeds from MSA case 5 resulted almost exclusively in the formation of a different type of filament, which I called type 3. Type 3 filaments are thinner, more bendy and longer than filaments of types 1 and 2. Type 3 filaments have a crossover of 900 Å and widths of 55–65 Å. I solved their structure to 3.2 Å resolution (Figures 2.12; Figure appendix A.2 and A.3). A minority of filaments (< 2%) comprised a doublet of the type 3 filaments. Throughout this section, I will use blue colours for fold A and green for fold B of type 1 and type 2 filaments, and I will use purple for type 3 filaments.

2.2.3 Cryo-EM structures of type 1 and type 2 α-synuclein filaments

Most type 1 and type 2 filaments that formed with seeds from MSA case 1, and all the filaments that formed with seeds from MSA case 2, consisted of two protofilaments of fold A that were related by C2 symmetry. Filaments of types 1 and 2 differed in their inter-protofilament packing (Figure 2.5). In type 1 filaments, two salt bridges between E46 and K58 held the protofilaments together, by creating a large solvent-filled channel. The inter-protofilament interface in type 2 filaments was formed by two salt bridges between K45 and E46 of each protofilament. The smeared reconstructed densities at the points furthest away from the helical axis suggest that the inter-protofilament interface of type 2 filaments is more flexible than that of type 1 filaments. Protofilament fold A consists of 8 β-sheets: β1-6 form a roughly Z-shaped hairpin-like structure, with glycines or KTK motifs between the β-sheets at the bends; β7-8 fold back against β4, leaving a small triangular cavity between β5, β6 and β7. This fold is unlike any of those of the MSA type I and type II protofilaments. It is almost identical to the protofilament fold that was reported for in vitro aggregated recombinant E46K α-synuclein, although the inter-protofilament interface was
different from the interfaces observed here for type 1 and type 2 filaments (Figure 2.5 and 2.6; 2.8 and 2.9) (Boyer et al., 2019). A minority of type 1 and 2 filaments that formed with seeds from MSA case 1 consisted of two symmetry-related copies of protofilaments with fold B. Although the reconstructions of type 1 and type 2 filaments with two protofilaments of fold B (Figure 2.6 and A.2) were less well defined than those for filaments with two protofilaments of fold A, the maps revealed that fold B is nearly identical with the structure of filaments assembled from wild-type recombinant α-synuclein. As such I could build and refine an atomic model for the protofilaments with fold B. The resulting model from the type 2 filament has a root-mean-square deviation (r.m.s.d.) of 1.38 Å with the structure of assembled wild-type α-synuclein (Figure 2.8). Again, protofilament fold B was unlike any of the four protofilaments from MSA type I and type II filaments. An asymmetric reconstruction from a subset of the images suggested that asymmetric type 2 filaments may also form from one protofilament with fold A and one protofilament with fold B (Figure 2.6H). However, one cannot exclude the possibility that this reconstruction is an artefact arising from sub-optimal classification of filament segments. Folds A and B are almost identical at residues G36-V55, and V63-A78, with some flexibility in the β-turn at residues E57-E61. Comparing the more compact fold B to fold A, a flip in K80 from the hydrophobic core towards the solvent results in a sharp turn at T81 and a shift by three residues in the packing of β4 against β7 (Figure 2.6 and 2.10).

2.2.4 Cryo-EM structure of type 3 α-synuclein filaments

Type 3 filaments consist of a single protofilament that extends from G36-Q99 and comprises 10 β-sheets (β1-10) (Figure 2.11). Residues 46–99 form a Greek key motif, as described before, with a salt bridge between E46 and K80. This motif is preceded by a β-arch formed by residues Y39-T44 and Y39-E46. The density between residues 36 and 39 is more smeared. Two stretches of elongated, smeared densities, possibly originating from parts of the N-terminus of α-synuclein, are observed in front of β1 in the β-arch and β4 in the Greek key motif. An additional fuzzy density is observed in front of the side chains of K43, K45 and H50. It is likely that this density is anionic, to compensate the charges of the lysines and histidine. Whereas filament types 1 and 2 did not resemble the four protofilaments observed in MSA, type 3 filaments were almost identical to protofilament IIB2, with an r.m.s.d. between atomic coordinates of 1.02 Å (Figure 2.12). However, in MSA filaments, K58 is flipped away from the core of the protofilament to form a salt bridge with T33 of the opposing protofilament, whereas K58 forms part of the protofilament core in type 3 filaments. Minor rearrangements occur near V40, which is also involved in inter/protofilament packing in MSA filaments. Interestingly, the position of the density of the unidentified co-factor at
Fig. 2.6 Cryo-EM structures of type 1 and type 2 filaments with protofilament fold B assembled using seeds from MSA case 1. a Primary sequence of α-synuclein with β-strands and loop regions shown from dark green (N-terminal) to light green (C-terminal). b Central slice of the 3D map for type 1 filaments with protofilament fold B. c Cryo-EM density (transparent grey) and fitted atomic model (with the same colour scheme as in a for type 1 filaments. d Cartoon view of three successive rungs of the type 1 filament. e-g As b-d, but for type 2 filaments. h-j As (b-d), but for the putative type 2 filament that contains a mixture of protofilament folds A and B.
Fig. 2.7 Cryo-EM structures of type 1 and type 2 filaments with protofilament fold A assembled using seeds from MSA case 1. 

(a) Central slice of the 3D map for type 1 filaments.
(b) Side view of the 3D reconstruction of type 1 filaments.
(c-d) As (a-b), but for type 2 filaments.
Fig. 2.8 Comparison of protofilament A with PDB-entry 6SSX of recombinant wild-type α-synuclein. a Atomic model of protofilament fold B (green) overlaid with one protofilament from PDB entry 6SSX (grey). b Comparison of the interface between two protofilaments with fold B in type 1 filaments and those from PDB entry 6UFR, with the same colour scheme as in a. c Zoomed-in view of the interface, with salt bridges between K45 and E47 in PDB-entry 6UPR and between E46 and K58 in type 1 filaments highlighted in grey and green, respectively.
Fig. 2.9 **Comparison of protofilament fold A with PDB-entry 6UFR of assembled recombinant E46K α-synuclein.** 

a. Atomic model of protofilament fold A (blue) overlaid with one protofilament from PDB entry 6UFR (grey).

b. Comparison of the interface between two protofilaments with fold A in type 1 filaments and those from PDB entry 6UFR, with the same colour scheme as in a.

c. Zoomed-in view of the interface, with salt bridges between K45 and E47 in PDB entry 6UFR and between E46 and K58 in type 1 filaments highlighted in grey and blue, respectively.
Fig. 2.10 Comparison of protofilament folds A and B. 

**a** Atomic model of protofilament fold A (blue) overlaid with protofilament fold B (green) 

**b,c** As in **a**, but showing all-atom representation for different residues. 

**d,e** Schematic representations of protofilament folds A and B. Each amino acid residue is represented with its one-letter code in a circle. Positively charged amino acids are shown in blue, negatively charged ones in red, polar ones in green, hydrophobic ones in white, and glycines in pink.
the inter-protofilament interface of type II filaments coincides with the fuzzy density in front of K43, K45 and H50. Type 3 filaments are almost identical to the narrow protofilament formed upon in vitro assembly of recombinant H50Q α-synuclein, with an r.m.s.d. between atomic coordinates of 0.62 Å (Figure 2.12 and 2.14 (Boyer et al., 2019)).

Fig. 2.11 Cryo-EM structure of type 3 filaments assembled using seeds from MSA case 5. a Primary amino acid sequence of α-synuclein with β-strands and loop regions shown from dark violet (N-terminal) to light pink (C-terminal) b Central slice of the 3D map for the type 3 filament c Cryo-EM density (transparent grey) and the fitted atomic model (with the same colour scheme as in a. d Cartoon view of three successive rungs of the type 3 filament.

2.2.5 Cryo-EM structures of α-synuclein filaments from second-generation seeded aggregation

To further explore the effects of buffer conditions on seeded aggregation, I incubated seeds from MSA case 5 with recombinant human α-synuclein in phosphate-buffered saline (PBS). It’s been previously shown that the density for the additional molecules at the interface between protofilaments in our reconstructions of MSA filaments overlaps with similar densities in reconstructions of in vitro aggregated recombinant α-synuclein, which have been attributed to phosphate ions. Since the additional density in MSA filaments could accommodate two phosphate ions, I supplemented PBS with 1 mM of pyrophosphate. However, by negative-stain imaging, the seeded assemblies were indistinguishable from those formed using PBS without pyrophosphate. I then performed second-generation seeded assembly, in which the aggregates from the assembly in PBS-pyrophosphate were used as seed. Cryo-EM structure determination of the seeded assemblies confirmed the faithful propagation of type 3 filaments, with a larger proportion of type 3 doublet filaments (5%) (Figure 2.11 and 2.13).
Seeded assembly of recombinant α-Synuclein

Fig. 2.12 Comparison of type 3 α-synuclein filaments with protofilament IIB from MSA case 5

a Atomic model of the type 3 filament (purple) overlaid with the model of protofilament IIB2 from MSA case 5. The additional density at the protofilament interface of MSA type II filaments is shown in orange. b Cartoon view of one rung of type 3 filaments overlaid with one rung of protofilament IIB and three rungs of protofilament IIA of MSA case 5. Residues on MSA protofilament IIA that interact with the rung of protofilament IIB shown are highlighted with sticks. c Close up all-atom view of the hydrogen-bonding network (yellow dashed) between K58, E61 and T72 in type 3 filaments. d As in c, but for protofilaments IIA and IIB in MSA filaments.
2.2 Results

Fig. 2.13 Second-generation type 3 filaments. a Central slice of the 3D map of the type 3 filaments from the second generation of seeding. b Side view of the 3D reconstruction of the same type 3 filaments. c-d As in a-b, but for the doublets of type 3 filaments.

Fig. 2.14 Comparison of type 3 filament with PDB entry 6PEO of assembled recombinant H50Q α-synuclein. a All-atom view of the type 3 filament (purple) aligned with PDB-entry 6PEO (grey). b Schematic representation of the type 3 filament.
2.3 Discussion

I show here that the structures of the seeded assemblies of wild-type recombinant human \( \alpha \)-synuclein differ from those of seeds that were extracted from the brains of individuals with MSA (Figure 2.15). I used the assembly conditions of Shahnawaz et al. who reported that PMCA, using CSF as seed and recombinant \( \alpha \)-synuclein as substrate, can discriminate between PD and MSA. It remains to be seen if \( \alpha \)-synuclein seeds from PD brain yield structures that are different from those described here. Nevertheless, these results raise important questions for the study of amyloid structures and prion processes. Amyloid filaments are structurally versatile, with the same amino acid sequences being able to adopt different structures. Moreover, the cryo-EM structures of tau, \( \beta \)-amyloid and \( \alpha \)-synuclein filaments from human brain are different from those of recombinant proteins assembled in vitro Arseni et al., 2021; Falcon, Zhang, Murzin, et al., 2018; Y. Shi, Zhang, et al., 2021; Yang, Arseni, et al., 2022. These findings demonstrate that, even when using brain-derived filament preparations to seed in vitro assembly, the resulting structures do not necessarily replicate those of the seeds.

When using seeds from MSA cases 1 and 2, which contain a mixture of type I and type II filaments, and recombinant human \( \alpha \)-synuclein as substrate, I observed the formation of type 1 and type 2 filaments. When using seeds from MSA case 5, with only type II filaments, I observed the formation of filaments of type 3. These observations suggest that in seeded assemblies, type I filaments overshadow type II MSA filaments, despite the observation that seeds of case 5 resulted in a faster and stronger increase in thioflavin-T fluorescence compared to seeds from cases 1 and 2. The possibility that different conformational strains have different seeding potencies has implications for the interpretation of prion propagation assays.

It is commonly assumed that self-propagation of strains occurs through templated incorporation of monomers at the ends of amyloid filaments. Indeed, following sonication, \( \alpha \)-synuclein filaments had increased seeding potencies. However, it is unclear how this could explain the formation of type 1 and type 2 filaments with markedly different protofilament folds, when compared to MSA filaments. Each prion strain is believed to comprise a large number of conformationally distinct assemblies (also known as clouds), often with a dominant conformer that propagates under host selection. Previous work on tau and \( \alpha \)-synuclein assemblies has shown the presence of only one or two major filament types in the brains from patients at end-stage disease. It is possible that type 1 and type 2 filaments were present in the filament preparations from MSA brains, but not numerous enough to be detected by cryo-EM. It has previously been demonstrated that tau structures that only made up around
3% of filaments can be detected, indicating that, if present in MSA brains, type 1 and type 2 α-synuclein filaments are infrequent.

Type 3 filaments, which assembled from MSA type II seeds, fit the model of structural equivalence between seeds and seeded assemblies better than type 1 and type 2 filaments, because their structure overlaps almost completely with that of type IIB protofilaments from the putamen of patients with MSA. The additional cryo-EM densities at the inter-protofilament interfaces of type I and type II MSA filaments are negatively charged non-proteinaceous molecules. It is possible that the absence of these molecules in the seeded assembly experiments led to the formation of a structure that represents only half of the seed structures. These findings indicate that protofilament IIB, but not IIA, can form from recombinant α-synuclein through seeded assembly without added cofactor.

Abundant glial cytoplasmic inclusions (GCIs) in oligodendrocytes are the major neuropathological hallmark of MSA (Papp et al., 1989). Thus, differences in the cellular milieu between oligodendrocytes and other brain cells may play a role in the seeded aggregation of MSA filaments. Oligodendrocytes have been shown to transform misfolded α-synuclein into a GCI-like strain (Peng et al., 2018).

Fig. 2.15 Summary of MSA seeded aggregation experiments. Cartoon illustrations show the structures of MSA type I and type II filaments and their relative quantities in MSA cases 1, 2 and 5 at the top, and the products of seeded aggregation underneath.
Besides the possible incorporation of other molecules in α-synuclein filaments from human brain, it is also conceivable that recombinant α-synuclein is not able to form MSA filaments. Truncation and post-translational modifications of α-synuclein may be needed. In α-synuclein filament preparations from the putamen of patients with MSA, mass spectrometry identified N-terminal acetylation, C-terminal truncation, ubiquitination at K6 K12, K21, acetylation at K21 K23 K32 K34 K45 K58 K60 K80 and K96 and phosphorylation at Y39, T59, T64, T72 and T81. It’s unknown if these modifications occur prior to, during or after filament assembly, and if or how they may affect filament conformations. Assembly of recombinant wild-type human α-synuclein using seeds of α-synuclein phosphorylated at Y39 gave rise to filaments with a different fold from that of the seeds (Zhao et al., 2023). Moreover, C-terminal truncation of recombinant α-synuclein has been shown to promote filament assembly in vitro; inhibiting C-terminal truncation in transgenic mouse models of MSA has been reported to reduce pathology (Bassil et al., 2016; Sorrentino & Giasson, 2020). It has also been shown that interactions with lipids, DNA, RNA, iron and phosphate promote α-synuclein aggregation in vitro, and similar interactions could be important for the formation of MSA filaments in brain.

Identification of the factors that govern the replication of conformational prion strains will be essential for our understanding of propagation of the distinct proteinopathies. Similar methods are used for the detection of vCJD caused by the PrPs (McGuire et al., 2012). The relevance of structures of amyloids assembled from recombinant protein seeds and the results of self-propagation studies should be interpreted with care. The molecular mechanism of seeded assembly remains poorly understood. In the future, cryo-EM may represent a powerful tool to elucidate the structural mechanisms of templated seeding. This would be valuable in understanding the molecular mechanisms of templated seeding in neurodegeneration.
2.4 Conclusions

The *in vitro* seeded assembly of recombinant \( \alpha \)-synuclein using seeds from MSA brains did not reproduce the structures of the initial seeds. After this study was published, several other PMCA and RT-QuIC reactions using seeds from MSA or PD showed the same results (Burger *et al.*, 2021). It is possible that the chemical environment of the *in vitro* assembly conditions determines which structures form in the *in vitro* assembly of \( \alpha \)-synuclein into the filaments. PTMs and truncations may also be required for faithful amplification of the correct disease structures.

Our interpretations of templated seeding may thus need reconsideration. PMCA and RT-QuIC assays show that recombinant proteins can be seeded. In disease, additional densities which are not attributed to the filament are also observed, often orienting positively charged lysines. It is possible that non-proteinaceous cofactors are required for providing charge neutrality of the substrate to allow its incorporation into the filament. While the molecular mechanisms of seeded assembly remain unclear, PMCA and RT-QuIC results need to be interpreted with care.
Chapter 3

Assembly of recombinant tau

3.1 Introduction

This study was initiated in 2021. In human disease, full-length tau assembles into filaments. However, full-length tau is highly soluble and requires the addition of poly-anionic molecules to promote its assembly (Friedhoff et al., 1998; Goedert, Jakes, et al., 1996; Kampers et al., 1996; Pérez et al., 1996; Wilson & Binder, 1997). The in vitro assembled structures of recombinantly expressed full-length 3R and 4R tau filaments in the presence of heparin are polymorphic and different to those observed in disease (Zhang, Falcon, et al., 2019).

Truncated proteins may present an alternative route to full-length WT protein, as they readily assemble into amyloid filaments in the absence of anionic molecules. For example, fragments corresponding to amino acid residues 250-378, but excluding R2, form filaments using a hanging drop approach (R. A. Crowther, Jakes, et al., 1998). Fragments K11 and K12, corresponding to amino acid residues 244-394, with and without R2, respectively, also readily assemble (Wille et al., 1992). Fragments corresponding to the ordered cores from certain tauopathies including AD (residues 307-378), PiD (residues 254-378 but excluding R2) and CBD (residues 274-389), spontaneously assemble into filaments (Carlomagno et al., 2021). Whether their structures reminisce those in disease is unknown. The most widely used constructs are K18 and K19 (Gustke et al., 1994; Metrick et al., 2020; Mukrasch et al., 2005; Shammas et al., 2015; von Bergen et al., 2000; Yu et al., 2012). These fragments comprise the MTBD plus four additional amino acids at the C-terminus (residues 244-372, with and without R2 for K18 and K19, respectively). The K18 and K19 fragments are seeding competent by RT-QuIC (Metrick et al., 2020). These structures for K18 and K19 cannot be the same as those observed in disease, as all the ordered cores of known post-mortem structures extend to amino acid residues 378. The 297-391 construct (dGAE) was described as the proteolytically stable core of PHFs extracted from AD neurofibrillary tangles.
Assembly of recombinant tau

(Wischik et al., 1988). dGAE has been shown to spontaneously assemble into filaments with morphologies similar to PHFs observed by negative stain EM and Atomic Force Microscopy (AFM) (Al-Hilaly, Foster, et al., 2020).

In this chapter I report conditions that lead to the formation of AD PHFs using a range of experimental conditions and various construct lengths of tau, starting from dGAE, 297-391. I show that the filament structures formed in vitro are polymorphic and sensitive to their buffer environments. The same constructs can also be used to form type I and type II filaments of CTE. I demonstrate the effects of cations and how they can affect the structures formed. I describe how tau fragments can be extended and shortened, while still forming PHFs and CTE filaments. This study is published in Lövestam et al., eLife (2022). Text and figures of this publication are used below. The methods required to solve these structures, 76 in total, required high-throughput cryo-EM approaches are discussed in the Methods section and published in Lövestam & S. H. Scheres, 2022 Faraday Discussions. Some of the cryo-EM micrographs were acquired at ThermoFisher Scientific in Eindhoven by Abhay Kotecha, and his team members.

3.2 Results

3.2.1 In vitro assembly of tau (297–391) into PHFs

I first performed in vitro assembly of tau (297–391) in 10 mM phosphate buffer (PB) containing 10 mM dithiothreitol (DTT), with shaking at 700 rpm, as described (Al-Hilaly, Pollack, et al., 2017). I will refer to this as assembly condition 1. Filaments formed within 4 hr, as indicated by ThT fluorescence. Cryo-EM imaging after 48 hr revealed a single type of filament comprising two protofilaments that were related by 2-start helical screw symmetry. Although the extent and topology of the ordered cores resembled those of the protofilaments of AD and CTE, the protofilament cores were extended, rather than C-shaped (Figure 3.7A-C; filament type 1a).

I then reduced the shaking speed to 200 rpm. This resulted in a slower assembly reaction, with ThT fluorescence appearing after 6 hr. Cryo-EM structure determination after 48 hr showed that the filaments were polymorphic. They shared the AD protofilament fold, spanning residues 305–378. However, besides AD PHFs, I also observed filaments comprising three or four protofilaments, which I called triple helical filaments (THFs) and quadruple helical filaments (QHFs) (Figure 3.7D; filament types 2b–d). THFs and QHFs have not been observed in brain extracts from individuals with AD (Falcon, Zhang, Schweighauser, et al.,...
2018; Fitzpatrick et al., 2017), possibly because the presence of the fuzzy coat would hinder their formation. SFs were not seen on the micrographs nor by 2D classification. THFs consist of a PHF and an additional single protofilament with the AD fold, whereas QHFs are made of two stacked PHFs that come in two different arrangements (types 1 and 2). The cryo-EM maps of QHF type 1 filaments was of sufficient quality to build atomic models. QHF type 1 and the THF share a common interface, where one protofilament forms a salt bridge at E342 with K343 from the third, adjoining protofilament. This protofilament remains on its own in THFs, whereas it forms a typical PHF interface with a fourth protofilament in QHFs. Although the cryo-EM reconstruction of the QHF type 2 filament was of insufficient resolution for atomic modelling, the cross-section perpendicular to the helical axis suggested that a salt bridge was present between E342 from one PHF and K321 from another (Figure 3.2; filament types 2b–d).

Next, I sought to optimise the assembly conditions. THFs and QHFs have not been observed in AD and their formation would thus confound the use of this assembly assay for screens or to model disease. In addition, filaments from the above experiments tended to stick together, which complicated their cryo-EM imaging, and could interfere with their use. Over longer incubation periods (>76 hr), filaments tended to precipitate, resulting in cloudy solutions. Changing from microplates to eppendorf tubes resulted in faster precipitation of the protein. These precipitates were also ThT positive but their insoluble nature precluded the ability to perform cryo-EM studies and structure determination. To assemble tau (297–391) into pure PHFs and reduce stickiness, I explored the addition of salts and crowding reagents. In particular, the addition of 200 mM MgCl$_2$, 20 mM CaCl$_2$, or 0.1 µg/ml dextran sulphate resulted in purer populations of PHFs (95%) (Figure 3.7E; filament types 3–5). I also observed that timings of the reactions was important. Over longer incubations (>76 hrs) filaments tended to stick and clump together forming white precipitates. Indeed, pronase treated PHFs extracted from AD brains, clump together in solution which may suggest that clumping is an inherent nature of the surface of the PHF filaments. Non-pronase treated PHFs from human brain do not clump. It is possible that the fuzzy coat prevents the clumping of these filaments. Nevertheless, cryo-EM structure determination of filaments made using these conditions confirmed that their ordered cores were identical, with a root mean square deviation (r.m.s.d) of all non-hydrogen atoms of 1.3 Å, to those of AD PHFs (Figure 3.7F-G).

Tau (297–391) comprises the C-terminal 9 residues of R2, the whole of R3 and R4, as well as 23 amino acids after R4. The equivalent 3R tau construct, which lacks R2, begins with the C-terminal 9 residues of R1, the first four of which (266LKHQ269) are different from R2. I also assembled tau (266–391), excluding R2, in the presence of 200 mM MgCl$_2$,
Fig. 3.1 New electron cryo-microscopy (cryo-EM) structures. Backbone traces for filaments with previously unobserved structures. Residues 244–274 (R1) are shown in purple; residues 275–305 (R2) are shown in blue; residues 306–336 (R3) are shown in green; residues 337–368 (R4) are shown in yellow; residues 369–441 (C-terminal domain) are shown in orange. The filament types (as defined in Table 1) are shown at the top left of each structure.
Fig. 3.2 **Cross-sections of electron cryo-microscopy (cryo-EM) reconstructions.** Projected slices perpendicular to the helical axis, and with a thickness of approximately 4.7 Å, are shown for all cryo-EM structures described in this paper. For each structure, the filament type (as defined in Table 1) is displayed in the bottom left, and the percentage of a given type in the cryo-EM data set is shown in the top right.
Fig. 3.3 **Electron cryo-microscopy (cryo-EM) maps and models of new structures (part 1).** Cryo-EM density maps (grey transparent) and atomic models are shown for filaments with previously unobserved structures. Residues 244–274 (R1) are shown in purple; residues 275–305 (R2) are shown in blue; residues 306–336 (R3) are shown in green; residues 337–368 (R4) are shown in yellow; residues 369–441 (C-terminal domain) are shown in orange. The filament types (as defined in Table 1) are shown at the top left of each structure.
3.2 Results

Fig. 3.4 Electron cryo-microscopy (cryo-EM) maps and models of new structures (part 2). Cryo-EM density maps (grey transparent) and atomic models are shown for filaments with previously unobserved structures. Residues 244–274 (R1) are shown in purple; residues 275–305 (R2) are shown in blue; residues 306–336 (R3) are shown in green; residues 337–368 (R4) are shown in yellow; residues 369–441 (C-terminal domain) are shown in orange. The filament types (as defined in Table 1) are shown at the top left of each structure.
Fig. 3.5 Schematics of the tau folds (part 1). Negatively charged residues are shown in red, positively charged residues in blue, polar residues in green, non-polar residues in white, sulphur-containing residues in yellow, prolines in purple, and glycines in pink. PHF, paired helical filament; CTE, chronic traumatic encephalopathy.
Fig. 3.6 **Schematics of the tau folds (part 2).** Negatively charged residues are shown in red, positively charged residues in blue, polar residues in green, non-polar residues in white, sulphur-containing residues in yellow, prolines in purple, and glycines in pink. PHF, paired helical filament; CTE, chronic traumatic encephalopathy.
Fig. 3.7 Assembly of recombinant tau into filaments like Alzheimer’s disease paired helical filaments (AD PHFs). A Schematic of 2N4R tau sequence with domains highlighted. The regions 1N (44–73), 2N (74–102), P1 (151–197), and P2 (198–243) are shown in increasingly lighter greys; R1 (244–274) is shown in purple; R2 (275–305) is shown in blue; R3 (306–336) is shown in green; R4 (337–368) is shown in yellow; the C-terminal domain (369–441) is shown in orange. B Amino acid sequence of residues 244–441 of tau, with the same colour scheme as in A. C-E: Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for several cryo-microscopy (cryo-EM) reconstructions. The filament types (as defined in Table 1) are shown at the bottom left and the percentages of types for each cryo-EM data set are given at the top right of the images. C Conditions of Al-Hilaly et al., 2017, with shaking at 700 rpm. D My adapted protocol, using 200 rpm shaking. From left to right; paired helical filament (PHF), triple helical filament, quadruple helical filament type 1, and quadruple helical filament type 2. E The optimised conditions for in vitro assembly of relatively pure PHFs. F Cryo-EM density map (grey transparent) of in vitro assembled tau filaments of type 4a and the atomic model colour coded according as in A. G Backbone ribbon of in vitro PHF (grey) overlaid with AD PHF (blue).
Fig. 3.8 Electron cryo-microscopy (cryo-EM) micrographs and density maps comparing *in vitro* assembled paired helical filaments (PHFs) and Alzheimer’s disease paired helical filaments (AD PHFs). A Cryo-EM micrographs with filaments from *in vitro* assembly conditions 1 (left), 2 (middle), and 4 (right). Scale bar represents 100 Å. B Cryo-EM density maps of *in vitro* PHF (filament type 4a, EMDB-14063) as top view (top) and side view (bottom). C Cryo-EM density maps of AD PHF (EMDB:0259) as top view (top) and side view (bottom). The black arrows point to densities in front of lysines 317 and 321; the green and pink arrows point to densities inside the C-shape.
as well as a 50:50 mixture of the 3R/4R tau constructs, in the absence of MgCl₂. I observed PHFs, THFs, and QHFs in the absence of MgCl₂, whereas assembly in the presence of 200 mM MgCl₂ gave rise to AD PHFs with a purity greater than 94% (Figure 3.2; filament types 6 and 7). It is important to note that these numbers are averages and can differ between preparations. These findings show that bona fide PHFs can be formed from only 3R or 4R, albeit truncated, tau. In AD and CTE, all six isoforms, each full-length, are present in tau filaments (Goedert, Spillantini, Cairns, et al., 1992; Schmidt et al., 2001). It remains to be determined if there are PHFs in human diseases that are made of only 3R or 4R tau.

The cryo-EM structures of in vitro assembled PHFs and of AD PHFs shared the same left-handed twist and 2-start helical screw symmetry. Moreover, there were similar additional densities in front of lysine residues 317 and 321, and on the inside of the protofilament’s C, which has previously been observed for AD PHFs (Figure 3.8). The assembly buffers contained only Na₂HPO₄, NaH₂PO₄, MgCl₂, and DTT. The density is also present in the absence of MgCl₂. Although one cannot exclude the possibility that negatively charged co-factors may have purified together with recombinant tau, it is likely that the additional densities arose from phosphate ions in the buffer. The phosphates’ negative charges may have counteracted the positive charges of stacked lysines. It remains to be established if similar densities in AD PHFs also correspond to phosphate ions, or if other negatively charged co-factors or parts of the fuzzy coat may play a role. It is possible that any negative charge, be it from peptide sequences or non-proteinaceous molecules can counter-act the positive charges from the lysine. The fuzzy coat, which consists of only a few residues on either side of the ordered core, is not visible in the cryo-EM micrographs of in vitro assembled tau (297–391) (Figure 3.8).

3.2.2 The effects of salts on tau filament assembly

During optimisation of the assembly conditions, I noticed by cryo-EM that different cations in the buffer caused the formation of filaments with distinct protofilament folds. Negative stain EM is not able to show the different types of filaments due to its limited resolution. The morphologies of the filaments formed during optimisation of buffer constituents were very similar. Cryo-EM reconstruction was essential to determine the contributions of the salts to the structures.

Besides MgCl₂ and CaCl₂, which led to the formation of AD PHFs, I also explored the effects of ZnCl₂, CuCl₂, NaCl, LiCl, and KCl (Figure 3.9A). Addition of ZnCl₂ led to a small amount of precipitation. Rapidly mixing the reagents and bringing buffers and protein to
RT prior to the mixing of a reaction, alleviates precipitation. The addition of ZnCl$_2$ resulted in the same fold as observed for filaments assembled using condition 1, whereas addition of CuCl$_2$ led to folds with little resemblance to previously observed tau folds (Figure 3.1 and Figure 3.2; filament types 11 and 12). Cu$^{2+}$ ions led to the formation of intermolecular disulphide bonds that were part of the ordered cores of these filaments probably because Cu$^{2+}$ is an oxidising agent.

Monovalent cations modulated the formation of protofilament folds that were similar or identical to AD and CTE folds. The CTE fold is similar to the AD fold, in that it also comprises a two-layered arrangement of residues 274/305–379; however, it adopts a more open C-shaped conformation and comprises a larger cavity at the tip of the C, or $\beta$-solenoid (amino acid residues 338–354), which is filled with an additional, unknown density (Falcon, Zivanov, et al., 2019). First, I will describe how different monovalent cations led to the formation of both C-shaped and more extended protofilament folds. I then present the effects of cations on the additional density in the cavity and the conformations of the surrounding residues.

The addition of 200 mM NaCl led to the formation of two types of filaments. The first type was identical, with an all-atom r.m.s.d. of 1.4 Å to CTE type II filaments (Figure 3.9A-C; filament type 8a); in the second type (filament type 8b), two identical protofilaments with a previously unobserved, extended protofilament fold packed against each other with 2-start helical symmetry. This fold resembled the extended fold observed when using condition 1. The extended fold concurred with a flipping of the side chains of residues 322–330, which were alternatively buried in the core or solvent-exposed in opposite manner to the CTE fold. Side chains before and after $^{364}$PGGG$^{367}$ had the same orientations, but formed a 90° turn in the CTE fold and adopted a straight conformation in the extended fold. Residues 338–354 had identical conformations, with an all-atom r.m.s.d for these residues of 1.3 Å, at the tips of the C-shaped and extended folds (Figure 3.1; Figure 3.10). When adding 200 mM LiCl, I observed two types of filaments, with either C-shaped or more extended protofilament folds (Figure 3.9A; filament types 10a and 10b). In the first type, two C-shaped protofilaments packed against each other in an asymmetrical manner. In the second type, two protofilaments with an extended conformation packed against each other with 2-start helical symmetry. As observed for the filaments obtained with NaCl, the side chain orientations of residues 322–330 differed between folds. However, whereas the side chain of H330 was buried in the core of the C-shaped protofilament formed with NaCl, it was solvent-exposed in the C-shaped protofilament formed with LiCl. This suggests that the conformation of the $^{364}$PGGG$^{367}$ motif defines the extended or C-shaped conformation (Figure 3.11). Addition of 200 mM KCl also
Fig. 3.9 Assembly of recombinant tau into filaments like chronic traumatic encephalopathy (CTE) type II filaments. A Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis are shown for different assembly conditions and filament types (as defined in Table 1), which are indicated in the bottom left. The percentages of types are shown in the top right of each panel. B Cryo-EM density map (grey transparent) of filament type 8a and the corresponding atomic model with the same colour scheme as in Figure 3.7. C-E Backbone ribbon views of protofilament and filament folds. C In vitro NaCl filament type 8a (grey) overlaid with CTE type II (orange). D Extended and C-shaped protofilaments aligned at residues 338–354 for LiCl, filament types 9a and 9b (left) and NaCl, filament types 8a and 8b (right). E Filament types 8b (NaCl), 9a (LiCl), 10a (KCl), and 4a (MgCl₂) aligned at residues 356–364. F Atomic view of residues 334–358. The distance between the Cα of L344 and I354 is indicated. Filament types 8a, 8b, 9a, 9b, and 10a are shown in light purple, dark purple, dark orange, light orange, and blue, respectively.
led to two different filaments with either extended or C-shaped protofilament folds. However, in this case, low numbers of filaments with extended protofilaments resulted in poor cryo-EM reconstructions. The filaments with C-shaped folds comprised three protofilaments, which packed against each other with C3 symmetry (Figure 3.1; Figure 3.9A; filament type 10a).

Fig. 3.10 **Extended and C-shaped NaCl protofilaments.** A Backbone ribbon view of C-shaped (filament type 8a) and extended (filament type 8b) protofilaments formed with NaCl, aligned at residues 338–354. B-D Close-up atomic view of regions highlighted in A. E Cryo-microscopy density map (transparent grey) and the atomic model at residues 338–354. Filament types 8a and 8b are shown in dark and light orange, respectively.

For each monovalent cation, residues 338–354 adopted identical conformations when comparing extended and C-shaped protofilament folds (Figure 3.9D). These residues surrounded the cavity at the tip of the fold, which was filled with an additional density in the CTE fold. Additional densities were also observed in filaments formed in the presence of NaCl, KCl, and in the extended filaments formed with LiCl. The cryo-EM reconstructions of the threefold symmetric filaments formed with KCl, with a resolution of 1.9 Å, showed multiple additional spherical densities inside the protofilament core. Besides additional densities for what were probably water molecules in front of several asparagines and glutamines, the cavity at the tip of the fold contained two larger, separate spherical densities per β-rung, which were 3.1 Å apart, and at approximately 3.0–4.5 Å distance from S341 and S352, the only polar residues in the cavity. Another pair of additional densities, similar in size to those inside the cavity, was present at a distance of 2.6 Å from the carbonyl
Fig. 3.11 **Extended and C-shaped LiCl protofilaments.** A Backbone ribbon view of extended (filament type 9a) and C-shaped (filament type 9b) protofilaments formed with LiCl, aligned at residues 338–354. B-D Close-up atomic view of regions highlighted in A. E Cryo-microscopy density map (transparent grey) and the atomic model at residues 338–354. Filament types 9a and 9b are shown in light and dark purple, respectively.
3.2 Results

Fig. 3.12 Cryo-microscopy (cryo-EM) densities inside the cavities of KCl and NaCl filaments. **A–H** Cryo-EM density map (transparent grey) and the corresponding atomic models for filament types 7a (assembled with NaCl) and 8a (assembled with KCl) are shown in orange and green, respectively. **A** Cryo-EM density at the tip of the protofilament of 8a with additional non-proteinaceous densities highlighted by arrows. Blue arrows represent putative water molecules; pink arrows represent putative K+ or Cl ions. K+ ions are shown in pink; Cl ions are shown in cyan. **B–D** Close-up view of putative K+ and Cl ion pairs. Distances between K+ and Cl ions, and Oγ of S341 or the carbonyl oxygen of G335 are indicated. **E** Cryo-EM density of the tip of the protofilament of 8a with putative Na+ and Cl ions fitted into the density. Purple arrow represents putative Na+ or Cl ions; Na+ ions are shown in purple; Cl ions are shown in cyan. **F–G** Close-up view of putative Na+ and Cl ions. Distances between the Na+ and Cl ions and the Oγ of S341 are indicated. **H** Cryo-EM density of chronic traumatic encephalopathy type I (EMD:0527), with putative Na+ and Cl ions modelled inside the additional density.
of G335 (Figure 3.12). These densities probably correspond to pairs of $K^+$ and Cl ions. Reconstructions for the filaments formed with NaCl were at resolutions of 2.8 and 3.3 Å. The additional density in these maps was not separated into two spheres, but was present as one larger blob per rung, with separation between blobs along the helical axis. Filaments formed with LiCl were resolved to resolutions of 3.1 and 3.4 Å. No additional densities were present inside the cavity of the C-shaped fold, but the cavity in the extended fold contained a spherical density that was smaller than the densities observed for NaCl and KCl filaments (Figure 3.12). Different cations also led to conformational differences in residues S356 and L357, which were akin to the differences observed previously between AD and CTE folds (Figure 3.9E, F) (Falcon, Zivanov, et al., 2019). In the AD fold, S356 is solvent-exposed and L357 is buried inside the protofilament core, whereas they adopt opposite orientations in the CTE fold. As mentioned, filaments formed with NaCl are identical to CTE filaments; in filaments formed with KCl, S356 and L357 are oriented in the same directions as in the AD fold; in filaments formed with LiCl, both residues are buried in the core.

Efforts to generate AD SFs using mixtures of salts were unsuccessful, but they did reveal novel protofilament interactions (Figure 3.2; filament types 13–17). Notably, using a buffer with MgSO$_4$ and NaCl, I obtained a minority of filaments with an SF interface (11%, filament type 15d). Probably because of the presence of NaCl, protofilaments adopted the CTE fold. Further exploration of the role of salts may lead to the assembly of recombinant tau into AD SFs and CTE type I filaments.

### 3.2.3 The effects of protein length on tau filament assembly

I also investigated if the N- and C-termini of tau (297–391) are required for its assembly into PHFs. As mentioned above, over longer incubations, 297-391 PHFs stick and clump together, which complicates downstream experiments. In AD, full-length tau assembles and these filaments do not clump.

I first made a series of protein fragments ending at residue 391, to explore the effects of the position of the N-terminus. Next, keeping the N-terminus at residue 297, I explored the effects of the position of the C-terminus (Figure 3.13A). Each recombinant tau fragment was assembled in 10 mM PB, 10 mM DTT, and 200 mM MgCl$_2$, at 200 rpm shaking. I assessed the presence of filaments by negative stain EM and used cryo-EM to determine their structures (Figure 3.13B; filament types 18–30).
3.2 Results

Proteins comprising the entire N-terminal domain (residues 1–391 of 0N4R tau) did not assemble into filaments. The same was true of proteins starting at residues 151, 181, or 231 in the proline-rich region. When using proteins starting at 244, the first residue of R1, I observed the formation of filaments. The quality of the cryo-EM reconstructions was not sufficient for atomic modelling, but the ordered filament cores adopted a more open conformation than in the Alzheimer fold, with two protofilaments interacting at their tips. Residues 338–354 probably adopted the same conformation as in the Alzheimer fold. Addition of NaCl or pyrophosphate did not lead to the formation of PHFs or CTE filaments (Figure 3.1; Figure 3.2; filament types 19–20). However, it is possible that this construct will form filaments with the Alzheimer fold after further optimisation of assembly conditions. Proteins that started at residues 258 or 266 in R1 formed PHFs. Adding NaCl to tau (297-391), or using phosphate-buffered saline (PBS), gave rise to CTE type II filaments, as well as to filaments consisting of either single CTE protofilaments or two protofilaments, related by C2 symmetry, which formed an interface at residues 327–336 (Figure 3.2; filament types 15b, 23b). For all filaments assembled in the presence of NaCl, I saw densities inside the tip of the fold’s cavity. Moreover, with the proteins starting at residues 258 and 266, I observed proteinaceous densities, which packed against the N-terminus of the protofilaments. This may have been the extension into R2 (Figure 3.2; filament types 23a–c).

Tau (297–391) can also be shortened from the N-terminus, as proteins starting at residues 300 or 303 still formed filaments with AD or CTE folds (Figure 3.2; filament types 25–26). When shortened to residue 306, I also observed filaments, but their stickiness precluded cryo-EM structure determination. No filaments were formed when shortening the protein to residue 310. It has previously been shown that the PHF6 motif (\(^{306}\text{VQIVYK}^{311}\) in tau) is essential for filament formation (W. Li & V. M.-Y. Lee, 2006; von Bergen et al., 2000).

Reminiscent of what I observed for the N-terminal domain, the tau starting at residue 297 and ending at 441 failed to assemble into filaments. The same was true of tau (297–431). However, proteins ending at residues 421, 412, 402, or 396 formed filaments, but their ordered cores are much smaller than in the Alzheimer fold, precluding cryo-EM structure determination to sufficient resolution for unambiguous atomic modelling (Figure 3.13B; filament types 28–31). I did observe filaments with the Alzheimer fold for a protein ending at residue 394 (Figure 3.2; Figure 3.13B; filament type 32).

As observed for proteins ending at residue 384, when tau (297-391) was shortened from the C-terminus, it could still form AD and CTE folds (Figure 3.13B; filament type 33). How-
Fig. 3.13 The effects of protein length on tau filament assembly. A Schematic representation of 2N4R tau and the constructs used in this study. A red cross indicates that no filaments were formed; an orange dash indicates that filaments with structures distinct from Alzheimer’s disease paired helical filaments (AD PHFs) were formed; a blue circle indicates that the AD protofilament fold was formed; a green tick indicates that AD PHFs were formed. Resulting filament types (as defined in Table 1) are indicated for each experiment. B Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed with the constructs in A. Filament types are indicated at the bottom left; percentages of filament types in each cryo-EM data set are shown in the top right. (*) 50:50 ratio of 266–391 (3R) and 297–391 (4R) tau.
ever, constructs ending at residue 380 formed straight ribbons, precluding cryo-EM structure determination.

As shown in Figure 3.7, I found that reducing the shaking speed from 700 to 200 rpm was important for the formation of filaments with the AD fold by tau (297–391). To show that this was not only the case for tau (297–391), I assembled tau (258–391) and tau (297–394) with a shaking speed of 200 rpm. For tau (297–394), besides structures similar to those formed by tau (297–391) at 200 rpm, I observed an additional filament with 2-start helical screw symmetry that comprised four protofilaments (Figure 3.1; Figure 3.2; filament type 34). In the presence of NaCl, I observed filaments with similar extended conformations, but with a CTE cavity. This was also the case when tau (300–391) and tau (303–391) were assembled at 700 rpm in PBS (Figure 3.1; 3.2; filament types 35–37). When assembling tau (258–391) at 700 rpm, I observed protofilaments with partial similarity to the GGT fold. The structures were nearly identical at residues 288–322, with an all-atom r.m.s.d. of 1.5 Å, (Figure 3.1; 3.2; 3.14; filament type 38a). It is possible that the absence of a non-proteinaceous co-factor, which was hypothesised to be present inside the GGT fold, precluded formation of bona fide GGT filaments. Addition of heparan sulphate or phosphoglycerate to the assembly buffer did not result in formation of the GGT fold (Figure 3.1; 3.14; filament types 39–41). Further experiments are thus needed to identify the co-factor of GGT.

### 3.2.4 The effect of pseudo-phosphorylation on tau filament assembly

Whereas *in vitro* filament assembly of recombinant tau repeats was inhibited by N- and C-terminal regions, it is predominantly or only full-length tau that assembles into filaments in AD (Goedert, Spillantini, Cairns, *et al.*, 1992; V. M. Lee *et al.*, 1991). This difference may be due to the fact that I used unmodified recombinant proteins for *in vitro* assembly, while tau undergoes extensive post-translational modifications in AD (Wesseling *et al.*, 2020). Immunochemistry and mass spectrometry have identified abnormal hyperphosphorylation of S396, S400, T403, and S404 in PHF-tau (Hasegawa *et al.*, 1992; Kanemaru *et al.*, 1992; V. M. Lee *et al.*, 1991; Morishima-Kawashima *et al.*, 1995). Because these sites are located in the C-terminal domain that prevents the formation of PHFs from recombinant tau, I hypothesised that their phosphorylation may modulate PHF formation and help to overcome the inhibitory effects of the C-terminal domain. I therefore mutated these residues to aspartate, with its negative charge mimicking the negative charges of phosphorylation, in the tau constructs ending at residues 408 and 441 (Figure 3.13; filament types 42–43).
Fig. 3.14 *Comparison of in vitro assembled tau filaments with the globular glial tauopathy (GGT) fold*. **A** Backbone ribbon view of GGT type 1 (PDB:7p66) and tau filaments from *in vitro* assembled filament types 27a, 28a, 29a, 32a protofilament (PF)1, and 32a PF2, aligned at residues 288–322. **B** Close-up atomic view. 7p66 is shown in grey, 27a in green, 28a in pink, 29a in orange, 32a PF-1 in blue, and 32a PF-2 in purple.
3.2 Results

Fig. 3.15 Coomassie stained sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel for constructs 297–391, 297–408, and 296–441 before and after filament assembly. About 3.5 g of protein was loaded into each well. f: after filament assembly; *: constructs with the four-phospho-mimetic mutations: S396D, S400D, T403D, and S404D.
As described above, tau (297–408) formed filaments with small ordered cores; its pseudophosphorylated mutant formed two types of filaments consisting of AD PHF protofilaments. Both filament types had the same inter-protofilament interface as that of THFs and type 1 QHFs, where E342 from one protofilament forms a salt bridge with K343 of the other. The second filament type was a doublet of the first. Importantly, tau (297–441), which does not assemble into filaments, did so in presence of the four phospho-mimetic mutations. Filaments consisted of a single protofilament with the AD fold.

It has been suggested that AD filaments may also be phosphorylated at S356 (Hanger, Betts, et al., 1998; Hanger, Byers, et al., 2007). I therefore tested phospho-mimetic mutation S356D. However, using tau (266–391) with S356D, I was unable to form PHFs. Instead, filaments consisted of two asymmetric protofilaments, both of which had features reminiscent of the GGT fold. In the presence of NaCl, this construct formed filaments with a novel fold, consisting of two identical protofilaments related by 2-start helical symmetry (Figure 3.1; Figure 3.2; Figure 3.14; filament types 44 and 45). Analysis by gel electrophoresis confirmed that the tau constructs forming these filaments remained intact (Figure 3.15).

### 3.2.5 The effects of anionic cofactors on tau filament assembly

The effects of anionic co-factors on tau filament assembly Addition of anionic co-factors leads to the formation of filaments from full-length tau (Chakraborty et al., 2021; Farid et al., 2014; Goedert, Jakes, et al., 1996; Kampers et al., 1996; Pérez et al., 1996; Wilson & Binder, 1997). However, it was previously shown that heparin-induced tau filament formation led to structures that were different from those observed in disease thus far, under the conditions used (Zhang, Falcon, et al., 2019). To further explore the effects of anionic co-factors, I solved cryo-EM structures of filaments formed using full-length recombinant tau in the presence of RNA (Kampers et al., 1996) or phosphoserine. The resulting tau filaments had structures that have not been observed previously.

Addition of RNA led to the formation of two asymmetrical and extended protofilaments, similar to heparin-induced 3R tau filaments (Zhang, Falcon, et al., 2019) (Figure 3.2; filament type 46). The resulting map was of insufficient resolution for atomic modelling. I also determined a 1.8 Å resolution reconstruction for full-length tau filaments that formed in the presence of 5 mM phosphoserine (Figure 3.2; filament type 47). At this resolution, the absolute handedness of the filament was obvious from the position of the main-chain carbonyl oxygens. Whereas most tau filaments described in this paper have a left-handed
twist (see Methods), phosphoserine-induced filaments are right-handed. Surprisingly, the ordered core of this filament comprised residues 375–441, encompassing only residues from the C-terminal domain. Filaments consist of two protofilaments that are related by 2-start helical symmetry and the protofilament fold contains eight $\beta$-strands. Several water molecules were observed, particularly in front of the side chains of serines, threonines, and asparagines. This is the first example of a region outside the microtubule-binding repeats of tau forming amyloid filaments.

### 3.3 Discussion

In the course of protein evolution, natural selection has produced amino acid sequences that fold into specific protein structures to fulfil the multitude of tasks required to maintain life. The observation that a given protein can adopt multiple different amyloid structures has made it clear that the paradigm by which a protein’s sequence defines its structure may not hold for amyloids. Apparently, the packing of $\beta$-sheets against each other in amyloid filaments can happen in many ways for a single amino acid sequence. The work described here and performed previously (Y. Shi, Zhang, et al., 2021; Zhang, Falcon, et al., 2019) illustrates this for tau. Similar observations have been made for amyloid-$\beta$ (Bertini et al., 2011; Kollmer et al., 2019; Lu et al., 2013; Wälti et al., 2016; Xiao et al., 2015; Yang, Arseni, et al., 2022), $\alpha$-synuclein (Guerrero-Ferreira, Taylor, Arteni, et al., 2019; Y. Li et al., 2018; Lövestam, Koh, Knippenberg, et al., 2021; Schweighauser et al., 2020; Tuttle et al., 2016), TAR DNA binding protein-43 (Arseni et al., 2021; Q. Li et al., 2021), and immunoglobulin light chain (Radamaker et al., 2019; Swuec et al., 2019). Because amyloid filaments of these proteins are typically associated with pathology, rather than function, their structural diversity could be disregarded. However, the observation that, for tau, and possibly also other proteins, specific amyloid structures define distinct neurodegenerative conditions, raises important questions on what drives their formation.

The lack of in vitro assembly models for recombinant tau replicates the amyloid structures observed in disease has hampered progress. Here, I identified conditions for the in vitro assembly of AD PHFs and type II CTE filaments and established the formation of these structures by cryo-EM structure determination to resolutions sufficient for atomic modelling. The latter is crucial. Biochemical methods that discriminate between the different structures in solution do not exist, and negative stain EM or atomic force microscopy does not provide
sufficient resolution to unequivocally distinguish between them.

The ability to make AD PHFs and CTE type II filaments in vitro opens new avenues for studying tauopathies. In vitro assembly assays could be used to screen for compounds that inhibit filament formation. Alternatively, in vitro generated filaments may be used in high-throughput screens for small-molecule compounds that bind specifically to a single type of filament. Such amyloid structure-specific binders could be developed into ligands for positron emission tomography to differentiate between tauopathies in living patients. Moreover, specific binders could be explored for use in therapeutic approaches that aim to degrade filaments inside neurons through their coupling with the protein degradation machinery inside the cell. Ultimately, specific binders could even obviate the need for cryo-EM structure determination to confirm the formation of the correct types of filaments in new model systems for disease.

Besides their use in screens, in vitro assembly of tau filaments also provides a model system, amenable to experimental perturbation, for the study of the molecular mechanisms that underlies amyloid formation. Coupled to cryo-EM structure determination, this provides a promising model system for studying the formation of different tau folds that define distinct tauopathies. The work presented in this paper provides two examples, as outlined below.

A major difference between Alzheimer and CTE tau folds is the presence of a larger cavity that is filled with an additional density in the CTE fold. Based on the relatively hydrophobic nature of this cavity, it was previously hypothesised that this additional density may correspond to an unknown hydrophobic co-factor that assembles with tau to form CTE filaments (Falcon, Zivanov, et al., 2019). However, I observed the in vitro assembly of CTE type II filaments in the absence of hydrophobic co-factors. Instead, whether CTE type II filaments or AD PHFs formed in vitro was determined by the presence or absence of NaCl. Moreover, using different monovalent cations changed the additional densities, as well as the conformation of residues surrounding the cavity. The 1.9 Å map for the structure formed with KCl showed two spherical blobs of additional density per β-rung that were in an arrangement that would be consistent with a pair of K⁺ and Cl ions; a similar pair of spherical densities was also observed in front of G335 (Figure 3.12). Therefore, it is possible that the extra density in the cavity of filaments formed with NaCl also corresponds to Na⁺ and Cl ions. The continuous nature of the additional density along the helical axis could arise from limited resolution of the reconstructions with NaCl, or from the ions not obeying the 4.75 Å helical rise that is imposed on the reconstruction. The observation that the filaments formed with
NaCl are identical to CTE type II filaments, which show a similar extra density, suggests that, rather than being filled with hydrophobic co-factors, the cavity in CTE filaments may contain NaCl. Na+ and Cl levels in neurons are typically much lower than the 200 mM NaCl used here, but it could be that brain trauma somehow leads to increased levels of these ions in the brain regions where tau filaments first form. Additional studies will explore this further.

My results have also identified how regions outside the repeats interfere with the in vitro assembly of AD PHFs. In particular, the presence of the N-terminal domain, including its proline-rich region, and residues beyond 421 in the C-terminal domain, inhibited the spontaneous assembly of recombinant tau. In addition, the presence of residues 396–421 led to the formation of filaments with much smaller ordered cores. The C-terminal region has been reported to inhibit anionic co-factor-induced assembly of full-length tau, with pseudophosphorylation overcoming inhibition (Abraha et al., 2000; Haase et al., 2004).

Interestingly, mutating serine or threonine residues at positions 396, 400, 403, and 404 to aspartate, to mimic phosphorylation, overcame these inhibitory effects. Although I did not yet examine the effects of pseudophosphorylation of the N-terminal domain, I note that the proline-rich region upstream of the repeats is positively charged, which may inhibit amyloid formation. The introduction of negative charges through phosphorylation, or the removal of positive charges by, for instance, acetylation, may be necessary for filament assembly. In AD, filamentous tau is hyperphosphorylated, especially in the proline-rich region and the C-terminal domain (Grundke-Iqbal et al., 1986; Iqbal et al., 2016). Hyperphosphorylation inhibits the ability of tau to interact with microtubules (Bramblett et al., 1993; Yoshida & Ihara, 1993), which may be necessary for filament formation, since the physiological function of tau and its pathological assembly requires the repeat region. Other post-translational modifications, such as acetylation and ubiquitination of positively charged residues in the microtubule-binding repeats, may also play a role (Wesseling et al., 2020). It is therefore possible that post-translational modifications of tau lead to a higher propensity to form amyloid filaments in disease. Thus, the enzymes causing these modifications could be therapeutic targets.

Moreover, most of the structures in Figure 3.1, all of which are of recombinant tau proteins over-expressed in E. coli, share features with the Alzheimer and CTE folds: a cross-β packing of residues from near the start of R3 against residues from near the start of the C-terminal domain, and a turn in R4. Other tau folds are markedly different. Unmodified tau may have a tendency to form filaments that resemble Alzheimer and CTE folds, whereas
specific post-translational modifications and/or associated molecules may play a role in driving the formation of tau folds associated with other tauopathies.

This work illustrates how cryo-EM structure determination can guide the development of better experimental models by showing that the filaments generated from recombinant tau are identical to those formed in disease. As a start, I identified in vitro assembly conditions for the formation of tau filaments like those of AD and CTE. Future work will explore which factors, including post-translational modifications and/or associated molecules, determine the formation of other tau filaments. These results do not only have implications for the in vitro assembly of tau. Similar strategies could be applied to the study of other amyloid-forming proteins, and insights from in vitro assembly studies may carry over to the development of better experimental models of disease, including in cell lines and in animals. Whereas solving a cryo-EM structure previously constituted a scientific project in itself, current through-put allows the use of cryo-EM structure determination as a tool in projects with a wider scope. I envision that this type of high-throughput cryo-EM structure determination will play a crucial role in future investigations into the structural diversity of amyloids.
Chapter 4

Time-resolved studies of tau filaments

4.1 Introduction

The mechanisms of amyloid assembly remain unknown. The cryo-EM structures of amyloid filaments from post mortem brains have shown that specific amyloid structures correspond to different diseases (Schweighauser et al., 2020; Y. Shi, Zhang, et al., 2021; Yang, Arseni, et al., 2022). However, since post mortem brains represent the end stage of the disease, they provide limited knowledge of how these amyloids assembled initially. The in vitro assembly of recombinant proteins into amyloid filaments provides a unique opportunity to study the molecular mechanisms of amyloid formation. I showed in the previous chapter that I can assemble tau into filaments identical to those in AD and CTE by using recombinant tau protein with amino- and carboxy-terminal truncations. In this chapter, I use time-resolved cryo-EM to investigate the process of tau filament assembly, specifically the formation of the structures observed in AD and CTE. At early time intervals, tau assembles into small filaments with a core corresponding to amino acid residues $^{302}$GGSVQIVYKPVDS$^{316}$. This First Intermediate Amyloid (FIA), has the ability to rearrange into multiple different intermediate filament structures prior to maturing into end-stage filaments. I received help from Stephan McLaughin for analysing AUC data. Abhay Kotecha at Thermo Fisher Scientific acquired part of the cryo-EM images. The results in this section will be published after submission of this thesis.
4.2 Results

4.2.1 Monomeric tau

I used the tau 297-391 construct described in chapter 2. Protein purification of tau results in a single band by SDS-PAGE with the expected molecular weight (MW) of a monomer (10.1 kDa). AUC of the 297-391 construct at 4 ºC and 20 ºC, shows a single sedimenting protein species with a sedimentation coefficient of 0.6 S. The value of 0.6 S is normalised...
Fig. 4.2 **ThT and cryo-EM micrographs shown of time-resolved filament formation** 

a Assembly was monitored by thioflavin T fluorescence of recombinant 297-391 tau in the presence of MgCl$_2$ (blue) and NaCl (green). b Cryo-EM micrographs of filaments at 120, 240 and 360 min in the presence of MgCl$_2$. Long cross-over PHF (lcoPHF). Scale bar, 50 nm.
to standard conditions \((Sw,20 = 1.0 \text{ S})\) (Figure 4.1). The sedimentation coefficient is lower than 1.0 S which suggests that the protein is relatively small or compact. The fitted frictional coefficients for interference and absorbance data were 1.784 and 2.106, respectively. These values are consistent with an extended protein conformation. The calculated mass from the absorbance and interference data is 10.7 kDa, which is close to the expected mass for a monomer of the 297-391 protein construct. These results shows that the protein is not aggregated and is present as a single species in solution.

### 4.2.2 Cryo-EM structures of early tau intermediate filaments

To study the molecular mechanisms of amyloid assembly, I used the \textit{in vitro} assembly of tau (297-391) as described (Lövestam, Koh, van Knippenberg, et al., 2022). I used two different reaction conditions: condition m \((10 \text{ mM PB pH 7.2 10 mM DTT 100 mM MgCl}_2)\) and condition n \((20 \text{ mM PB pH 7.2 10 mM DTT 100 mM NaCl})\). I used ThT to measure the presence of filaments during assembly. To increase the sensitivity of the ThT signal, I set a high gain to the reader. ThT signal was observed after 180 min in the presence of NaCl or MgCl\(_2\) (Figure 4.2). This proceeded with a steep increase in ThT fluorescence. Due to the high gain settings, saturation was observed after 600 and 800 min for NaCl and MgCl\(_2\) respectively.

I used cryo-EM to sample different time-points during the assembly of 297-391 in both conditions. In the lag phase, after 120 min for NaCl and MgCl\(_2\), I observed small assemblies in the cryo-EM micrographs (Figure 4.3). In the Fourier transform (FT) of the micrographs containing the small aggregates, I observed a 4.75 Å signal. The 4.75 Å signal in the FT is indicative of \(\beta\)-strands in the structures. I therefore proceeded to solve their cryo-EM structures using helical reconstruction.

Reference-free 2D class averaging of condition m and n after 120 min showed that the small elongated assemblies were structured and had a characteristic "beads-on-a-string-like" pattern with a clear 4.75 Å separation along the helical axis (Figure 4.4). I will call the small elongated assemblies the first intermediate amyloid (FIA). The bead-like appearance arises from the fast twist and short crossover distance of the the tau FIA (Figure 4.4). The crossover distance of the tau FIA was 135 Å with a twist of -6.32 ° per rung. I solved the structure to 2.1 Å and 2.2 Å resolution, in the presence of MgCl\(_2\) and NaCl, respectively, allowing for unambiguous atomic modeling (Figure 4.5). Cryo-EM showed the presence of two symmetrical protofilaments that are related by pseudo-\(2_1\) helical screw symmetry and that are composed of \(302^{\text{GGSVQIVYKPV}}\). The discernable backbone carbonyls in the reconstruction indicate that the filaments adopt a left-handed twist. The tau FIA
adopts identical structures in the presence of MgCl$_2$ and NaCl suggesting that the structure is independent of the chemical environment of assembly and follows a common mechanism of primary nucleation.

Fig. 4.3 Cryo-EM micrographs and the corresponding Fourier transform of filaments which had formed after 120 min in condition m a short FIAs and b long FIAs. Scale bar represents 20 nm.

Fig. 4.4 Cryo-EM micrograph and the 2D reconstruction of early tau filaments: FIA a Cryo-EM micrograph showing early stages of filament formation. Top left shows a 2D class average of tau FIA. Scale bar represents 20 nm. b Cryo-EM reconstructed electron density map of the FIA. The Z plane is shown at high contour (top panel) and low contour (bottom panel) levels. Scale bar represents 135 Å and represents a single crossover distance of the the tau FIA.
Fig. 4.5 **Structure of FIAS** a Amino acid sequence of residues 297-331 of tau. Residues are shown in purple with the core of the FIA highlighted in purple. β-strands are shown as thick arrows underneath the amino acid sequence. b Cryo-EM density map (grey transparent) of early assembled tau filaments and the atomic model coloured in purple a. c Top (XY) cartoon view of 8 successive rungs of tau FIA. d Side (Z) cartoon view of 8 successive rungs of tau FIA. The separation between individual stacks is 4.75 Å and the twist is -6.32°.
The two opposing \( \beta \)-strands in FIA pack against each other through a steric zipper that is predominantly stabilised by residues \( 306^{\text{VQIVYK}}^{311} \), whereas amino acids before and after this region are less structured, as evidenced by smears in the electron density map. The interaction between the \( \beta \)-strands is predominantly non-polar. In the case of FIA, the structure is similar structures to those previously described, with two sheets closely interlocking to create a dry interface (Sawaya et al., 2007). The N and C-termini form the fuzzy coat (84\%), which gives rise to the amorphous or aggregate morphologies observed in the cryo-EM micrographs. The shared FIA structure in conditions containing NaCl or MgCl\(_2\) suggests there is a common mechanism of primary nucleation.

I then studied the time-resolved maturation of the tau FIA in the two different conditions. Filament types throughout the text will be labelled according to their condition (m - MgCl\(_2\); n - NaCl), their corresponding times, 120, 180, 240, 300, 360, and 720, the type of filament in the time course (a-z), and the protofilament type (1-9). I used the following colour code for the structures sampled at time-points of 120, 180, 240, 300, 360, and 720 min: purple, blue, green, yellow, orange, and red, respectively. The colour codes will be used throughout this result section for illustrating the atomic structures from their respective time points.
4.2.3 Time-resolved filament formation in the presence of MgCl₂

In the presence of MgCl₂, during the 120-180 min time-frame, I observed filaments with a range of morphologies. However, due to the flexible characteristics and large number of polymorphs, I was unable to solve their structures. These filaments appeared unfolded or swollen in the micrographs. It is unclear why this happened.

Fig. 4.6 Structures of tau filaments formed in the presence of MgCl₂. Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed in the presence of MgCl₂. Time in min is indicated at the top left of the projected slices. Each slice has a coloured box which shows the time point when the condition was sampled. The filament IDs are labelled at the top left of each projected slice. Four time course experiments were conducted (Experiments 1-4). Experiment 1 was sampled at 120, 180, 300, and 720 min, Experiment 2 was sampled at 120, 240, 360, and 720 min, Experiment 3 was sampled at 120, 240, 300, and 720 min, and Experiment 4 was sampled at 180, 240, 300, 360, and 720 min.

For later time-points, I solved the cryo-EM structures of assembled tau that were sampled at time intervals of 180, 240, 300, 360 and 720 min. I conducted four replicate experiments (Experiments 1-4). Experiment 1 was sampled at 120, 180, 300, and 720 min, experiment 2 was sampled at 120, 240, 360, and 720 min, experiment 3 was sampled at 120, 240, 300, and 720 min, and experiment 4 was sampled at 180, 240, 300, 360, and 720 min. All filament structures that were solved from these experiments are shown in Figure 4.6.
Fig. 4.7 *Structures of tau filaments formed after 180 min in the presence of MgCl₂*

a Backbone ribbon of *in vitro* tau filaments formed after 180 min. Structures are aligned at amino acid residues S305-G324 and D358-E380. b Structure of protofilament from filament type m180-a. c Structure of protofilament 1 (m180-c1) from filament type m180-c. d Structure of protofilament 2 (m180-c2) from filament type m180-c. e Protofilament interface for filament types m180-a and f m180-a. Filament type m180-a is shown in indigo, protofilament m180-c1 is shown in light blue and protofilament m180-c2 is shown in dark blue. Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed after 180 min.
After 180 min, I observed three distinct filament types (m180-a, m180-b and m180-c; Figures 4.6 and 4.7). Filament types m180-a and m180-c were of sufficient resolution for atomic modelling. Although the cryo-EM reconstruction of filament type m180-b was of insufficient resolution, the cross-section perpendicular to the helical axis shows similar morphologies to filament types m180-a and m180-c but it is flexible at amino acid residues 338-357, as evidenced by smears in the electron density map. The protofilaments of filament types m180-a, m180-b and m180-c adopt extended conformations and comprises amino acid residues 306-380 in the filament core (Figure 4.7). Filament type m180-a consists of two symmetrical protofilaments which pack against each other by pseudo-2\_1 helical screw symmetry. Filament type m180-c consists of two asymmetrical protofilaments which pack against each other asymmetrically. The protofilaments in type m180-a, and the protofilaments in type m180-c (m180-c1 and m180-c2 Figure 4.7) differ at residues N327-G336 and the orientation of the side chain S356. In the protofilaments of filament type m180-a, N327 is solvent-exposed and \textsuperscript{328}IHH\textsuperscript{330} adopts "in-out-in" conformations, in which an "out" conformation means that the side chains are solvent-exposed and an "in" orientation means the side-chain is buried inside the protofilament core. In protofilaments types m180-c1 and m180-c2 in filament type m180-c, N327 is buried inside the protofilament core, which results in an out-in-out conformation at residues \textsuperscript{328}IHH\textsuperscript{330}. In protofilament types m180-a and m180-c1, S356 is solvent-exposed whereas in protofilament type m180-c2, S356 is buried inside the protofilament core. For all protofilaments, L357 is buried inside the protofilament core. The protofilament interface differs for filament types m180-a and m180-c. The protofilament interface for filament type m180-a is symmetrical and comprises residues 329-336 of one protofilament and residues V313-K321 of the opposing protofilament. The interface for filament type m180-c is asymmetrical and is formed by residues H330-E338 in protofilament m180c-1 and residues E342-R349 in protofilament m180c-2 (Figure 4.7 e-f).

After 240 min, I observed seven different filament types (m240-a, m240-b, m240-c, m240-d, m240-e, m240-f and m240-g; Figures 4.6 and 4.8). Filament type m240-a was of insufficient resolution for atomic modelling. All other filament types comprised of residues V306-R379 and consist of two protofilaments. Filament types m240-b, m240-c and m240-d consist of two asymmetric protofilaments. Protofilaments m240-b1, m240-c1 and m240-d1 adopt extended conformations whereas protofilaments m240-b2, m240-c2 and m240-d2 protofilaments are identical to each other and adopt C-shaped conformations. The extended protofilament, m240-b1 is a mixture, possibly of m240-c1 and m240-d1. 3D classification was unable to separate out the mixture. Protofilament types m240-c1 and m240-d1, differ in the orientations of the oxygen in the carbonyl groups at glycine residues 333 and 334. In
4.2 Results

Protofilament type m240-c1, G333 is oriented towards the core of the protofilament whereas in m240-d1, G333 is solvent-exposed. This results in a shift of the amino acid residue I354 towards residues $^{344}$LDF$^{346}$ and a smaller head cavity when compared to m240-d1. The difference between the extended and C-shaped protofilaments lie in the opposite in-out orientations of residues $^{327}$NIHH$^{330}$, the orientations of the oxygen in the carbonyl of G333 and G334 and the H-bond of S320 with G368 or N369. In the C-shaped protofilament, S320 forms an inter-sheet H-bond with N369. In the extended conformation, S320 forms a H-bond with G368. This is followed by opposite carbonyl orientations in the last $^{364}$PGGG$^{367}$ motif. Although opposite orientations in the $^{327}$NIHH$^{330}$ may favour a C-shaped conformation, this does not appear to determine whether a protofilament adopts an extended or C-shaped fold (Figure 4.9). Filament type m240-a has identical $^{327}$NIHH$^{330}$ in-out orientations as the C-shaped protofilaments, yet adopts an extended conformation. The protofilament interface is similar for filament types m240-b, m240-c, and m240-d, consisting of residues K311-K321 and S324-K331 from the extended protofilament and residues H329-K340 and G323-H329 from the C-shaped protofilament. The structure of filament type m240-g consists of two asymmetrical protofilaments and is similar to that of m180-c. The difference between protofilaments m240-b1, m240-c1 and m180-c1 lies in the orientations of the carbonyl backbones of G333 and G334. The carbonyl of G335 has the same orientation in all protofilaments.

After 300 min, I observed five different filament types (filament types m300-a, m300-b, m300-c, m300-d, m300-e, m300-f and m300-g; Figures 4.6 and 4.18). Filament type m300-a consists of two symmetrical protofilaments which adopt a C-shaped conformation with the AD fold. Filament types m300-b, m300-c and m300-e are PHFs. Filament types m300-d and m300-f consist of two protofilaments, where one adopts a C-shaped and the other an extended conformation. Filament type m300-g consists of four protofilaments and is a doublet of filament type m300-f. The C-shaped protofilament in filament types m300-f and m300-g is identical to the C-shaped protofilament in filament types m240-b, m240-c and m240-d. The extended protofilament, m300-d1 in filament type m300-d is identical to the extended protofilament in filament type m240-d; however, it adopts a different protofilament interface with the C-shaped protofilament. Instead, the extended protofilament in type m300-d forms an interface with the C-shaped protofilament at the C-terminus of the protofilament at residues 349-375 whilst the C-shaped protofilament interacts with the extended protofilament at residues 327-342. The extended protofilament in m300-f adopts identical conformations to the protofilaments in filament types m180-c and m240-g. Although filament type m300-g was reconstructed to insufficient resolution for de novo atomic modelling, the cross-section perpendicular to the helical axis of the reconstruction shows that this filament
Fig. 4.8 **Structures of tau filaments formed after 240 min in the presence of MgCl₂**  

a. Backbone ribbon view of the extended protofilaments from filament types m240-b and m-240-c which were formed after 240 min. Protofilaments are aligned at amino acid residues S305-G324 and D358-E380.  
b. Structure of protofilament 1 (m240-b1) from filament type m240-b.  
c. Structure of protofilament 1 (m240-c1) and  
d. Structure of protofilament 2 (m240-b2, m240-c2 and m240-d2) from filament types m240-b, m240-c and m240-d.  
e. Backbone ribbon view of the extended and C-shaped protofilaments from filament type m240-b which are aligned at residues Q336-G355.  
f-g. Protofilament interface for filament type m240-b and  
h-i. Protofilament interface for filament type m240-c.
Fig. 4.9 **Comparison of extended and C-shaped protofilaments formed at time-intervals 180-240 min in the presence of MgCl\(_2\)**  

(a) Backbone ribbon of protofilaments m180-c1 and m240-c1 from filament types m180-c and m240-c, respectively. Protofilament structures are aligned at amino acid residues S305-G324 and D358-E380. 

(b) Structures of protofilament types m180-c1 and m240-c1 showing amino acid residues K331-L357 from filament types m180-c and m240-c, respectively. 

d) Backbone ribbon view of protofilament types m180-c1, m240-b2 and m240-c2 which are aligned at residues Q336-G355 from filament types m180-c, m240-b and m240-c. 

e) Structures of protofilament types m180-c1 and m240-b, m240-c2 showing amino acid residues G323-E342.
type is a doublet of filament type m300-f. As such, I docked filament type m300-f into the density. The doublet filament interface is symmetrical and comprises residues 336-343 of the extended protofilament. The C-shaped protofilament in filament types m300-d and m400-f is similar to the C-shaped protofilaments in PHFs and adopts identical side chain orientations. The differences lie in the side chain inter- and intra-sheet interactions, and in the protofilament interfaces. In the presence of a second extended protofilament, for example in the C-shaped protofilament in filament types m300-d and m300-f, \( \text{PGGG}^{335} \) is extended, which results in a more open C-shaped conformation. Moreover, the interface of the C-shaped protofilament with the extended protofilament stabilises the open conformation. As such, the presence of an extended protofilament may open the C-shaped protofilament. Meanwhile, the protofilaments in filament type m300-a form a salt bridge at residues E342 and K343 and have no extended protofilament packing at the \( \text{PGGG}^{335} \), which results in a more closed C-shape conformation. In the cryo-EM micrographs of early formed PHFs, I observed that PHFs could form a range of cross-over distances. By using 3D classification and performing angular searches for optimal rise and twist, I could separate five different cross-over distances for PHFs within a single data-set (Figure 4.10). The PHF crossover distance for this time point ranged from 1100-2950 Å. Aligning the five different PHFs shows no major differences in the side chain rotamers nor in the curvature of the C shape of the protofilament. The distance between the carbonyl of R348, which sits at the tip of the C, and the carboxyl O of F378 at the C-terminus is 24-25 Å for all PHFs solved at this time point. However, the inter-protofilament packing of \( \beta \)-sheets along the helical axis varies. In the PHFs with longer cross-over distances, such as 2950 Å PHFs, amino acid residues at the N and C-terminus from layer (i) form \( \beta \)-sheet interactions with the layer above (i+1). In this case, N-terminal amino acid residues 305-321 in one layer (i) forms \( \beta \) inter-sheet interactions with the C-terminal amino acids 368-378 in the layer below it (i-1). In the 1100-1300 Å cross-over distances, this shift is less pronounced and appears to be in-between layers i and i-1. At later time points, PHFs adopt shorter cross-over distances, which ranges from 750-900 Å, the N-terminus and C-terminus are aligned so that \( \beta \)-sheets are formed within the same layer (i). The crossover distances for PHFs extracted from post mortem human brains, ranges from 700-850 Å.

After 360 min, I observed five different filament types (m360-a, m360-b, m360-c, m360-d, m360-e, m360-f, m360-g, m360-h, m360-i) (Figure 4.6). Types m360-a, m360-d and m360-h were PHFs. Filament types m360-b and m360-i were triple helical filaments (THFs). The latter consist of a PHF and an additional single protofilament with the Alzheimer C-shaped fold, with the additional protofilament forming a salt bridge at E342 with K343 from the third,
Fig. 4.10 **Comparison of PHF crossover distances formed after 300 min in the presence of MgCl$_2$.**

* **a** Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed in the presence of MgCl$_2$ after 300 min. Crossover distances are indicated at the bottom right of each projected slice. From left to right, crossover distances are: 2950 Å; 1280 Å; 1220 Å; 1170 Å; 1100 Å; 855 Å.

* **b-g** Secondary structure of three successive rungs in the PHF protofilament parallel to the helical axis showing N-terminus and C-terminus β-sheet packing. Crossover distances corresponding to **a** are shown: b 2950 Å; c 1280 Å; d 1220 Å; e 1170 Å; f 1100 Å; g 855 Å.

* **h-k** Atomic models showing the N-terminus and C-terminus β-sheet packing for crossover distances h 2950 Å; i 1280 Å; j 1100 Å; k 855 Å.
Fig. 4.11 *Structures of C-shaped tau protofilaments formed after 300 min in the presence of MgCl₂* a-b Backbone ribbon of tau filaments formed after 300 min of protofilaments from filament types m300-a, m300-c, m240-b2 and m240-c2. Structures are aligned at amino acid residues S305-P332 a and D358-E379 b. c-d Structures of protofilaments from filament types m300-a c and m300-c d.
adjoining protofilament. Filament type m360-c is a mixture of PHFs and THFs. Filament types m360-e and m360-g are identical to each other and consist of an extended and a C-shaped protofilament that is identical to filament types observed previously, filament type m240-d. Filament type m360-f is identical to filament type m300-a.

After 720 min, all assembly reactions had formed PHFs, with 2-10% THFs appearing in experiments 1-3, although, not enough for a 3D reconstruction.
4.2.4 Time-resolved filament formation in the presence of NaCl

I proceeded to study the rearrangement of the FIA's also in the presence of NaCl at later time-points. In the 120-180 min time-frame, I observed similar changes in the morphology of the FIA as in the experiments containing MgCl$_2$. Filaments had a range of morphologies, appeared unfolded and swollen; as a result, I was unable to to solve any of the structures within this time-frame.

![Fig. 4.13 Structures of tau filaments formed in the presence of NaCl](image)

**Fig. 4.13 Structures of tau filaments formed in the presence of NaCl** a Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis of the filaments formed in the presence of NaCl. Time in min is indicated at the top left of the projected slices. Each slice has a coloured box, which shows the point of sampling. Filament IDs are labelled at the top left of each projected slice. Four time course experiments were carried out Experiment 1 was sampled at 120, 240, 300 and 720 min; experiment 2 at 180, 360, and 720 min; experiment 3 at 180, 240 and 360 min; experiment 4 at 240, 300 and 720 min.

I continued to study the rearrangements of the filaments that had formed after 180 min. At 180 min, I observed four different filament types (filament types n180-a, n180-b, n180-c, n180-d) (Figure 4.13 and 4.14). Similar to what was found in experiments with MgCl$_2$, all filament types consisted of extended protofilaments comprising residues 305/306-378/380. Filament types n180-a, n180-b, n180-c and n180-e consist of two protofilaments, whereas filament type n180-d only has a single protofilament. Types n180-a and n180-d are symmetrical, whereas n180-b and n180-c are identical and asymmetrical. The protofilaments in
filament type n180-a is the same in n180-b1, n180-c1 and in the protofilaments of filament types n180-d and n180-e. The second protofilament in filament types n180-b and n180-c, protofilaments n180-b2 and n180-c2, adopts a more open conformation as a result of the opposite orientation of S356. In protofilaments n180-b2 and n180-c2, S356 is solvent-exposed. The protofilament interface in filament type n180-a is symmetrical and comprises residues 369-377 from one protofilament and 343-348 from the other. Protofilament interfaces in n180-b and n180-c are asymmetrical and comprise residues 331-338 from one protofilament and residues 343-348 from the other. The protofilament interface for filament type n180-e is symmetrical and comprises residues 362-367 from one protofilament and residues 348-351 from the other (Figure 4.14d-e).

After 240 min, I observed five filaments with three different filament types (filament types n240-a, n240-b, n240-c, n240-d, n240-e; Figure 4.13 and 4.15), belonging to three different structures. All filament types consist of two protofilaments in extended conformations. The filament types are similar to those observed after 180 min and mainly differ in their inter-protofilament arrangements. Filaments n240-b and n240-d, are identical; filament n240-a is similar to n240-b and n240-d, with identical side chain orientations and interprotofilament interactions, but a weak turn in the PGGG motif, resulting in a straighter appearance. The protofilament interfaces for n240-a, n240-b and n240-c are symmetrical and consist of residues 311-317 from one protofilament of n240-a and 326-331 from the other. Filament type n240-c is similar to the protofilament in filament type n180b-1, as it has identical side chain orientations. The protofilaments in n240-c differ from those of n180-b1 in the orientation of G334, which in n240-c protofilaments points out towards the solvent, whereas it points towards the layer above (i+1) in n180-b1. This leads to a slightly more open conformation in n240-c. The protofilament interface in n240-c is symmetrical and comprises residues 326-331 from both protofilaments. Filament n240-e is identical in structure to n180-b and n180-c.
Fig. 4.14 Structures of tau filaments formed after 180 min in the presence of NaCl

a Backbone ribbon of protofilaments from filament types n180-a, n180-b2 and n180-c2. The protofilaments are aligned at residues 305-324 and 358-378.
b-c Atomic structures of protofilament types n180-a and n180-b2, n180-c2.
d-e Protofilament interface for filament types n180-a and n180-b.
f Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed in the presence of NaCl of filaments formed after 180 min.
Fig. 4.15 Structures of tau filaments formed after 240 min in the presence of NaCl
Backbone ribbon of protofilament structures a n180-a, n240-a or n240-d and b n180-b1, n240-b aligned at residues 305-324 and 358-378. c-d Atomic structure of protofilament types c n240-d and d n240-c. e-f Protofilament interface for filament types e n240-c and f n240-d. g Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed in the presence of NaCl of filaments formed after 240 min.
After 300 min, I observed three different types of filaments (filament types n300-a, n300-b, n300-c, n300-d; Figure 4.13 and 4.16). All filaments comprise of two protofilaments. Filament type n300-a is identical to filament types n240-a and n240-b. Filament types n300-b and n300-c are identical to each other and are asymmetrical consisting of an extended protofilament (n300-b1; n300-c1) and a C-shaped protofilament (n300-b2; n300-c2). The extended protofilament was also observed at time-points 180 and 240 min in protofilaments from filament types n180-a, n180-b1, n180-c1, n180-d, n180-e, n240-a and n240-e1 and n-300a. The second protofilament (n300-b2; n300-c2) adopts a C-shaped conformation. The difference between the C-shaped and the extended protofilament lies in the last 364PGGG367 motif and in the opposite in-out orientations of side chains 327NIHH330. Similar observations were made for filaments assembled in the presence of MgCl₂. The protofilament interface for filament types n300-b and n300-c consists of residues 327-331 from the C-shaped protofilament (n300-b2; n300c2) and residues 347-351 from the extended protofilament (n300-b1; n300-c1). The fourth filament type, n300-d, consists of two identical protofilaments, which are C-shaped and pack against each other asymmetrically. The protofilaments in filament type n300-d are similar to the C-shaped protofilaments in filament types n300-b and n300-c but differs in the side chain orientation of L357. In the protofilaments of filament type n300-d, L357 is solvent exposed, whereas in the C-shaped protofilaments of filament types n300-b and n300-c (or n300-b2; n300-c2), L357 is oriented into the core of the protofilament. The outward orientation of L357 results in a more open C-shape. The protofilament interface is asymmetrical and comprises residues 333-340 from one protofilament and 320-325 from the other.

After 360 min, I observed five different filaments (filament types n360-a, n360-b, n360-c, n360-d, n360-e; Figure 4.13). Filament types n360-a, n360b, n360e consist of two protofilaments whereas filament types n360-c and n360-f consist of three protofilaments which are related pseudo-3₁ screw symmetry. Filaments n360-a is identical to filaments n180-a, n180-b1, n180-c1, n180-d, n180-e, n240-a, n240-e1 and n-300a. Filaments n360-b and n360-d are identical to filament type n300-d. Filaments n360-c and n360-f are similar to filaments n360-b and n360-e, but they contain an additional C-shaped protofilament which arranges itself by pseud-3₁ helical screw symmetry.

After 720 min, I observed four filaments, which fell into three distinct filament types (filament types n720-a, n720-b, n720-c, and n720-d; Figure 4.13). They comprised two identical protofilaments, like those in filament types n300-d, n360-b, and n360-c. Differences between the filament types were seen in their inter-protofilament interfaces, which were
Fig. 4.16 **Structures of tau filaments formed after 300 min in the presence of NaCl.** a-b Backbone ribbon of protofilament structures n300-b or n300-c1 and n300-d. a Aligned at residues 305-324 and 358-378 and b aligned at residues 342-355. c-d Atomic model of protofilaments e n300-b2 or n300-c2 and d n300-d. e-f Atomic models showing residues 319-325 and 361-368 for protofilament types e n300b-2 or n300c-2 and f n180-a n300-b1 n300-c1. g-h Protofilament interfaces for filament types g n300-b or n300-c and h n300-d. i Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed in the presence of NaCl of filaments formed after 300 min.
Fig. 4.17 **Structures of tau filaments formed after 720 in the presence of NaCl**

- **a** Backbone ribbon of protofilament structures n720-a, n720-c and n720-d.
- **b-d** Protofilament interface for filament types b n720-a, c n720-c and d n720-d.
- **e** Projected slices, with a thickness of approximately 4.7 Å, orthogonal to the helical axis for the filaments formed in NaCl of filaments after 720 min.
Fig. 4.18 **Comparison of extended and C-shaped protofilament structures formed in the presence of NaCl**

(a) Backbone ribbon of protofilaments n240-c with n300-b2/n300-c2 and b n180-1/n300-b1/n300-c1 with n300-b2/n300-c2 aligned at residues G355-E342. c Backbone residues 323-343 shown for protofilament types n180-a/n300-b1/n300-c1, n240-c, n240-d, n300-b2/n300-c2, n720-e.
Filament types n720-a and n720-b had a protofilament interface consisting of residues 323-329, while in filament type n720-c, the protofilament interface comprised residues 327-331 from one protofilament and residues 333-338 from the other. In filament type n720-d, the protofilament interface comprised residues 331-338. Filament types n720-a and n720-b were identical to CTE type I filaments, whereas filament type n720-d was identical to CTE type II filaments.

4.2.5 Fold conversion

FIAs represent an early stage in the formation of tau amyloid filaments, consisting of amino acid residues $^{302}$GGSVQIVYKVPDLS$^{316}$. These structures form under conditions containing either NaCl or MgCl$_2$, indicating that the structure of FIAs is independent of their presence. Once formed, FIAs can arrange into multiple different filament types. Filaments formed between 120 and 180 min are mostly untwisted, precluding their structure determination. It remains unclear how FIAs convert into the filament types observed at 180 min. As the core of the FIA represents only 16% of the sequence of the tau 297-391 construct, the fuzzy coat may play an important role in this re-arrangement. Filaments formed after 180 min under both NaCl and MgCl$_2$ conditions consisted mostly of two protofilaments with extended conformations. Although the tip of the protofilaments differed along the time course of the experiments and the conditions used, some residues, including S356, L357 and the $^{332}$PGGG$^{335}$ motif, had the ability to modulate the conformation of the protofilament tip. For instance, an outward orientation of S356 resulted in a more compact tip, whereas an inward orientation towards the protofilament core resulted in a more open conformation. The $^{332}$PGGG$^{335}$ motif residue modulated the interaction of S356 with the carbonyl group of G334 and G335. From 240 min, the extended protofilament collapsed into a C-shaped fold and, in some cases, filaments contained a mixture of extended and C-shaped protofilaments, particularly in conditions using MgCl$_2$. The primary differences between the C-shaped and extended protofilaments were in residues $^{364}$PGGC$^{367}$ and S320. In extended protofilaments, G356 formed a hydrogen bond with S320 within the same layer (i), while in C-shaped protofilaments, G355 formed a hydrogen bond with the layer below itself (i-1) (Figure 4.19). The C-shaped fold was stable and did not revert back to an extended conformation, possibly due to additional inter-sheet interactions that locked the C-shaped conformation in place.
Fig. 4.19 **Comparison of extended and C-shaped protofilaments at residues S320, G355 and G356**

**a** Cartoon view of protofilaments types m180-c1 and m720-a. Protofilaments are aligned at the N-terminus (306-320) and the C-terminus (367-376). Residues S320, G355 and G356 are shown as atoms. **b-c** Side view of two layers of protofilaments from filament types b m180-c1 and c m-720a. **d** Cartoon view of protofilaments from filament types n240-c, n240-d and n720-a, aligned at the N-terminus (306-320) and the C-terminus (367-376). **e-g** Side view of two layers of protofilaments from filament types e n240-c, f n240-d and g n720-a.

### 4.3 Discussion

The assembly of amyloid filaments comprises of a lag phase, a primary nucleation event that is followed by non-linear growth, and a plateau phase. During the lag phase, small nuclei develop that can elongate into fibrils. Fibrils are reported to appear after ThT fluorescence, however, it has been observed that in the lag phase, smaller assemblies are present which have been shown to contain -strands and β-sheets by circular dichroism, but these did not bind ThT. These smaller assemblies are often referred to as oligomers or protofibrils and have been described to contain distinct structures compared to mature fibrils (Chiti & Dobson, 2006). My results for the assembly of tau construct 297-391 show that these entities may in fact be small amyloid filaments, which I named FIAs. They fit previous descriptions of protofibrils (Harper et al., 1997; Walsh et al., 1999). FIAs are composed of a small ordered core that contains residues $^{302}$GGSVQIVYKPVDLS$^{316}$ of tau297-391. FIAs are formed in the presence of NaCl and MgCl$_2$, even though their end-stage filaments are different. This suggests a common mechanism of primary nucleation under the two conditions. FIAs are ThT-negative and have the ability to arrange into different filament structures, according to the chemical environment. Notably, certain folds are shared between conditions. At early time-points, the protofilaments adopted extended conformations, primarily comprising
two protofilaments. These extended folds are distinct from the C-shaped folds, mainly at residues $^{364}$PGGG$^{367}$ and S320. In the extended fold, these residues interact within the same layer, while in the C-shaped fold, they interact with the layer below. Moreover, the in-out orientations of residues $^{327}$NIHH$^{330}$ may favour the C-shaped folds, while the extended folds can accommodate either orientation. The C-shaped folds appears to be stable and irreversible under conditions containing NaCl or MgCl$_2$. The transition from FIAs into extended protofilaments remains unclear and requires further research.

The finding that FIAs do not bind ThT highlights the need to reassess the interpretation of data from ThT fluorescence measurements. The concentration of amyloid filaments has previously been assumed to correlate directly with ThT fluorescence intensity. However, the molecular basis of ThT binding to amyloid filaments is only poorly understood and may depend on the presence of specific filament folds. Moreover, fluorescence intensity, which has previously been used to distinguish between different filament polymorphs (Shahnawaz, Mukherjee, et al., 2020), may be influenced by the number of ThT binding sites that a particular fold can accommodate. Therefore, it is important to gain a better understanding of the molecular mechanisms of ThT binding to amyloid filaments and how this relates to fluorescence intensity. Meanwhile, it is important to approach the interpretation of amyloid assembly kinetics through ThT fluorescence data with caution.

It has been shown that residues $^{306}$VQIVYK$^{311}$, also known as PHF6, are essential for the assembly of tau into filaments (Goedert, 2018; W. Li & V. M.-Y. Lee, 2006; von Bergen et al., 2000). The FIA structure contains the PHF6 motif, with residues $^{302}$GGSVQIVYKPV$^{316}$ comprising the core of the early intermediate filament. All known tau filament structures from human brains also contain the PHF6 sequence in their cores (Falcon, Zhang, Murzin, et al., 2018; Falcon, Zivanov, et al., 2019; Fitzpatrick et al., 2017; Y. Shi, Murzin, et al., 2021; Zhang, Tarutani, et al., 2020). Some MAPT mutations, such as P301S/L/T, cause inherited forms of tauopathies (Goedert, 2018). Moreover, G303V, G304S and S305L, S305N and S305S also cause inherited forms of disease, although these mutations have been shown to have primary effects at the mRNA level (Ghetti et al., 2015). However, these mutations are all located close to the start of the FIA core and may affect the ability of FIAs to form. Therefore, investigating the impact of these mutations on the formation of tau FIAs and their correlation to mature tau filaments may provide further insights into the molecular mechanisms underlying familial cases of tauopathies.
The experiments described in this chapter used a tau fragment and were carried out \textit{in vitro}; it remains to be determined if they reflected the mechanisms of amyloid assembly in brain. Future time course experiments using full-length tau constructs, which can replicate disease relevant structures, will be important to understand amyloid filament formation.
Chapter 5

Discussion

Neurodegenerative diseases are characterised by the accumulation of protein inclusions composed of amyloids. Cryo-EM studies of amyloid filaments isolated from human brains have revealed that specific conformers of tau and α-synuclein are associated with disease (Fitzpatrick et al., 2017; Y. Shi, Zhang, et al., 2021; Yang, Arseni, et al., 2022; Yang, Garringer, et al., 2022). These findings suggest that specific mechanisms may lead to the disease-specific conformers. However, studying the molecular mechanisms of amyloid assembly in post mortem samples is impossible, as they represent the end-stage of the disease. Therefore, the in vitro assembly of amyloids provides unique opportunities to gain insights into the underlying mechanisms of these diseases and may lead to the development of future therapeutic interventions. In this PhD thesis, I have set out a path towards studying disease-relevant, in vitro amyloid filament formation.

5.1 Overview

In Chapter 2, I showed the in vitro seeded assembly of MSA extracts into recombinant α-synuclein protein. I observed that the amplified structures did not accurately replicate the structures of the seeds. A subsequent study that employed seeds obtained from the cerebrospinal fluid (CSF) of individuals at various stages of PD, resulted in the same structures (Burger et al., 2021). Both studies used the same PMCA assay and aggregation conditions, suggesting that the in vitro conditions, rather than the structure of the seeds, determined which structures were formed (Shahnawaz, Mukherjee, et al., 2020). Nevertheless, Shahnawaz et al. reported that one can distinguish between MSA and PD after PMCA. Further studies are required to understand the molecular mechanisms of seeded aggregation, possibly using time-resolved cryo-EM. This may provide insight into the relationship between the chemical
environment and the function of seeds in the formation of amyloid filaments.

In chapter 3, I described the *in vitro* assembly of N and C-terminally truncated tau constructs into filaments. This study showed that tau can form multiple different amyloid structures and that the amyloid assembly of tau is influenced by the assembly environment. Assembly conditions containing NaCl yielded structures identical to those observed in CTE brains, while the addition of MgCl$_2$, CaCl$_2$, and dextran sulfate resulted in structures identical to those in AD brains. The tau constructs could be extended towards 258 and 396 in the N and C-termini, respectively, while still forming the same structures. Extending the construct further at the C-terminus required mutating serines and threonines, particularly at residues 396, 402, 404 and 408 into aspartates, mimicking the negatively charged phosphate group of a phosphorylated serine and threonine. In AD and CTE, full-length tau comprising all six tau isoforms probably assembles into filaments. Although, truncation of a small amount of tau giving rise to a seed, cannot be excluded. Future *in vitro* assembly experiments using phospho- and acetylation mimetics on full-length tau will be required. Further exploration of *in vitro* assembly of constructs and assembly conditions will also be required for the development of other disease conformers.

This study also led to the development of high-throughput approaches for the structure determination of amyloids. High-throughput cryo-EM reconstruction methods are necessary for the rapid screening of large numbers of samples, which is particularly important for studying the structural heterogeneity of amyloid fibrils. By rapidly screening multiple samples, we can identify the role of the chemical environment in amyloid assembly for the development of disease-specific amyloid conformers. While significant progress has been made in amyloid reconstruction, there is still a long way to go, particularly in automation. For example, automated picking often requires user input, trial and error and is not always successful. Manual picking remains superior. Further developments in automated picking will greatly aid the throughput of amyloid reconstruction. Furthermore, *in vitro* assembly of amyloids often yields a large number of polymorphic structures, which can have similar structures that may be missed and unresolved. Additional 2D classification algorithms or clustering methods capable of identifying specific polymorphs and mixtures of polymorphs would be helpful. Another major bottleneck in amyloid reconstruction is generating initial 3D-models. Currently, a projection model representing one full cross-over distance of the amyloid is generated using 2D classes, 2D back-projection then leads to a reconstruction of the filament XY cross-section. This process often involves considerable trial and error. Stochastic gradient descent, which is used to generate initial models for SPA, could poten-
5.1 Overview

Initially assist in automating initial 3D model generation. Moreover, a considerable number of amyloid structures have now been solved, presenting possibilities to train machine models on 2D classifications and their corresponding 3D reconstructions. Deep neural network models could potentially learn the relationships between the 2D images and the 3D structures, and use this knowledge to generate a 3D model directly from a new set of 2D images.

In chapter 4, I showed that the in vitro assembly of N and C-terminally truncated tau is a dynamic and step-wise process. During the lag phase of assembly, I observed filaments which comprise of a small filament core. I named these filaments, First Intermediate Amyloids (FIA) as they are the first structures that appear in the time-resolved assembly of tau. The tau FIA can rearrange into multiple different amyloid structures in a time dependent manner, and their structures are dependent on the environmental conditions. This highlights the complexity of amyloid assembly and indicates that this is not a straightforward "add-on", process in which monomers are added to the ends of mature filaments for elongation. I also found that the tau FIA does not bind to ThT. This suggests that the relationship between ThT fluorescence and filament concentration may not be as straightforward as previously assumed. Different filament structures may have different binding capacities and affinities for ThT. A detailed characterisation of how ThT binds to various filament structures is required to accurately model the kinetics of amyloid assembly. Subsequently, I showed that the tau FIA has the ability to rearrange into specific filament types during the maturation of the filaments. Although AD and CTE folds are different, they share certain conformations during their maturation. Initially, the filaments consist mostly of two protofilaments that take on extended conformations. Gradual changes in the orientations of the side chains may promote the folding of the extended protofilament into a C-shaped structure. However, the exact changes required for this curvature are not entirely clear. A key difference between the extended and C-shaped protofilaments lies in the hydrogen bonds between β-sheet layers. While the extended protofilament forms hydrogen bonds only within its own layer, the side chains of the C-shaped protofilament form hydrogen bonds with additional layers. Residues S320 and GG366 play a significant role in the transformation of the extended protofilament into a C-shaped one. Once the extended protofilament has converted into a C-shaped protofilament, the C-shaped protofilament does not revert back to an extended conformation. This is likely due to the additional hydrogen bonds the C-shaped protofilament forms with other layers that locks in the C-shaped conformation. Additional time-course experiments and further processing of the existing data-sets will be necessary to fully understand how filament maturation occurs at the atomic level.
5.2 Outlook

Neurodegeneration is a complex process that may have to progress for several decades until it results in symptoms of disease, with in vitro assembly being unable to provide a complete understanding of what happens in the disease. In the future, it will therefore be important to develop new model systems to study amyloids including cellular and animal models. It is important to ensure that the models used, accurately recapitulate the amyloid conformers observed in disease. The constructs and conditions identified in chapter 3 may guide future cellular and animal models that recapitulate the filament structures that form in disease. In chapter 4, I described First Intermediate Amyloids (FIAs) as the earliest stage in the amyloid filament formation of tau. It will be important to determine if tau FIAs also form inside cells. Studying filament formation in a cellular context can help us to understand how cells handle filament formation and identify the point at which toxicity sets in. Specific tau conformers are associated with particular tauopathies, and selectivity for distinct strains to propagate in certain cell types has been observed in disease. Therefore, understanding the propagation of strains in different cell types is important. Additionally, investigating whether particular strains are more toxic than others, even if they are not associated with disease, will provide insights into the mechanisms underlying the toxicity of conformers. This could inform the development of novel therapeutic strategies for neurodegenerative diseases by identifying the cell types to target in the early stages of disease.

Ultimately, neurodegenerative diseases impact much of the brain. Having disease models, in C. elegans, Drosophila melanogaster (fruit fly) and mice will aid our understanding of neurodegeneration and will be useful for early-stage drug trials. Many C. elegans and fruit fly models have been developed to study the aggregation of tau and α-synuclein and subsequent neurodegeneration. However, these models lack cryo-EM structures of the amyloid filaments, and it is uncertain whether they form filaments that are the same as those observed in human disease. Developing C. elegans and fruit fly models with amyloid filaments identical to those in disease will be important for understanding the molecular basis of neurodegenerative diseases in an easily perturbable system. Several mouse models have also been developed for tau and α-synuclein, however the structures of the filaments that form in these mice remain unknown. Furthermore, the majority of mouse models in use are those with familial mutations in the genes encoding the amyloids. As a result, the structures generated by these models may represent only a minority of the diseases. Despite this, two amyloid structures of Aβ have been solved from mouse models. The Aβ filaments of the mouse model APPNL-F were found to be identical to type II Aβ filaments observed in human brains (Yang, Arseni, et al., 2022). This model can be used to study how Aβ assembles into filaments and how
these filaments affect the mouse brain over time. However, the Aβ filaments extracted from App\textsuperscript{NL−G−F} knock-in mice had different structures to those observed in human brains, including in individuals with the Arctic mutation, thus raising questions about its relevance (Yang, Zhang, et al., 2023). Further research will be required to develop mouse models that recapitulate the structures of tau and α-synuclein filaments observed in disease.

Improved diagnostic and therapeutic strategies for neurodegenerative diseases may result in better detection of amyloid and a greater reduction in its levels in brain. Positron emission tomography (PET) is a non-invasive imaging technique that enables the detection and mapping of amyloid accumulation in the brain. To aid diagnosis, PET ligands have been developed for detecting Aβ plaques and tau filaments in AD. Among the most commonly used PET ligands are Pittsburgh compound B (PiB) for Aβ plaques and flortaucipir for assembled tau (La Joie et al., 2020). In vitro assembly of tau into filaments that are identical to those found in AD and CTE can facilitate the development of PET ligands by enabling mapping of their binding sites and measuring their affinity to the filaments. This may lead to the development of high-affinity PET ligands, and more accurate readouts of the particular strains that are affecting the human brain, thereby helping the recruitment of clinical trial participants. Various drugs have been developed to target amyloid deposits. They work through distinct mechanisms, including gene silencing, protein stabilisation, and antibody binding to amyloid. For example, RNA interference (RNAi) therapeutics, such as Onpattro and Tegsedi, silence the production of transthyretin protein by targeting its gene (Coelho et al., 2013). In contrast, aducanumab and lecanemab are humanised antibodies that target Aβ plaques and have demonstrated promise in reducing Aβ amyloid levels and cognitive decline in individuals with early-stage AD (Selkoe, 2019; M. Shi et al., 2022; Swanson et al., 2021; van Dyck et al., 2023). Future research on amyloid-targeting drugs may help amyloid related diseases. I identified FIAs formed by regions in the tau sequence, providing new considerations for disease intervention. FIAs were identical in conditions resulting in end-stage AD and CTE tau folds, and they contain the amyloid-prone region previously described for tau (\textsuperscript{306}VQIVYK\textsuperscript{311}, PHF6) that has been shown to be essential for tau filament formation (von Bergen et al., 2000). Targeting tau FIAs may prevent both AD and CTE amyloid filament formation. Amyloid-prone regions have also been described for other amyloid-forming proteins (Sabate et al., 2015), and similar mechanisms may thus occur in other amyloid-related diseases. Targeting amyloid-prone regions which form FIAs, through small molecule binders, could inhibit primary nucleation of amyloids and prevent initiation of amyloid-related diseases.
The relationship between amyloids and neurodegeneration is complex and multi-faceted. To understand and treat neurodegenerative diseases, collaboration and cooperation among experts are essential. It is important to recognise that the study of amyloids and neurodegeneration is a long-term endeavour that will require a sustained effort and investment from all disciplines. Ultimately, the goal is to prevent these diseases from happening.
Chapter 6

Methods

This chapter uses some text and figures from my publications, Lövestam et al. (2021) FEBS OpenBio, Lövestam et al. (2022) eLife and Lövestam and Scheres (2022) Faraday discussions.

6.1 Molecular cloning

For all the cloning procedures for α-synculein and tau, I used in vivo assembly (IVA) (García-Nafría et al., 2016). IVA utilises recA-independent recombination pathway, which exists in the majority of E. coli systems, thus removing enzymatic assemblies and limiting the reaction to a single test tube. The reaction only takes 2 hrs and requires minimal molecular cloning. Figure 6.1 outlines the concept of IVA. In brief, for insertions, the new sequence is included in both the forward and the reverse primers, which will act as homologous regions during transformation. For deletions, the forward primer will anneal upstream to the site of deletion, whilst the reverse strand will simply anneal to the plasmid with the homologous regions, removing the undesired sequence. For mutagenesis, the reverse primer will flank the mutation and the replacement will be encoded in the forward primer. After the PCR, DpnI is added to the PCR product for 30 min at 37 °C, to remove any methylated template DNA. The digested product is transformed into XL10-gold (Agilent) and a few colonies are picked, grown overnight (ON) in 2xTY supplemented with 100 mg/l ampicillin and sent for sequencing. All sequences were forward and reversed sequenced using the T7 promoter.
Methods

6.2 Protein purification for α-synuclein

I expressed and purified α-synuclein, essentially as previously described (Morgan et al., 2020). Briefly, plasmid pRK172 encoding a cDNA for full-length, wild-type human α-synuclein was transformed into E. coli BL21 (DE3)-gold (Agilent Technologies LDA UK Limited, Stockport, UK). Cells were cultured in 2xTY, 5 mM MgSO_4 and 100 mg/l ampicillin at 37 °C until an Optical Density (OD)_{600} of 0.7 was reached; α-synuclein expression was then induced with 1 mM IPTG. After 4 hrs, cells were harvested by centrifugation and resuspended in cold buffer A: 50 mM Tris/HCl, pH 7.5, 10 mM EDTA, 2.5 mm TCEP (Sigma-Aldrich, Gillingham, UK), 0.1 mM AEBSF (Sigma-Aldrich), 40 µg/ml DNase and 10 µg/ml RNase (Sigma-Aldrich), supplemented with cOmplete EDTA-free Protease Inhibitor Cocktail (Roche, Welwyn Garden City, UK). They were lysed by sonication on ice using a Sonics VCX-750 Vibra Cell Ultra Sonic Processor for 5 min (5 s on, 10 s off) at 40% amplitude. The lysates were centrifuged at 20,000 xg for 40 min at 4 °C, filtered with a 0.45 µm cut-off filter. The pH of the lysate was brought to a pH of 3.5 using HCl and stirred at RT for 30 min. The lysate was centrifuged at 50,000 xg for 1 hr at 4 °C and the supernatant was loaded onto an anion exchange HiTrap Q (GE Healthcare, Chalfont Saint Giles, UK) and eluted with a 0–1 M NaCl gradient, collecting fractions of 1 ml. Fractions containing α-synuclein were precipitated using ammonium sulphate (0.3 g/ml) for 30 min at 4 °C and centrifuged at 20,000 xg for 30 min at 4 °C. The resulting pellets were resuspended in buffer B.
(PBS, 0.1 mM AEBSF, supplemented with cOmplete EDTA-free Protease Inhibitor Cocktail), loaded onto a HiLoad 16/60 Superdex (GE Healthcare) column equilibrated in buffer B and eluted using a flow rate of 1 ml/min. The purity of α-synuclein was analysed by 4-20 % Tris-glycine SDS/PAGE and protein concentrations determined spectrophotometrically using an extinction coefficient of 5600 M$^{-1}$·cm$^{-1}$.

### 6.3 Protein purification for tau

I expressed and purified tau, essentially as described Studier, 2005 Expression of tau was carried out in *E. coli* BL21 (DE3) gold or BL21(DE3)pLysS. For expression, I used two different approaches. Approach one: A single colony was inoculated into 500 ml lysogeny broth (LB) auto-induction media and grown for 8 hr at 37°C, followed by subsequent expression for 16 hr at 24°C. Approach two: A plate of freshly transformed cells was resuspended in 10 ml of 2xTY and inoculated in 500 ml of 2xTY supplemented with 5 mM MgSO4 and grown until an O.D$_{600}$ of 0.8 was reached. Expression was subsequently induced with IPTG for 3-4 hrs at 37 °C or ON at 20-24 °C.

Purification proceeded the same for both expression systems. Cells were harvested by centrifugation (4000 ×g for 20 min at 4°C), and resuspended in washing buffer (WB: 50 mM MES at pH 6.0; 10 mM EDTA; 10 mM DTT, supplemented with 0.1 mM PMSF and cOmplete EDTA-free protease cocktail inhibitors, at 10 ml/g of pellet). For full-length and C-terminal containing constructs of tau, additional protease inhibitors such as AEBSF and chymostatin were required. This is likely due to the large number of serines and threonines clustered in the proline rich region and in the C-terminus of tau. Cell lysis was performed using sonication (either at 40% or 95% amplitude using a Sonics VCX-750 Vibracell Ultra Sonic Processor for 7 min, 5 s on/10 s off). Lysed cells were centrifuged at 20,000 ×g for 40 min at 4°C, filtered through 0.45 μm cut-off filters and loaded onto a HiTrap Capto S HP 5 ml column (GE Healthcare) for cation exchange. The column was washed with 10 column vol of WB and eluted using a gradient of WB containing 0–1 M NaCl. Fractions of 3.5 ml were collected and analysed by SDS-PAGE bis-TRIS 4–12% or Tris Glycine 4–20%. Protein-containing fractions were pooled and precipitated using 0.3 g/ml ammonium sulphate and left on a rocker for 30 min at 4°C. Precipitated proteins were then centrifuged at 20,000 xg for 35 min at 4°C, and resuspended in 2 ml of 10 mM PB, pH 7.2–7.4, with 10 mM DTT, and loaded onto a 60/10 Superdex size exclusion column. Size exclusion fractions were analysed by SDS-PAGE, protein-containing fractions pooled and concentrated to 6-50 mg/ml using molecular weight concentrators with a cut-off filter of 3 kDa. Purified protein samples were flash frozen in 50-100 μl aliquots and stored at -20 °C for future use.
6.3.1 Considerations for the purification of amyloid proteins

For amyloid studies, it is essential to obtain highly pure protein for reproducible studies of the self-assembly reactions for kinetic and structural analysis. My studies on α-synuclein and tau, as well as other studies (Linse, 2020) have shown that amyloid proteins are highly polymorphic. Protein purity, sequence homogeneity and buffer purity are therefore critical for us to decipher the molecular determinants of a reaction. As such, it is important to consider all facets of protein purification when carrying out an assembly reaction. For example, the use of tags such as x6His will dramatically affect the sensitivity of the polypeptide chain to its pH environment. Purifying as described above without a tag also comes with disadvantages: ion exchange uses high NaCl concentrations which will stick to the protein throughout the purification process. Ammonium sulfate, which is used in the precipitation step will also stick to the protein. This is observed subsequently in the SEC, where after a column run has been performed, an increase in conductivity is observed.

6.4 Purification of α-synuclein filaments from MSA brains

The filament preparations used in this study have been described (Schweighauser et al., 2020). Briefly, frozen putamen from MSA cases 1, 2 and 5 was homogenised in 20% vol (w/v) extraction buffer (10 mM Tris/HCl, pH 7.5, 0.8 M NaCl, 1 mM EGTA, 10-20% sucrose, 1-2% sarkosyl, pH 7.5) and incubated for 30 min at 37 °C. The homogenates were centrifuged for 10 min at 10,000 xg at room temperature, followed by a 20 min spin of the resulting supernatants at 100,000 xg. The pellets were resuspended in 500 µl/g extraction buffer and centrifuged at 3,000 xg for 5 min to remove large contaminants. The supernatants were diluted in 50 mM Tris/HCl, pH 7.5, containing 150 mM NaCl, 10% sucrose and 0.2% sarkosyl, and centrifuged at 166,000 xg for 30 min. Sarkosyl-insoluble pellets were resuspended in 50 µl/g tissue and filament concentrations estimated by negative-stain EM. Prior to seeded assembly experiments, pellets were centrifuged at 2,000 xg for 5 min, the resulting supernatants were diluted 10-fold, and sonicated in an Eppendorf tube using a VialTweeter (Hielscher Ultrasonics GmbH, Teltow, Germany) at a cumulative power of 100 W. Sonication did not alter the structure of the seeds, as suggested by negative-stain EM (see chapter 2), and as confirmed by cryo-EM 2D class averages of the seeds before and after sonication.
6.5 Seeded assembly of recombinant α-synuclein with MSA seeds

Purified recombinant α-synuclein was centrifuged at 20,000 xg for 1 hr to remove potential aggregates. 70 µm recombinant α-synuclein was incubated with 2 µm MSA seeds (as assessed by negative-stain EM) in 100 mM PIPES pH 6.5, 500 mM NaCl, 0.05% NaN3 and 5 µm thioflavin-T, in a final volume of 200 µl per experiment. Controls used buffer without seeds. A low concentration of recombinant α-synuclein was used to prevent its spontaneous assembly. Seeded assembly proceeded for 120 h at 37 °C in a FLUOstar Omega (BMG Labtech, Aylesbury, UK) microplate reader where the samples were alternately shaken for 1 min at 400 rpm, and left to rest for 1 min, during which fluorescence was measured.

For cryo-EM, seeded assembly conditions were identical, but no thioflavin-T was added to the buffer and samples were shaken continuously for 72 h. Seeded assembly experiments for cryo-EM were also performed in PBS buffer, supplemented with 1 mm pyrophosphate and 0.05% NaN3. The resulting filaments were pelleted, resuspended in 200 µl and sonicated as described above, and then used as seeds (2 µm) for a second-generation seeded assembly experiment with recombinant α-synuclein (70 µm) in the same PBS buffer.

6.6 Solubility assays

To test whether in vitro seeded α-synuclein filaments would be stable in the conditions used for extraction of seeds from MSA brains, filaments were briefly homogenised in A68 buffer (10-20% Sucrose 50 mM Tris pH 7.4 800 mM NaCl 5 mM EGTA 5 mM EDTA supplemented with EDTA-free Complete protease inhibitors) incubated with 1% of sarkosyl or SDS at 37 °C for 30 min. The reaction was centrifuged at 100,000 xg for 30 min and the pellet was resuspended in 100 mM PIPES pH 6.5, 500 mM NaCl. The supernatant and pellet for each procedure was checked separately and were loaded on an SDS-PAGE Tris Glycine 4–20%.

6.7 Assembly of recombinant tau

Purified protein samples were thawed at RT and centrifuged 20,000 xg for 10 min at 4 °C. See table 1 for conditions. Buffers were filtered through 0.22 µM nylon filters. Assembly reactions were made in a total volume of 150 L. Reactions were mixed at RT to prevent any precipitation of phosphates and salts at 4 °C. Mixing was performed in order of H2O, PB,
DTT, salt* and last the protein. All components of the reaction were at RT prior to mixing. The reactions were filtered through 0.22 µM filters (corning costar Spin-X centrifuge tube filters, 0.22 µM). Note, in some cases, for example ZnCl₂, a small amount of precipitation was observed during mixing. Thoroughly mixing the solution removed the aggregates. 98 µL of the mixture was mixed with 2 µl of filtered ThT (150 µM) for measurements. Triplicates of 30 µl of the ThT containing mixture was aliquoted into pre-washed 384 well plates and a control of 30 µl not containing ThT was used for cryo-EM studies. Assembly was carried out in a FLUOstar Omega (BMG Labtech, Aylesbury, United Kingdom) using double orbital shaking continuously at 37 ºC. The gain was set by using the saturation of a previous reaction set to 70 - 80 % of the maximum fluorescence. ThT fluorescence was measured every 10 min of selected wells (using 20-100 flashes) with an excitation filter of 440 nm and an emission filter of 480 nm.

6.8 Cryo-EM data acquisition

Cryo-EM grid preparation proceeded similarly for all studies described in this thesis. For the seeded assembly of α-synuclein and tau constructs, samples were centrifuged 3000 xg prior to grid preparation. For time-resolved cryo-EM of intermediates, microplates were taken directly from the shaker to the vitrobot (Thermo Fisher Scientific, TFS). The microplates were removed from the plate reader for a maximum of 10-20 minutes at a time. 3 µl aliquots were applied to glow-discharged R1.2/1.3 or R0.6/1.0, 300 mesh carbon Au grids that were plunge-frozen in liquid ethane using a Thermo Fisher Scientific Vitrobot Mark IV. Cryo-EM data were acquired at the MRC Laboratory of Molecular Biology (LMB) and at the Research and Development facility of Thermo Fisher Scientific in Eindhoven by members of the team of Abhay Kotecha. All images were recorded at a dose of 30–40 electrons per Å², using EPU software (Thermo Fisher Scientific), and converted to tiff format using relion_convert_to_tiff prior to processing.

At LMB, images were recorded on a Krios G1 with a K2 or K3 camera (Gatan), using an energy slit of 20 eV on a Gatan energy filter, or on a Krios G2 with a Falcon 4 camera (Thermo Fisher Scientific), without an energy filter. At TFS, images were recorded on a Krios G4 with a CFEG, a Falcon 4i camera, and a Selectris X energy filter that was used with a slit width of 10 eV. At TFS, cryo-EM grid screening and part of data acquisition were performed using EPU-Multigrid on a Glacios microscope equipped with a Selectris X energy filter (Thermo Fisher Scientific) and a Falcon 4 camera. Eight grids were loaded for each 48 hr EPU-Multigrid run. Each grid was loaded to the stage for session set-up, grid
square selection, and ice-filter preparation. The session information was stored in EPU and associated with the grid position in the autoloader. After all the grids sessions were created and stored in queue, twofold astigmatism was corrected and the beam tilt was adjusted to the coma-free axis using automatic functions within EPU. Before starting the EPU Multigrid queue, and once for each 48 hr session, the Selectris X filter slit was cantered, and the filter tuned for isochromaticity, magnification, and chromatic aberrations, using Sherpa software (Thermo Fisher Scientific, TFS). During the fully automated EPU Multigrid runs, each grid was loaded onto the stage, grid squares were brought to eucentric height, and holes were selected with the stored ice filter settings. For each grid, 2000 images were collected with a 10 eV energy filter slit width. Images were collected in electron event registration (EER) mode using the aberration-free image shift method in EPU version 2.12. Under these conditions, a throughput of 200–250 images per hour was achieved depending on the number of holes available per grid square.

### 6.9 Cryo-EM image processing

#### General Strategy for amyloid structure determination in Relion

Structure determination of amyloids in Relion follows a broadly similar workflow as in conventional single-particle analysis of globular proteins. Because this thesis describes more than 200 structures, this section provides an overview of the general strategy and relevant differences between the two modalities. However, every data set is different, and it is hard to provide one approach that is suitable for all.

**Micrograph inspection**

Some data sets are not worth acquiring. Compared to many globular protein complexes, amyloids are relatively sturdy objects, but sometimes they do get damaged during sample preparation. Processing images of such filaments is likely to fail. One feature to look out for when acquiring data are what I call ‘swollen filaments’, which appear “blobby” with inconsistent widths along the filament axis, and result in poor 2D class averages. In addition, although it is in principle possible to solve the structure of filaments that do not twist, in practice this is often hampered by strong preferred orientations. Fig. 2 shows examples of good filaments, swollen filaments and non-twisting, flat ribbons.

**Micrograph pre-processing and filament picking**

Once suitable images have been recorded, micrograph movies are motion-corrected. I recommend using Relion’s own implementation of the UCSF MotionCor2 program (Zheng et al., 2017) which communicates metadata with particle polishing. Next, contrast transfer
Methods

Fig. 6.2 Cryo-EM Micrograph inspection of amyloid filaments Examples of micrographs are shown for good filaments (A), swollen filaments (B) and non-twisting ribbons (C). Insets on the top right show representative 2D class averages for the filaments shown.

function (CTF) parameters are estimated, preferably using the open-source CTFFIND4 program (Rohou & Grigorieff, 2015). Filaments are picked manually in the micrographs by clicking start–end coordinate pairs or picked automatically using the modified Topaz approach described above. The start–end coordinate pairs define straight lines. Curved filaments are picked as multiple shorter lines. Individual particles images are extracted along these lines with an inter-particle distance that is defined by the helical rise (at this point one can provide a value of 4.75 Å) and the number of unique asymmetrical units (a value of 3 works well in most cases). It is often better to pick fewer, good filaments than larger numbers of suboptimal filaments. Filaments with anisotropically shaped cross-sections display alternating strong and weak signals in the micrographs. To avoid missing the weaker parts, a relatively low Topaz threshold may be necessary. I recommend to visually check auto-picking results in several micrographs.

2D classification

2D class averaging serves to assess the quality of the data, to remove sub-optimal particles, and to detect the presence of multiple filament types. The latter requires user expertise if the filament types are similar. To aid novel users, I elaborate on this step for the second and third data sets that are described in the Results section. The mask diameter for 2D classification is set close to the box size; the angular sampling rate to 1–2°. The VDAM algorithm in Relion-4.0 (Kimanius et al., 2021) works less well for amyloids than it does for globular proteins and better results are often obtained with the (slower) default algorithm. It is often useful to calculate 2D class averages with a few different box sizes. Many final reconstructions are calculated in box sizes of approximately 250–300 Å, but earlier 2D classifications with box sizes in the range of 500–1000 Å are also useful. Although typically
of lower resolution, larger 2D class averages facilitate identification of distinct filament types, can help measuring cross-over distances, and reduce the number of variables to optimise in the relion_helix_inimodel2d program for initial model generation.

**Initial model generation**

Subsets of 2D class averages that correspond to a single filament type are used to calculate initial 3D models in the relion_helix_inimodel2d program. This approach has been described in detail (S. H. W. Scheres, 2020; S. H. Scheres et al., 2020). Nevertheless, reliable initial model generation remains the biggest hurdle in many amyloid structure determination projects. The initial model generation consists of a 2D reconstruction of the XY cross-section of the filaments. Therefore, the results are best assessed by visual inspection of 2D images, rather than the 3D model. The output files rec.spi (the 2D reconstruction), before_reproject.spi (the summed 2D class averages along the cross-over) and after_reproject.spi (the projected 2D reconstruction along the cross-over) are over-written at every iteration and can be monitored with a display program that re-reads the images every time they change on disk (I use Xmipp-2.4). The algorithm should converge, with few changes to these images in the last iteration. The rec.spi image of good models typically has higher contrast (white signal against a black background) than that of suboptimal models. Good models also show continuous main-chain density, possibly even with density for bulky side chains. Multiple disconnected densities and streaks of density that extend into the solvent area are typical of suboptimal solutions. The before_reproject.spi image should have 2D class average images along the entire cross-over, without large discontinuities between them, and the after_reproject.spi image should resemble the before_reproject.spi image.

2D class averages displaying filaments that are not centred in the Y-direction or that are not oriented horizontally can be aligned during the 2D reconstruction process using the arguments –search_shift, –search_angle and –step_angle. In difficult cases, pre-alignment of the images (in Relion or other software) may give better results. If the rec.spi image appears symmetrical, rotational symmetry can be imposed (using the –sym argument) to aid convergence. Using a mask on the 2D reconstruction (–mask_diameter) and limiting the resolution of the 2D reconstruction (–maxres) speed up the calculations and may facilitate convergence. The program has also been parallelised (–j), with multiple threads each aligning subsets of the 2D class averages. Quickly performing multiple runs and interactively monitoring their results facilitates parameter tuning. The most important variables to tune are the –crossover_distance parameter (because it may be difficult to estimate the cross-over distance from the micrographs or the 2D class averages) and the selection of which 2D class
averages to use (because it may be difficult to recognise distinct filament types).

**3D refinement**

The initial model is used for 3D auto-refinement. The initial low-pass filter applied to the initial model varies with the quality of the model, but is typically around 10 Å. The initial model is not on the same grey scale as the particles extracted for the refinement. At this stage, a helical rise of 4.75 Å (or the adjusted value if the pixel size calibration was off, see the Pitfalls section below) is used, together with a helical twist that is calculated as \((4.75 \times 180°)/d\), with \(d\) being the cross-over distance (in Å) as measured in the micrographs. Typically, no point group symmetry is applied during the first refinement. Because the resolution of the model does not yet extend beyond 4.75 Å, helical twist and rise parameters are also not refined at this stage.

Refinement should result in a substantial gain in resolution over the initial model and, more importantly, result in amyloid-like features in the map, including separated β-strands and connected main-chain density with convincing side chains. If this is not the case, a better initial model may be necessary. If additional symmetry is present in the map, the symmetry operators need to be determined. For example, if two symmetrical protofilaments are visible, the individual molecules could be related by C2 point group symmetry, or by pseudo-21 helical screw symmetry. If these symmetries cannot be distinguished from the reconstruction, subsequent refinements with either of these options should be explored. The correct symmetry will typically lead to an increase in resolution, whereas the incorrect symmetry will often prevent the map from acquiring good separation of the β-strands. Once refinement with good β-strand separation is achieved, the helical twist and rise may be optimised. For this purpose, an initial reference with good β-strand separation is used for another 3D auto-refinement job with an initial low-pass filter close to 5 Å. To avoid overfitting, one needs to provide one of the two half-maps from a previous refinement as the initial reference. How initial models are dealt with in 3D auto-refinement has changed in Relion-4.0. If the filename contains the substring half1 or half2, then both half-maps are read and set as the initial models for the two separate half-map refinements. In previous versions of Relion, the same initial model was always used for both halves. This could lead to severe overfitting, as explained in the Pitfalls section below. In rare cases, typically with relatively noisy data, the Sidesplitter program, which is invoked through the –external_reconstruct argument, leads to better reconstructions than the default auto-refinement algorithm.
3D classification
The separation of particle images into structurally homogeneous subsets does not work as well for amyloids as it does for single-particle analysis of globular proteins. Nevertheless, 3D classification is useful for the separation of filament types that are relatively similar to each other, in particular when used without further alignment of the individual particle images. Varying the regularisation parameter \(T = 4–100\) may help. One particularly useful application of 3D classification is the identification of suboptimal particles, which tend to separate from the good particles into different classes. Again, including fewer, better particles often yields better reconstructions than using more, suboptimal ones.

Particle polishing and CTF refinement
Beam-induced motion correction by particle polishing is typically done earlier in the structure determination process of amyloids than it is for globular proteins. The rationale behind this is that beam-induced motions can often be detected even when the reference map does not yet show all the expected features of an amyloid. If \(\beta\)-strand separation in early refinements is suboptimal, it is often helpful to perform polishing prior to 3D classification. Early polishing results in an early increase in the signal-to-noise ratio of the particles, which facilitates subsequent refinements. If deemed necessary, the polishing can be repeated once a better map is available. CTF refinement, in particular optimisation of the defoci of individual particles and astigmatism for micrographs may lead to further increases in resolution. Optimisation of higher-order optical aberrations may be affected by the absence of signal at spatial frequencies in between the helical layer lines and may require better data than for globular proteins. If attempted, visual inspection of the colourful phase difference images in the log files is recommended. If a large gain in resolution is achieved in the 3D refinement after CTF refinement, executing a second CTF refinement, in particular optimising the defoci of individual particles, may further improve resolution.

Post-processing
Resolution estimation based on Fourier Shell Correlation (FSC) between the two independently refined half-maps and sharpening of the final map are performed in the post-processing step. A soft mask is generated using the ‘Mask creation’ jobtype, with the central Z length set to 20 or 30% of the box size. Post-processing for amyloids is run in the same automated manner as for globular proteins. However, although the resulting map will have some helical averaging applied in the Fourier domain, the real-space map will not obey helical symmetry. To impose the latter, the post-processed map is symmetrised using the relion_helix_toolbox
Because real-space symmetrisation leads to a further increase in the signal-to-noise ratio of the map, the estimated resolution from the post-processing tends to be somewhat under-estimated. This is preferred to over-estimating the resolution, which would result from convolution effects if one would attempt to measure resolution from half-maps that are symmetrised in real-space. To maximise the information content in the map, it is sometimes useful to run additional post-processing jobs, fixing the map sharpening B-factor to the value obtained from the first, automated run, but applying a higher-resolution ad hoc low-pass filter than the resolution estimated in the automated procedure. Caution is needed when interpreting these maps, as under-filtering may lead to high noise levels. The reported resolution of the map should be the one estimated by the automated procedure.

6.9.1 Pitfalls

Getting stuck in local minima
This probably continues to be the largest pitfall of amyloid structure determination. How to recognise and circumvent getting stuck in local minima has been described in detail previously. The possibility of ending up in local minima of refinement means that the user needs to remain highly critical of unexpected features in the map. Although this is equally true for single-particle analysis of globular proteins, artefacts with important implications for their interpretation are much more common with amyloid reconstructions. Continuing developments in the field, like better detectors or more robust optimisation techniques, will hopefully ameliorate this situation in the future.

Incorrectly calibrated pixel size
In many electron microscopes, the calibrated pixel size deviates from the correct value by several percent. Incorrectly calibrated pixel sizes will lead to deviations from the expected helical rise of 4.75 Å. Such deviations can be detected from 2D class averages with sufficient resolution to separate the β-strands (looking for a peak in the spectral signal-to-noise of those images in the model.star file), or from the optimised helical rise in 3D auto-refinements. If an incorrectly calibrated pixel size is suspected, processing may be continued with the incorrect pixel size, but the helical rise in subsequent steps will need to be adjusted accordingly. The correct pixel size can be provided at the end of processing, through the ‘Post-processing’ jobtype, to generate a final map at the correct scale. Only for cases where the resolution extends substantially beyond 2 Å would this procedure be suboptimal, as at those resolutions
higher-order effects start to become significant. Another, computationally more expensive option would be to restart processing with the correct pixel size from the beginning (although picking results could be re-used).

**Estimating cross-over distances with higher-order symmetries**

For filaments with higher than 2-fold additional symmetry, it may be difficult to estimate the cross-over distance from alternating patterns of the width of the filaments in the micrographs. An example of this is shown for the second data set described in the Results section. Therefore, when calculating the initial model in the relion_helix_inimodel2d program, it may be useful to attempt a wider range of cross-over distances than suggested by the micrographs.

**Overfitting**

In overfitting, noise artefacts in the reference map lead to systematic errors in the particle orientations. The artefacts are then enhanced in subsequent reconstructions with the incorrect orientations. Artefacts are most likely to appear at high spatial frequencies, where signal-to-noise ratios are low. In 3D auto-refinement, the iterative deterioration of the reconstruction is prevented by refining two maps independently against two halves of the data and low-pass filtering both half-maps at every iteration based on the FSC between them. Previous versions of Relion used the provided initial model as the reference for both half-sets and relied on the user to choose a suitably low-resolution initial low-pass filter to prevent overfitting. However, as described above, it is often necessary to provide relatively high-resolution initial models for successful optimisation of the helical twist and rise parameters. By providing a single, high-resolution initial model, the benefits of refining two half-sets are diminished. To address this problem, Relion-4.0 sets pairs of independently refined half-maps as initial models for the two half-sets, provided the input filename contains a half1 or half2 substring. Because procedures in the previous versions of Relion are amenable to accumulating high-resolution artefacts in the maps, users are urged to upgrade to Relion-4.0 and use only half-map references going forward.

A related problem exists with performing 3D classification with a single class as an alternative to 3D auto-refinement. In 3D classification, iterative overfitting is not prevented by separation of the data set into two halves. Instead, resolution is estimated from the power spectrum of the map itself, with higher values of the regularisation parameter $T$ leading to higher resolution estimates. Thereby, high-resolution artefacts in the map may lead to inflated resolution estimates and the further accumulation of noise. I therefore strongly advise against this use of 3D classification. If 3D auto-refinement does not give the expected resolution
of the final map, I note that the 3D auto-refinement job also responds to the regularisation parameter (through providing –tau_fudge as an additional argument). Using values higher than 1 will lead to higher resolution estimates during refinement, which in rare cases may improve convergence. However, as both half-maps are still refined independently, an estimate of the true resolution may still be obtained by post-processing.

**Z-shifted half-maps**
Independent refinement of two halves of the data may lead to a shift between the two maps in the (Z-) direction of the helical axis. This will lower the FSC between the two maps and thus hamper convergence onto a high-resolution solution. When the two half-maps are provided again as initial models for subsequent refinements (as described above), it will be difficult to escape from this situation. In such cases, one may align the two half-maps with respect to each other and replace one of the original half-maps with the aligned version before performing the next refinement. I use UCSF Chimera or ChimeraX for this alignment.

**Handedness**
Because cryo-EM reconstruction does not provide information on the absolute hand, the final map may need to be inverted. Most amyloid filaments solved to date have a left-handed twist, but filaments with right-handed twists have also been observed, including for filaments extracted from diseased tissue. At resolutions beyond 2.9 Å, the handedness may be determined directly from the map through densities for the carbonyl oxygens of the main chain. For maps at lower resolutions, handedness may be inferred from the conformation of parts of the structure that have been observed previously, or from additional experiments, like atomic force microscopy or rotary shadowing electron microscopy. If this is not possible, one may also build a model in maps of both hands and compare the corresponding FSCs between the models and the maps, but it may be safer to explicitly state that the handedness remains unclear.
References


References


Friedreich. *Friedreich: Zur amyloidfrage*


References


Appendix A

Fourier Shell Correlation curves

A.1 FSCs for chapter 2
A.2 FSCs for chapter 3
Fig. A.1 **Additional cryo-EM data on type 1 and type 2 filaments with protofilament fold A.** (a) Electron micrograph of seeded assemblies using filament preparations from MSA case 2 as the seed. Type 1 and type 2 filaments are indicated with white arrows. Scale bar, 50 nm. (b) 2D class averages of type 1 (left) and type 2 (right) filaments with two protofilaments of fold A in a box spanning 280 Å. (c) Local resolution maps for type 1 (top) and type 2 (bottom) filaments, with the legend indicating resolutions in Å. (d) Side view of the 3D reconstructions for type 1 (left) and type 2 (right) filaments, showing clear separation of β-strands along the helical axis. (e) FSC curves for type 1 filaments (left) and type 2 filaments (right) with two protofilaments of fold A between two independently refined half-maps (black), of the final cryo-EM reconstruction and refined atomic model (red), of the first half map and the atomic model refined against the first half map (blue), and of the atomic model that was refined against the first half-map against the second half-map (yellow dashed).
Fig. A.2 Additional cryo-EM data on type 1 and type 2 filaments with protofilament fold B. a 2D class averages of type 1 filaments with two protofilaments of fold B. b Local resolution map for type 1 filaments with two protofilaments of fold B with the colour map indicating resolutions in Å. c Side view of the 3D reconstructions of type 1 filaments with two protofilaments of fold B. d-f as a-c but for type 2 filaments with two protofilaments of fold B. g-i as a-c but for type 2 filaments with one protofilament of fold A and one protofilament of fold B. j-l Fourier shell correlation curves for type 1 filaments with two protofilaments of fold B, type 2 filaments with two protofilaments of fold B, and type 2 filaments with one protofilament of fold A and one protofilament of fold B. Fourier shell correlation curves are shown between two independently refined half-maps (black) of the final cryo-EM reconstruction and refined atomic model (red), of the first half map and the atomic model refined against the first half map (blue), and of the atomic model that was refined against the first half-map against the second half-map (yellow dashed).
Fig. A.3 **Additional cryo-EM data on type 3 filaments.**

a Electron micrograph of case 5. The scale bar indicates 50 nm.

b 2D class averages of type 3 filaments in a box spanning 280 Å.

c Local resolution map, with the colour map indicating resolutions in Å.

d Fourier shell correlation curves between two independently refined halfmaps (black), of the final cryo-EM reconstruction and the refined atomic model (red), of the first half map and the atomic model refined against the first half map (blue), and of the atomic model that was refined against the first half-map against the second half-map (yellow dashed).

e Side view of the 3D reconstruction, showing separation of β-strands along the helical axis.
Data collection and processing for chapter 3

<table>
<thead>
<tr>
<th>Type 1A</th>
<th>Type 2A</th>
<th>Type 1B</th>
<th>Type 2B</th>
<th>Type 2A/B</th>
<th>Type 3</th>
<th>Type 3 doublet</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 5</th>
<th>Case 5</th>
<th>Case 5</th>
<th>Case 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(PDB 7NCA)</td>
<td>(PDB 7NCG)</td>
<td>(PDB 7NCH)</td>
<td>(PDB 7NCI)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
<td>(PDB 7NCK)</td>
</tr>
</tbody>
</table>

- **Magnification**: 105 000x, 105 000x, 105 000x, 105 000x, 105 000x, 105 000x, 105 000x, 105 000x, 105 000x, 105 000x
- **Voltage (kV)**: 300, 300, 300, 300, 300, 300, 300, 300, 300, 300
- **Detector**: K2, K2, K2, K2, K2, K2, K2, K2, K2, K2
- **Electron exposure (e−/Å²)**: 32.6, 32.6, 36.7, 36.7, 36.7, 37.5, 37.0, 37.0, 37.0, 37.0
- **Defocus range (μm)**: −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8, −1.5 to −2.8
- **Pixel size (Å)**: 1.14, 1.14, 1.1, 1.1, 1.1, 1.1, 1.1, 1.1, 1.1, 1.1
- **Micrographs**: 1,294, 1,294, 2172, 2172, 2172, 2172, 2172, 2172, 2172, 2172
- **Symmetry imposed**: C2, C2, C2, C2, C1, C1, C1, C2, C2, C2
- **Initial particle images (no.)**: 2,873,646, 2,873,646, 441,592, 441,592, 441,592, 122,831, 127,003, 127,003, 127,003, 127,003
- **Final particle images (no.)**: 67,619, 82,474, 33,479, 87,092, 57,358, 69,490, 18,691, 82,474
- **Map resolution (FSC = 0.143) (Å)**: 3.47, 3.43, 3.84, 3.55, 4.23, 3.18, 3.54, 4.40
- **Map resolution range (Å)**: 3.8–11, 3.2–6.3, 3.5–10, 3.3–18, 4.0–14, 2.7–5.5, NA, NA
- **Helical twist (°)**: −0.94, −0.95, −0.86, −0.77, −0.86, −0.95, −0.95, −1.52

Fig. A.4: Statistics on structures solved for the seeded aggregation of recombinant α-synuclein using seeds from MSA. Modified from Lövestam, Schweighauser, et al., 2021
Fig. A.5 **Fourier shell correlation curves (part 1).** FSC curves are shown for two independently refined cryo-microscopy (cryo-EM) half-maps (black); for the final refined atomic model against the final cryo-EM map (red); for the atomic model refined in the first half-map against that half-map (blue); and for the refined atomic model in the first half-map against the second half-map (yellow). The resolutions where the black line drops below 0.143 and the red line drops below 0.5 are indicated. The corresponding filament type (as defined in Table 1) is shown at the top left of each graph. See Supplementary file 1 - Tables 1–25 for further details on data processing.
Fig. A.6 **Fourier shell correlation curves (part 2).** FSC curves are shown for two independently refined cryo-microscopy (cryo-EM) half-maps (black); for the final refined atomic model against the final cryo-EM map (red); for the atomic model refined in the first half-map against that half-map (blue); and for the refined atomic model in the first half-map against the second half-map (yellow). The resolutions where the black line drops below 0.143 and the red line drops below 0.5 are indicated. The corresponding filament type (as defined in Table 1) is shown at the top left of each graph. See Supplementary file 1 - Tables 1–25 for further details on data processing.