compounds such as LMT can interfere with this process. This would enable the development of a structure-based hypothesis for ligand design.

Summary

In an attempt to address the above-mentioned needs, the present inventors devised multistep computational protocols to explore the conformational flexibility of a previously reported truncated tau fragment corresponding to one of the species isolated from proteolytically stable PHF preparations (residues ^{297}Ile-–Glu^{391}) and referred to as "dGAE" (Wischik et al., 1988; Novak et al., 1991; Novak et al., 1993). dGAE assembles spontaneously in physiological conditions *in vitro* into filaments that closely resemble native PHFs and straight filaments from AD brain morphologically (Al-Hilaly et al., 2017; Al-Hilaly et al., 2020). The analyses were aimed at understanding the flexibility of dGAE, its aggregation into PHFs and the potential of small molecules to inhibit PHF assembly and to cause PHF disaggregation in vivo. In an effort to understand the mechanism of how small molecules could inhibit dGAE assembly into fibrils, the conformation space of a single Tau97 monomer (comprising dGAE and preceding residues ^{295}Asp^{296}Asn) was evaluated with atomistic molecular simulation. The inventors hypothesized that LMT could complex to dGAE monomers and prevent fibril formation. To test this hypothesis, representative protein structures were sampled from the conformational ensemble of apo Tau97. From this, the inventors were able to identify a unique cryptic pocket which would bind to LMT and form a stabilised complex. From the conformation of the complex a pharmacophore model was built and a series of *de novo* designed compounds were synthesised in order to assess whether stabilising dGAE oligomers would result in inhibition of Tau aggregation. Compounds were assessed in a cell-based tau aggregation assay. More than half of the assayed molecules showed sub micro molar activity. From this, the inventors were able to rationalise the structure-activity relationship from this series of Tau assembly inhibitors.

Thus, the present disclosure is based, in part, on the discovery of a ligand binding pocket in the Tau protein structure, wherein binding of a ligand in the pocket stabilises the Tau protein in a conformation that is not prone to aggregation into paired helical filaments (PHFs). Interactions between the ligand and any of residues Leu315, Ser341, Glu342, Lys343, Phe346, Lys347, Val350, Ser352, Ile354, Lys369, Ile371, Glu372, Phe378 and Thr373 of Tau were found to be particularly important in stabilising the ligand bound conformation.

The formation of PHF was then analysed through immunostaining dot blots and ELISA assays. These showed that in vitro assembly of PHFs is associated with loss of immunoreactivity with antibodies directed against the PHF core. Furthermore, loss of

immunoreactivity could be prevented by including LMT during assembly. Starting from the stabilised conformation of dGAE73 (residues 306–378), a molecular simulation was performed to examine assembly onto a preformed PHF. This simulation supported findings from the dot blots for the assembly of PHFs from stabilized dGAE oligomers. The simulation identified 3 key stages in assembly of PHFs. The first is the electrostatic attraction of oligomers to PHFs and the anchoring of a hairpin turn. This is followed by a proline trans-cis-trans switch which allows for the final step of the formation of stabilising cross-β sheets through hydrophobic zippering of the N- and C-terminal tails of the protein structure. Taken together, the structure-activity results and the dot blots provide strong evidence for the binding of small molecules to the cryptic pocket proposed in our dGAE stabilised conformation, which finds uses in a structure-based approach to design further inhibitors of tau assembly.

Thus the present disclosure is based, in other examples, on the discovery of a folding process through which the Tau protein in the compact folded state that includes the binding pocket aggregates with a paired helical filament, where binding of a ligand that inhibits the formation of intermediates in the folding process may stabilise the Tau protein in a conformation that is not prone to aggregation into PHFs. A series of conformational changes from a compact folded state to an aggregated state were found to be particularly important. These conformational changes are such that:
(i) residues Val337-Gln355 form a hairpin loop that moves to align with alternating positively charged and negatively charged sidechain stacks in the hairpin loop of the PHF;
(ii) residue Pro332 switches between a trans and a cis configuration;
(iii) residues 355-378 and 306-318 move to form stabilising cross-β sheets with corresponding residues of the PHF through hydrophobic zippering.

In examples, residue Pro332 switching between a trans and a cis configuration causes residues His329, His330 and Lys331 to move close enough to PHF to establish interactions with the partner residues in the next layer of the PHF stack through hydrophobic stacking and a strong hydrogen bond between the side chains of His330 in a preformed layer of the PHF stack and Thr361 of the newly formed dGAE layer on the PHF stack.

In examples, step (iii) comprises the following steps: (iii)(1) residues 355-360 move to establish a zipper with the corresponding anti-parallel residues of the PHF; (iii)(2) residues 361-367 move to establish a zipper with the corresponding anti-parallel residues of the PHF; and (iii)(3) residues 306-318 and 368-378 move to form a cross-beta sheet conformation with corresponding residues of the PHF.

According to a first aspect, there is provided a method for selecting or designing a compound for modulating the aggregation of a Tau protein or a truncated form thereof, the method comprising using computer-implemented molecular modelling means to:

compare the three-dimensional structure of a candidate compound with a three-dimensional structure of at least a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or a variant or derivative thereof that is structurally equivalent to said part; and

determine whether the candidate compound is able to simultaneously form non-covalent interactions with two or more of Leu315, Ser341, Glu342, Lys343, Phe346, Lys347, Val350, Ser352, Ile354, Lys369, Ile371, Glu372, Phe378 and Thr373, of SEQ ID NO:1, or equivalent amino acids in a variant or derivative, wherein a candidate compound that is able to form said interactions is predicted to modulate the aggregation of the Tau protein or truncated form thereof.

According to a related example, there is provided a method for selecting or designing a compound for modulating the aggregation of a Tau protein or a truncated form thereof, the method comprising using computer-implemented molecular modelling means to:

compare the three-dimensional structure of a candidate compound with a three-dimensional structure of at least a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 as defined in Table 1, or a variant or derivative thereof that is structurally equivalent to said part; and

determine whether the candidate compound is able to form non-covalent interactions with the part of the Tau protein having the structure defined in Table 1 which stabilise said structure, wherein a candidate compound that is able to form said interactions is predicted to modulate the aggregation of the Tau protein or truncated form thereof.

As used herein, the step of comparing the three-dimensional structure of a candidate compound with a three-dimensional structure of at least a part of the Tau protein comprises providing the three-dimensional structure of the candidate compound, providing the three-dimensional structure of the part of the Tau protein, and obtaining a molecular model that includes both three-dimensional structures.

In examples, the non-covalent interactions comprise interactions with one or more, optionally with two or more of: Leu315, Ser341, Glu342, Lys343, Phe346, Lys347, Val350, Ser352, Lys369, Ile371, Glu372, and Thr373, of SEQ ID NO:1, or equivalent amino acids in a variant or derivative. Non-covalent interactions with the above-mentioned residues of the Tau protein were found to stabilise a compact conformation of the Tau protein (an example of which is provided by the coordinates defined in Table 1), reducing the propensity of the Tau protein to aggregate.

In examples, the non-covalent interactions comprise at least a non-covalent interaction with Lys347. Indeed, a potent inhibitor of Tau aggregation, compound 16, was identified by the inventors as binding the Tau protein by establishing an interaction with Lys347, thereby stabilising a conformation of the protein that has reduced propensity to aggregate. Further potent inhibitors of Tau aggregation were also found to interact with Lys347, including LMT, compound 17, compound 12, compound 1, compound 7, compound 5, compound 14, compound 2, compound 8, compound 13 and compound 18.

In examples, the non-covalent interactions comprise at least a non-covalent interaction with Lys343. Indeed, a potent inhibitor of Tau aggregation, compound 11, was identified by the inventors as binding the Tau protein by establishing an interaction with Lys343, thereby stabilising a conformation of the protein that has reduced propensity to aggregate. Further potent inhibitors of Tau aggregation were also found to interact with Lys343, including LMT, compound 17, compound 12, compound 3, compound 14, compound 10, compound 13 and compound 18.

In examples, the non-covalent interactions comprise at least a non-covalent interaction with Thr373. Indeed, a potent inhibitor of Tau aggregation, compound 9, was identified by the inventors as binding the Tau protein by establishing an interaction with Thr373, thereby stabilising a conformation of the protein that has reduced propensity to aggregate. Further potent inhibitors of Tau aggregation were also found to interact with Thr373, including LMT, compound 3, compound 4, and compound 15.

In examples, the non-covalent interactions comprise an interaction with two or more of Lys343, Leu315, Lys347, Glu372, and Thr373, of SEQ ID NO:1.

In examples, the non-covalent interactions comprise an interaction with one or more of Leu315, Val350, and Ile371 of SEQ ID NO:1. Interactions with these residues are preferably hydrophobic interactions, involving the side chain of the residue. Indeed, multiple potent inhibitors of Tau aggregations, such as compounds 16, 9 and 11, were identified by the inventors as binding the Tau protein by establishing an interaction with these residues.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys347, Lys343, Ser352 and Thr373 of SEQ ID NO: 1. A known inhibitor of Tau aggregation, LMT was identified by the inventors as binding the Tau protein in a cryptic binding pocket, where it establishes interactions with the above-mentioned

residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys347, Lys343, Ser341, Leu315, Ile371, Glu372 and Phe346 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 17, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys347, Lys343, Ile371, and Glu372 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 12, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys347, Ile371, Glu342, Phe346 and Glu372 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 1, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Leu315, Lys347, Ile371, Glu342, Phe346 and Glu372 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 7, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with at least Lys343 and Thr373 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 3, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Glu372, Lys369, and Thr373 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 7, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys343, Lys347, Ile371, Glu342, Phe346 and Glu372 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 5, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Leu315, Lys347, and Lys343 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 14, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys343, Lys347, and Glu372 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 2, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys347, Glu342, Phe346 and Glu372 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 8, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Thr373, Lys369, and Lys343 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 10, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Leu315, Ile371, and Lys343 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 6, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Ile371, Glu372, Lys343, Phe346, and Lys347 of SEQ ID NO: 1.

A potent inhibitor of Tau aggregation, compound 13, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Glu372 and Thr373 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 15, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

In examples, the non-covalent interactions comprise an interaction with two or more (such as all of) Lys343, Phe346, and Lys347 of SEQ ID NO: 1. A potent inhibitor of Tau aggregation, compound 18, was identified by the inventors as binding the Tau protein by establishing interactions with the above-mentioned residues, thereby stabilising a conformation of the protein that has reduced propensity to aggregate.

The methods of the first and second example may comprise determining whether the candidate compound is able to simultaneously form non-covalent molecular interactions with Lys343 and Glu372 of SEQ ID NO:1.

Multiple compounds found to have a strong effect in modulating Tau aggregation were predicted to form stabilising interactions with Lys343 and Glu372, stabilising the Tau protein in a conformation that include the binding pocket exposing these residues.

In examples, a non-covalent molecular interaction with Lys343 comprises a cation-pi interaction and/or a hydrogen bond and/or a pi-H interaction. In some examples, a cation-pi interaction with Lys343 is between an aromatic ring of the candidate compound and the sidechain amino group of Lys343. In some examples, a hydrogen bond with Lys343 is between an acceptor group of the candidate compound and the backbone amino group of Lys343. In some examples, a hydrogen bond with Lys343 is between a donor group of the candidate compound and the backbone carbonyl of Lys343. In some examples, a pi-H interaction with Lys343 is between an aromatic ring of the candidate compound and the sidechain Cβ of Lys343. In some examples, a pi-H interaction with Lys343 is between an aromatic ring of the candidate compound and the backbone amino group of Lys343. In some examples, a pi-H interaction with Lys343 is between an aromatic ring of the candidate compound and the sidechain Cε of Lys343.

In examples, a non-covalent interaction with Ser352 comprises a pi-H interaction. In some such examples, a pi-H interaction with Ser352 is between an aromatic ring of the candidate compound and the backbone carbonyl of Ser352.

In examples, the non-covalent interaction with any of Leu315, Ser341, Phe346, Lys347, Ile371, Glu372, and Thr373, of SEQ ID NO:1 is a hydrogen bond.

In examples, the non-covalent interaction with Glu372 is a hydrogen bond. In some such examples, a hydrogen bond with Glu372 is between a H donor group of the candidate compound and the sidechain carboxyl (OE1) of Glu372. In some examples, a hydrogen bond with Glu372 is between an acceptor group of the candidate compound and the backbone amino group of Glu372. In some examples, a hydrogen bond with Glu372 is between an acceptor group of the candidate compound and the sidechain amino group of Glu372.

In examples, the non-covalent interaction with Lys347 is a hydrogen bond. In examples, the hydrogen bond with Lys347 is between an acceptor group of the candidate compound and the backbone amino group of Lys347.

In examples, the non-covalent interaction with Thr373 is a hydrogen bond. In examples, a hydrogen bond with Thr373 is between a donor group of the candidate compound and the hydroxyl group of the side-chain of Thr373.

In examples, the non-covalent interaction with Leu315 is a hydrogen bond. In examples, a hydrogen bond with Leu315 is between a donor group of the candidate compound and the backbone carbonyl of Leu315.

In examples, the candidate compound is able to form a non-covalent interaction with Phe378. In examples, the non-covalent interaction is a pi-interaction.

In examples, the non-covalent interaction with Ile371 is a hydrogen bond. In examples, a hydrogen bond with Ile371 is between an acceptor group of the candidate compound and the backbone Cα of Ile371.

In examples, the non-covalent interaction with Ser341 is a hydrogen bond. In examples, a hydrogen bond with Ser341 is between an acceptor group of the candidate compound and the sidechain Cβ of Ser341.

In examples, the non-covalent interaction with Phe346 is a hydrogen bond. In examples, a hydrogen bond with Phe346 is between an acceptor group of the candidate compound and the backbone Cα of Phe346.

In examples, the non-covalent interaction with Glu342 is a hydrogen bond. In some examples, a hydrogen bond with Glu342 is between an acceptor group of the candidate compound and the backbone amino group of Glu342.

In examples, the non-covalent interaction with Lys369 is a hydrogen bond. In examples, a hydrogen bond with Lys369 is between an acceptor group of the candidate compound and the backbone Cε of Lys369. In examples, a hydrogen bond with Lys369 is between an acceptor group of the candidate compound and the backbone amino group of Lys369.

In examples, the interactions with any of Val350, Leu315, Ile354, Ile371, Phe378 and Phe346 are hydrophobic interactions.

In examples, the compound is for inhibiting the aggregation of a Tau protein or a truncated form thereof into paired helical filaments.

The compound may be a small molecule, a peptide, a polypeptide or a combination thereof. Preferably, the compound is a small molecule.

In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 comprises a binding pocket that is capped by Phe378 and Phe346, exposes the hydrophobic side chains of residues Val350, Leu315, Ile354 and/or Ile371, and contains residues Leu315, Lys343, Lys347, Glu372, and/or Thr373 capable of forming hydrogen bonds to a molecule within the binding pocket.

In examples, the candidate compound is able to simultaneously form non-covalent molecular interactions with one or more of residues Leu315, Ser341, Glu342, Lys343, Phe346, Lys347, Ser352, Lys369, Ile371, Glu372, and Thr373, of SEQ ID NO:1, or equivalent amino acids in a variant or derivative, without exposing hydrophilic groups at a distance below 4 Å from the hydrophobic side chains of residues Val350, Leu315, Ile354 and Ile371.

In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 comprises a hairpin loop between residues Val337 and Gly355 of SEQ ID NO:1.

The present inventors have identified that a compact folded state that is able to interact with inhibitors of Tau aggregation such as LMT, and is stabilised by this interaction, comprises a hair pin loop between residues Val337 and Gly355 of SEQ ID NO:1.

In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 comprises the sequence Pro364-Gly367 located between a loop formed by the sequence Tyr219-Lys331 and the sequence Pro332-Gly335.
In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 is stabilised at least in part by hydrogen bonds between Glu342 and Val318 and/or Thr319, between Gly367 and Lys340, between Asn368 and Lys340, between Lys369 and Glu372, between Lys370 and Asp358, between Ile371 and Ser316, between Glu372 and Ser356, between Thr373 and Gln351, Ser316, between His374 and Gln351, and between Lys375 and Gln351.

The three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 may be that represented by the structure co-ordinates in Table 1, or a structure modelled on these coordinates.

The three-dimensional structure of the part of the Tau protein may be obtainable by performing molecular dynamics simulations of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 in the presence of LMT to obtain one or more complex conformations and selecting a three-dimensional structure of the part of the Tau protein by applying a stability criterion and a binding affinity criterion to the one or more complex conformations.

Advantageously, the stability criterion may apply to the distance between complex conformations in consecutive frames of the molecular dynamics simulation after a predetermined amount of time, and/or the binding affinity criterion may apply to the value of a placement scoring function <-80 kcal/mol or a scoring function <-8.5 as determined by the GB IV scoring function within the CCG MOE docking software.

In examples, the stability criterion applies to the root-mean-square deviation (RMSD) of atomic positions, for atoms in the backbone of the part of protein in the complexes, between consecutive frames of the molecular dynamics simulation.

In examples, the stability criterion applies to the root-mean-square deviation (RMSD) of atomic positions, for atoms in LMT. In examples, a stability criterion may be defined as a threshold on the average RMSD of atomic positions, for atoms in the backbone of the part of

protein in the complexes, between consecutive frames of the molecular dynamics simulation, where the average is calculated over a predetermined set of frames of a MD simulation. For example, the predetermined set of frames may be defined as the last 10% of frames, the last 5% of frames, or the last 1% of frames in a MD simulation. The predetermined set of frames may be defined as the last 30k frames, the least 25k frames, the last 20k frames, the last 15k frames, or the last 10k frames. The threshold on the average RMSD may be chosen as about 0.5 Å, about 0.6 Å, about 0.7 Å, about 0.8 Å or about about 0.9 Å.

In particular, the present inventors have found that a stable LMT-bound conformation could be obtained by running molecular dynamics simulation of a part of the Tau protein in the presence of LMT, and excluding those complexes where after a predetermined amount of simulation time, the LMT was no longer tightly bound (as indicated by the docking score or ligand RMSD fluctuations) and/or the RMSD fluctuations between consecutive frames of the simulation indicated that the conformation was not stable.

In examples, the predetermined amount of time is at least 50ns, at least 100ns, or about 100ns.

In examples, the methods described herein further comprises designing a pharmacophore model, wherein the pharmacophore includes features representative of non-covalent molecular interactions with two or more of: Leu315, Ser341, Glu342, Lys343, Phe346, Lys347, Val350, Ser352, Ile354, Lys369, Ile371, Glu372, Phe378 and Thr373, of SEQ ID NO:1.

In examples, designing a pharmacophore model comprises computing the interaction energy between the three-dimensional structure of each candidate compound in a library of candidate compounds and the three-dimensional structure of at least a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or a variant or derivative thereof that is structurally equivalent to said region, selecting a subset of the library that has an interaction energy below a threshold, and deriving a pharmacophore model based on the selected subset.

The part of the Tau protein may comprise amino acids 306-378 of SEQ ID NO:1. The part of the Tau protein may comprise amino acids 297-391 of SEQ ID NO:1. The part of the Tau protein may comprise amino acids 295-391 of SEQ ID NO:1. The part of the Tau protein may consist of amino acids 297-391 of SEQ ID NO:1. The part of the Tau protein may consist of amino acids 295-391 of SEQ ID NO:1. The part of the Tau protein may consist of amino acids 306-378 of SEQ ID NO:1.In examples, the part of the Tau protein comprises

amino acids 306-378 of SEQ ID NO:1 (also referred to herein as dGAE73, SEQ ID NO: 4) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, the part of the Tau protein comprises amino acids 297-391 of SEQ ID NO:1 (also referred to herein as dGAE, SEQ ID NO: 3) or a variant or derivative thereof that is structurally equivalent to said part. In examples, the part of the Tau protein comprises amino acids 295-391 of SEQ ID NO:1 (also referred to herein as Tau97, SEQ ID NO: 5) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, the part of the Tau protein comprises the full sequence of the human Tau isoform 2N4R (SEQ ID NO: 1) or a homolog or variant thereof. In example, a variant has at least 90% amino acid identity with the sequence of SEQ ID NO: 1, or the portion thereof that is represented in the three-dimensional structure.

The methods may further comprise repeating the steps of comparing and determining with a further candidate compound that differs from the previous candidate compound in at least one substituent.

In examples, one or more candidate compounds may be modelled such that their properties can be compared to thereby define the features that advantageously stabilise the ligand-bound conformation of the part of the Tau protein. Comparing the properties of multiple candidate compounds may comprise comparing, using computational molecular modelling means, the characteristics of the complexes comprising the part of the Tau protein and each of the candidate compounds in terms of e.g. stability, binding affinity, etc. Comparing the properties of multiple candidate compounds may comprise performing one or more experiments to quantify the Tau aggregation modulation associated with each candidate compound.

In examples, such comparisons may be performed sequentially to progressively optimise a candidate compound, where the information obtained by comparing a candidate compound to a previously obtained candidate compound is used to design a new candidate compound in a subsequent iteration. Instead or in addition to this, such comparisons may be performed in parallel where multiple candidate compounds are simultaneously compared to derive insights from common properties of subsets of candidate compounds tested. Such insights can be used to design one or more further candidate compounds.
Without wishing to be bound by theory, it is believed that such a process may help to define the features of a candidate compound that optimally modulates Tau aggregation by stabilising a compact folded conformation of the Tau protein.

Comparing the three-dimensional structure of a candidate compound with a three-dimensional structure of at least a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or a variant or derivative thereof that is structurally equivalent to said region may comprise computing the interaction energy between the candidate compound and the part of the Tau protein represented in the three-dimensional structure.

According to a further aspect, there is provided a computer-implemented method for evaluating the ability of a candidate compound to bind to a binding pocket of a Tau protein, the method comprising the steps of:

(a) receiving the three-dimensional structure coordinates of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or equivalent amino acids in a variant or derivative thereof that is structurally equivalent to said part, wherein the three-dimensional structure of amino acids 315-378 of SEQ ID NO: 1 comprises the binding pocket of the Tau protein;

(b) performing a fitting operation between a candidate compound and the binding pocket; and

(c) analysing the results of the fitting operation to determine whether the candidate compound is able to bind to the binding pocket.

In examples, step (c) comprises determine whether the candidate compound is able fit at least in part within the binding pocket and form non-covalent molecular interactions with one or more of Lys343, and at least one of Leu315, Ser341, Glu342, Lys343, Phe346, Lys347, Val350, Ser352, Ile354, Lys369, Ile371, Glu372, Phe378 and Thr373, of SEQ ID NO:1, or equivalent amino acids in a variant or derivative.

In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 is that represented by the structure co-ordinates shown in Table 1, or a structure modelled on these coordinates.

In examples, the compound is a small molecule, a peptide, a polypeptide or a combination thereof.

In examples, a candidate compound is considered able to bind to the binding pocket if it is able to establish non-covalent interactions with amino acids in the binding pocket. In examples, the non-covalent interactions are selected from: hydrophobic interactions, hydrogen bonds, pi-cation interactions, and pi-stack interactions.

In examples, a candidate compound that is considered able to bind to the binding pocket is predicted to inhibit the aggregation of the Tau protein or a truncated form thereof. In

examples, a candidate compound that is considered able to bind to the binding pocket is predicted to inhibit the aggregation of the Tau protein or a truncated form thereof into paired helical filaments.

In examples, the method further comprises evaluating the ability of a further candidate compound to bind the Tau protein at a different site from the binding site of the first candidate compound.

In examples, evaluating the ability of a further candidate compound to bind the Tau protein at a different site from the binding site of the first candidate compound comprises computing the interaction energy between the further candidate compound and the three-dimensional structure of the part of the Tau protein and determining whether the interaction between the further candidate compound and the Tau protein further stabilises the conformation of the complex between the first candidate compound and the Tau protein.
In examples, the first candidate compound is a small molecule or a peptide, and the further candidate compound is a peptide or a polypeptide, such as an antibody or fragment thereof.

In examples, the candidate compound is able to bind to the pocket if it is able to simultaneously form non-covalent molecular interactions with two or more of Leu315, Ser341, Glu342, Lys343, Phe346, Lys347, Val350, Ser352, Ile354, Lys369, Ile371, Glu372, Phe378 and Thr373, of SEQ ID NO:1.

In examples, the candidate compound is able to bind to the pocket if it is able to simultaneously form non-covalent molecular interactions with Lys343 and at least one of Leu315, Lys347, Glu372 and Thr373.

In some cases, the candidate compound is able to bind to the pocket if it is able to simultaneously form non-covalent molecular interactions with Lys343 and Glu372 of SEQ ID NO:1.

In examples, the binding pocket is capped by Phe378 and Phe346, exposes the hydrophobic side chains of residues Val350, Leu315, Ile354 and/or Ile371, and contains residues Leu315, Lys343, Lys347, Glu372, and/or Thr373 capable of forming hydrogen bonds to a molecule within the binding pocket.

In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 comprises a hairpin loop between residues Val337 and Gly355 of SEQ ID NO:1.

In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 comprises the sequence Pro364-Gly367 located between a loop formed by the sequence Tyr219-Lys331 and the sequence Pro332-Gly335.

In examples, the three-dimensional structure of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 is stabilised at least in part by hydrogen bonds between Glu342 and Val318, Thr319, between Gly367 and Lys340, between Asn368 and Lys340, between Lys369 and Glu372, between Lys370 and Asp358, between Ile371 and Ser316, between Glu372 and Ser356, between Thr373 and Gln351, Ser316, between His374 and Gln351, between Lys375 and Gln351.

In examples, the three-dimensional structure coordinates of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or equivalent amino acids in a variant or derivative thereof that is structurally equivalent to said part have been obtained by performing molecular dynamics simulations of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 in the presence of LMT to obtain one or more complex conformations and selecting a complex conformation using a stability criterion and a binding affinity criterion.

In examples, step (a) receiving the three-dimensional structure coordinates of a part of the Tau protein comprises obtaining the three-dimensional structure coordinates of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or equivalent amino acids in a variant or derivative thereof that is structurally equivalent to said part by:

(a) performing molecular dynamics simulations of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 in the presence of LMT to obtain one or more complex conformations that differ in their three-dimensional conformations; and

(b) selecting a complex conformation using a stability criterion and a binding affinity criterion, wherein the three-dimensional structure coordinates of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 are defined as the three-dimensional structure coordinates of the part of the Tau protein in the selected complex conformation;

optionally wherein the stability criterion applies to the distance between conformations in consecutive frames of the molecular dynamics simulation after a predetermined amount of time, and/or wherein the binding affinity criterion applies to the value of a docking score.

In examples, selecting a complex conformation using a stability criterion comprises:

computing the root-mean-square deviation (RMSD) of atomic positions, for atoms in the backbone of the part of protein in the complex conformations, between consecutive frames of the molecular dynamics simulation over a predetermined amount of time; and

selecting the complex conformation(s) that has/have a RMSD below a predetermined threshold and/or that have the lowest RMSD amongst the one or more complex conformations.

In examples, the predetermined amount of time is at least 50ns, at least 100ns, or about 100ns.

In examples, selecting a complex conformation using a binding affinity criterion comprises computing the interaction energy between the candidate compound and the part of the Tau protein represented in the three-dimensional structure.

In examples, performing molecular dynamics simulations of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 in the presence of LMT comprises obtaining the three-dimensional structure coordinates of a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or equivalent amino acids in a variant or derivative thereof that is structurally equivalent to said part, wherein the three-dimensional structure coordinates correspond to a conformation of the Tau protein as part of a PHF stack. In examples, the three-dimensional structure coordinates correspond to those of PDB ID: 5O3L or a structure modelled on this structure.

The part of the Tau protein may comprise amino acids 306-378 of SEQ ID NO:1. The part of the Tau protein may comprise amino acids 297-391 of SEQ ID NO:1. The part of the Tau protein may comprise amino acids 295-391 of SEQ ID NO:1. The part of the Tau protein may consist of amino acids 297-391 of SEQ ID NO:1. The part of the Tau protein may consist of amino acids 295-391 of SEQ ID NO:1. The part of the Tau protein may consist of amino acids 306-378 of SEQ ID NO:1.

In examples, the part of the Tau protein comprises amino acids 306-378 of SEQ ID NO:1 (also referred to herein as dGAE73, SEQ ID NO: 4) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, the part of the Tau protein comprises amino acids 297-391 of SEQ ID NO:1 (also referred to herein as dGAE, SEQ ID NO: 3) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, the part of the Tau protein comprises amino acids 295-391 of SEQ ID NO:1 (also referred to herein as Tau97, SEQ ID NO: 5) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, the part of the Tau protein comprises the full sequence of the human Tau isoform 2N4R (SEQ ID NO: 1) or a homolog or variant thereof. In example, a variant has at least 90% amino acid identity with the sequence of SEQ ID NO: 1, or the portion thereof that is represented in the three-dimensional structure.

In examples, the three-dimensional structure coordinates of the part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or equivalent amino acids in a variant or derivative thereof that is structurally equivalent to said part correspond to a conformation that has the following structural characteristics:

- hydrogen bonds between Glu342 and Val318 and/or Thr319; optionally wherein the hydrogen bonds are between the Glu342 carboxylic acid and the backbone NH of Val318 and the sidechain OH of Thr319;
- one or more hydrogen bonds between one or more of residues Lys369-Thr377 and one or more of residue Ser341-Gln351; optionally wherein the one or more bonds comprise:
    (i) a bond between Gln351 and Thr373, preferably wherein the bond is between the backbone carbonyl of Gln351 and the hydroxyl sidechain of Thr373;
    (ii) a bond between Gln351 and His374, preferably wherein the bond is between the sidechain carbonyl of Gln351 and the sidechain amine of His374;
    (iii) a bond between Gln351 and Lys375, preferably wherein the bond is between the sidechain carbonyl of Gln351 and the backbone amine of Lys375;
    (iv) a bond between Arg349 and Thr377, preferably wherein the bond is between a sidechain amine of Arg349 and the hydroxyl sidechain of Thr377, between a sidechain amine of Arg349 and the hydroxyl backbone of Thr377, and/or between the carbonyl backbone of Arg349 and the hydroxyl sidechain of Thr377;
    (v) a bond between Glu372 and Ser356, preferably wherein the bond is between the carboxylic acid side chain of Glu372 and the backbone NH of Ser356, or between the carboxylic acid side chain of Glu372 and the OH-sidechain of Ser356; and/or
    (vi) a bond between Glu372 and Lys369, preferably wherein the bond is between the carboxylic acid side chain of Glu372 and the NH sidechain of Lys369;
- no beta sheets;
- a hairpin loop comprising residues Val337-Gly355;

- the PGGG sequence formed by residues Pro364-Gly367 is within a distance of 13A of the PGGG sequence Pro332-Gly335 and/or within a distance of 2A of a loop formed by the sequence Thr319-Lys331;
- the PGGG sequence formed by residues Pro364-Gly367 is located between the PGGG sequence Pro332-Gly335 and a loop formed by the sequence Thr319-Lys331;
- residues Lys369-Thr377 are within a distance of 6A of residues Asp314-Ser316, optionally wherein the distance between the Ser316 beta-carbon and Thr373 backbone carbonyl is between 2.5 Å and 5.0 Å;
- residues Gly355-Gly367 and Asn368-Arg379 are within 2A hydrogen bonding distance;
- Glu338 is folded towards Val363, optionally wherein the distance (RMSD) between the carbonyl oxygen side chain of Glu338 and the backbone amine nitrogen of Val363 NH during the final 10ns of a 50ns simulation is between 2 Å and 4 Å, or below 5 Å;
- the total water accessible surface area calculated for the part of the protein is at least 20% lower than the corresponding values calculated for a conformation as provided by the three-dimensional coordinates with PDB identifier 5O3L; and/or
- the polar and/or hydrophobic accessible surface area(s) calculated for the part of the protein is/are at least 20% lower than the corresponding values calculated for a conformation as provided by the three-dimensional coordinates with PDB identifier 5O3L.

In examples, the three-dimensional structure coordinates of the part of the Tau protein is such that the distance (RMSD) between the backbone carbonyl oxygen of Gln351 and the sidechain hydroxyl oxygen of Thr373 during the final 10ns of a 50ns simulation is between 2 Å and 4 Å, or below 5 Å.

In examples, the three-dimensional structure coordinates of the part of the Tau protein is such that the distance (RMSD) between the sidechain carbonyl oxygen of Gln351 and the sidechain amine nitrogen of His374 during the final 10ns of a 50ns simulation is between 2 Å and 4 Å, or below 5 Å.

In examples, the three-dimensional structure coordinates of the part of the Tau protein is such that the distance (RMSD) between the sidechain carbonyl oxygen of Gln351 and the backbone amine nitrogen of Lys375 during the final 10ns of a 50ns simulation is between 2.5 Å and 5 Å, or below 5 Å.

In examples, the three-dimensional structure coordinates of the part of the Tau protein is such that the distance (RMSD) between the backbone carbonyl oxygen of Arg349 and the hydroxyl sidechain oxygen of Thr377during the final 10ns of a 50ns simulation is between 2.5 Å and 4 Å, or below 5 Å.

In examples, the three-dimensional structure coordinates of the part of the Tau protein is such that the distance (RMSD) between the carboxylic acid sidechain oxygen of Glu372 and the backbone NH of Ser356 during the final 10ns of a 50ns simulation is between 2 Å and 4 Å, or below 5 Å.

According to a further aspect, there is provided a method for selecting or designing a compound for modulating the aggregation of a Tau protein or a truncated form thereof, the method comprising:

using a three-dimensional structural model of at least a part of the Tau protein comprising amino acids 315-378 of SEQ ID NO:1 or a variant or derivative thereof that is structurally equivalent to said part, wherein the model is an intermediate in the aggregation process of the part of the Tau protein with a paired helical filament (PHF), wherein the model is generated by simulating the conformational changes of the part of the Tau protein from a compact folded state to a an aggregated state such that:

(i) residues Val337-Gln355 form a hairpin loop that moves to align with alternating positively charged and negatively charged sidechain stacks in the hairpin loop of the PHF;

(ii) residue Pro332 switches between a trans and a cis configuration;

(iii) residues 355-378 and 306-318 move to form stabilising cross-β sheets with corresponding residues of the PHF through hydrophobic zippering; and

generating a model of a complex of the compound and the intermediate.

In examples, residue Pro332 switching between a trans and a cis configuration causes residues His329, His330 and Lys331 to move close enough to PHF to establish interactions with the partner residues in the next layer of the PHF stack through hydrophobic stacking and a strong hydrogen bond between the side chains of His330 in a preformed layer of the PHF stack and Thr361 of the newly formed dGAE layer on the PHF stack.

In examples, step (iii) comprises the following steps: (iii)(1) residues 355-360 move to establish a zipper with the corresponding anti-parallel residues of the PHF; (iii)(2) residues 361-367 move to establish a zipper with the corresponding anti-parallel residues of the PHF; and (iii)(3) residues 306-318 and 368-378 move to form a cross-beta sheet conformation with corresponding residues of the PHF.

In examples, residues 355-360 move to establish a zipper with the corresponding anti-parallel residues of the PHF from the C- to N- direction.

In examples, residues 361-367 move to establish a zipper with the corresponding anti-parallel residues of the PHF in the N- to C- direction.

In examples, the cross-beta sheet conformation is formed simultaneously starting from Phe378 and closing the residues from C- to N-terminal, and joining residues 318-306 of the N-terminal in the C- to N- direction.

In examples, the compound is a small molecule, a peptide, a polypeptide or a combination thereof. In examples, the polypeptide is an antibody or a fragment thereof.
In examples, the method further comprises selecting or designing a further compound for modulating the aggregation of a Tau protein or a truncated form thereof by generating a model of a complex of the first compound, the further compound and the intermediate.
In examples, the further compound is a small molecule, a peptide, a polypeptide or a combination thereof. In examples, the polypeptide is an antibody or a fragment thereof.

The compact folded state may have any of the structural characteristics defined in relation to the preceding example.

The compact folded state may have the structure co-ordinates shown in Table 1, or a structure modelled on these coordinates.

Generating a model of a complex of the compound and the intermediate may comprise identifying the compound as binding to the intermediate and preventing the occurrence of any of steps (i) to (v).

Generating a model of a complex of the compound and the intermediate may comprise identifying the compound as binding to the intermediate and preventing the occurrence of any of steps (i) to (iii).

In examples, simulating the conformational changes of the part of the Tau protein from a compact folded state to an aggregated state comprises performing a molecular dynamics simulation of the assembly of a Tau protein or a part thereof with a PHF stack. In examples, the PHF stack comprises at least 2, at least 4, at least 6, at least 8 or at least 10 monomers. In examples, the PHF stack comprises the structure defined in PDB ID: 5O3L, or a structure modelled from the structure in PDB ID: 5O3L.

In examples, the part of the Tau protein comprises amino acids 306-378 of SEQ ID NO:1 (also referred to herein as dGAE73, SEQ ID NO: 4) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, the part of the Tau protein comprises amino acids 297-391 of SEQ ID NO:1 (also referred to herein as dGAE, SEQ ID NO: 3) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, the part of the Tau protein comprises amino acids 295-391 of SEQ ID NO:1 (also referred to herein as Tau97, SEQ ID NO: 5) or a variant or derivative thereof that is structurally equivalent to said part.

In examples, simulating the conformational changes of the part of the Tau protein from a compact folded state to an aggregated state comprises performing a first set of molecular dynamics simulation of the assembly of a Tau protein or a part thereof with a PHF stack, in which the anchoring of the part of the Tau protein to the PHF stack is determined by starting the molecular dynamics simulations in the set with different relative orientations of the part of the Tau protein and the PHF stack.

In examples, the molecular dynamics simulation in the set are each started with a random orientation of the part of the Tau protein. Preferably, the molecular dynamics simulation in the set are each started with the part of the Tau protein placed beyond hydrogen-bonding distance (such as e.g. at least 4Å away) of the PHF stack.

The present inventors have found that such molecular dynamics simulation enabled the identification of an intermediate in which the part of the Tau protein has anchored itself on the PHF stack. Such an intermediate may correspond to one where step (i) has occurred. In examples, simulating the conformational changes of the part of the Tau protein from a compact folded state to an aggregated state comprises performing a nudged elastic band molecular dynamics simulation starting from an intermediate in which step (i) has occurred.

According to a further aspect, there is provided a computing system comprising a processor and a memory storing machine-readable instructions that, when executed by the processor, cause the processor to implement the method of any example described above.

According to a further example, there is provided a computer readable storage medium (or a plurality of computer readable storage media) storing instructions that, when executed by

one or more processors, cause the processor(s) to implement the method of any example described above.

According to a further example, there is provided a computer software product comprising instructions that, when executed by one or more processors, cause the processor(s) to implement the method of any example described above.

Figures

Figure 1. Flowchart illustrating a method for identifying, selecting and/or designing a compound for modulating Tau aggregation according to the present disclosure.

Figure 2. Flowchart illustrating a method for identifying, selecting and/or designing a compound for modulating Tau aggregation according to the present disclosure.

Figure 3. Schematic illustration of the steps used to analyse the structure of Tau97 (AspAsn-dGAE) monomers and LMT-bound Tau97 (AspAsn-dGAE) monomers.

Figure 4. PCA Subspace comparison of 30 molecular trajectories obtained by independent molecular dynamics simulations of Tau97 monomers. A. Hierarchical clustering of PCA subspace of the 30 trajectories. B. Subspace overlap between the 30 molecular trajectories.

Figure 5. Heatmap of 178 pairwise RMSD measurements between Tau97 protein conformations derived from the MD sampling (ligand free) and PCA analysis of Figures 3-4.

Figure 6. PCA Subspace comparison of 22 molecular trajectories obtained by cycles of molecular dynamics simulations of LMT-bound Tau97 complexes followed by docking rescoring to identify tightly bound complexes.

Figure 7. PCA subspace comparison between the 22 molecular trajectories of Figure 6 and the 30 molecular trajectories of Figure 4.

Figure 8. LMT-Tau97 pose found to stabilise the monomer Tau97 protein. A. Cartoon representation of Tau97bound to LMT (sticks representation), the bottom left is the figure rotated by 90 degrees, and also showing the residues within the binding site. Top right is the LMT binding site. Bottom right is a space filling and surface representation of LMT-bound Tau97. The colours correspond to the colours in the primary sequence indicated at the top. B. Plot of the RSMD (root mean square deviation) of the residues surrounding LMT, D_{314}-S_{316}, E_{342}-K_{353}, and K_{370}-H_{374}. Line plot shows the RSMD from the first frame (top curve) and the RSMD from each preceding frame (bottom curve), plotted as the mean for every 200ps of simulation time. Error bars indicate the standard deviation for every 200ps of simulation.