



**HAL**  
open science

## A molecular mechanism for transthyretin amyloidogenesis

Ai Woon Yee, Matteo Aldeghi, Matthew Blakeley, Andreas Ostermann, Philippe Mas, Martine Moulin, Daniele de Sanctis, Matthew Bowler, Christoph Mueller-Dieckmann, Edward P. Mitchell, et al.

► **To cite this version:**

Ai Woon Yee, Matteo Aldeghi, Matthew Blakeley, Andreas Ostermann, Philippe Mas, et al.. A molecular mechanism for transthyretin amyloidogenesis. *Nature Communications*, 2019, 10 (1), pp.925-1-925-10. 10.1038/s41467-019-08609-z . hal-02085946

**HAL Id: hal-02085946**

**<https://hal.science/hal-02085946v1>**

Submitted on 23 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.








Distributed under a Creative Commons Attribution 4.0 International License

ARTICLE

<https://doi.org/10.1038/s41467-019-08609-z>

OPEN

# A molecular mechanism for transthyretin amyloidogenesis

Ai Woon Yee <sup>1,2</sup>, Matteo Aldeghi <sup>3</sup>, Matthew P. Blakeley <sup>2</sup>, Andreas Ostermann <sup>4</sup>, Philippe J. Mas<sup>5</sup>, Martine Moulin<sup>1,2</sup>, Daniele de Sanctis<sup>6</sup>, Matthew W. Bowler <sup>7</sup>, Christoph Mueller-Dieckmann<sup>6</sup>, Edward P. Mitchell<sup>1,6</sup>, Michael Haertlein<sup>2</sup>, Bert L. de Groot<sup>3</sup>, Elisabetta Boeri Erba<sup>5</sup> & V. Trevor Forsyth<sup>1,2</sup>

Human transthyretin (TTR) is implicated in several fatal forms of amyloidosis. Many mutations of TTR have been identified; most of these are pathogenic, but some offer protective effects. The molecular basis underlying the vastly different fibrillation behaviours of these TTR mutants is poorly understood. Here, on the basis of neutron crystallography, native mass spectrometry and modelling studies, we propose a mechanism whereby TTR can form amyloid fibrils via a parallel equilibrium of partially unfolded species that proceeds in favour of the amyloidogenic forms of TTR. It is suggested that unfolding events within the TTR monomer originate at the C-D loop of the protein, and that destabilising mutations in this region enhance the rate of TTR fibrillation. Furthermore, it is proposed that the binding of small molecule drugs to TTR stabilises non-amyloidogenic states of TTR in a manner similar to that occurring for the protective mutants of the protein.

<sup>1</sup>Faculty of Natural Sciences, Keele University, Staffordshire ST5 5BG, UK. <sup>2</sup>Institut Laue-Langevin, 71 avenue des Martyrs, 38042 Cedex 9 Grenoble, France. <sup>3</sup>Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany. <sup>4</sup>Heinz Maier-Leibnitz Zentrum (MLZ), Technische Universität München, 85748 Garching, Germany. <sup>5</sup>Univ. Grenoble Alpes, CEA, CNRS, IBS, 38000 Grenoble, France. <sup>6</sup>European Synchrotron Radiation Facility, 71 avenue des Martyrs, 38042 Cedex 9 Grenoble, France. <sup>7</sup>EMBL, Grenoble Outstation, 71 avenue des Martyrs, 38042 Cedex 9 Grenoble, France. These authors contributed equally: Ai Woon Yee, Matteo Aldeghi. Correspondence and requests for materials should be addressed to V.T.F. (email: [tforsyth@ill.eu](mailto:tforsyth@ill.eu))

**H**uman transthyretin (TTR) is a 55 kDa homotetrameric protein that transports thyroxine ( $T_4$ ) and retinol-binding protein (RBP) in the serum and cerebrospinal fluid (Fig. 1a–c). Wild-type (WT) TTR is intrinsically amyloidogenic, and causes amyloid fibril formation in elderly individuals, resulting in senile systemic amyloidosis (SSA)<sup>1,2</sup>. A large number (>100) of TTR mutants have been identified (<http://amyloidosismutations.com>). The majority of these are implicated in familial amyloid polyneuropathy/myocardiopathy (FAP/FAC) and several other forms of amyloidosis<sup>3–5</sup>. Of special interest are the T119M and S52P mutations. The T119M mutation is non-amyloidogenic, and heterozygous individuals carrying the T119M mutation and an amyloidogenic mutation, such as V30M, either remain asymptomatic of FAP or present a more benign form of the disease<sup>6,7</sup>. On the other hand, individuals carrying the S52P mutation develop aggressive and early-onset fatal amyloidosis<sup>8,9</sup>. Currently, only one drug—tafamidis (trade name *Vyndaqel*; Pfizer)—has been granted market authorisation for the treatment of TTR-FAP in adult patients having stage 1 symptomatic polyneuropathy<sup>10</sup>.

Considerable effort has been invested in seeking to identify the factors that underlie TTR amyloid formation including X-ray diffraction<sup>11–13</sup>, NMR<sup>14–16</sup>, solution scattering<sup>17</sup>, and mass spectrometry (MS)<sup>18</sup>. However, a molecular level understanding of the widely different behaviours of various TTR mutants has remained elusive<sup>11,12</sup>. In this paper, native MS, neutron crystallography, and computer simulations have been used to investigate the effects of the T119M and S52P mutations on TTR stability and the molecular determinants of amyloid formation. It is found that the S52P mutation destabilises TTR by altering the stability of the CD loop in the protein monomer. In contrast, the T119M mutation stabilises the dimer–dimer interface as well as TTR’s tertiary structure. Furthermore, it is shown how the stability of TTR tetramers is coupled to those of the monomers and dimers. Finally, it is suggested that tafamidis stabilises native tetrameric TTR in a similar way to that occurring for the T119M mutant in terms of thermodynamic, kinetic and structural features. A molecular mechanism is proposed by which the vastly different behaviours of the various TTR mutants can be understood.

## Results

**Kinetics of TTR tetramer dissociation.** Tetramer dissociation is thought to be the rate-limiting step in amyloid formation<sup>19–22</sup>. MS was used to follow the subunit exchange dynamics of TTR over several days so that the effect of the two mutations, and that of tafamidis, on the rate of dissociation of the tetramer species could be compared. These experiments relied on the use of hydrogenated (H) and deuterated (D) TTR to track otherwise identical protein chains. H- and D-TTR were mixed in an equimolar ratio and monitored over time to observe the rate of dissociation of the homo-tetramers (4H, 4D) and the rate of formation of the hetero-tetramers (2H2D, 1H3D, 3H1D; Fig. 1d and Supplementary Figure 1). The two 2H2D assemblies that can be formed may have different association constants, given that one requires only association of homo-dimers while the other requires the formation of hetero-dimers before assembly into tetramers. However, these two forms cannot be differentiated by MS, and are necessarily considered together. Since it has been shown that the kinetics of TTR subunit exchange is susceptible to isotope effects<sup>18</sup>, a deuterated TTR species [(D)S52P] was used as a reference to allow a meaningful comparison of the TTR variants. The change in relative abundance of each tetrameric species over time, and the relative association/dissociation rate constants are shown in Fig. 1e–h.

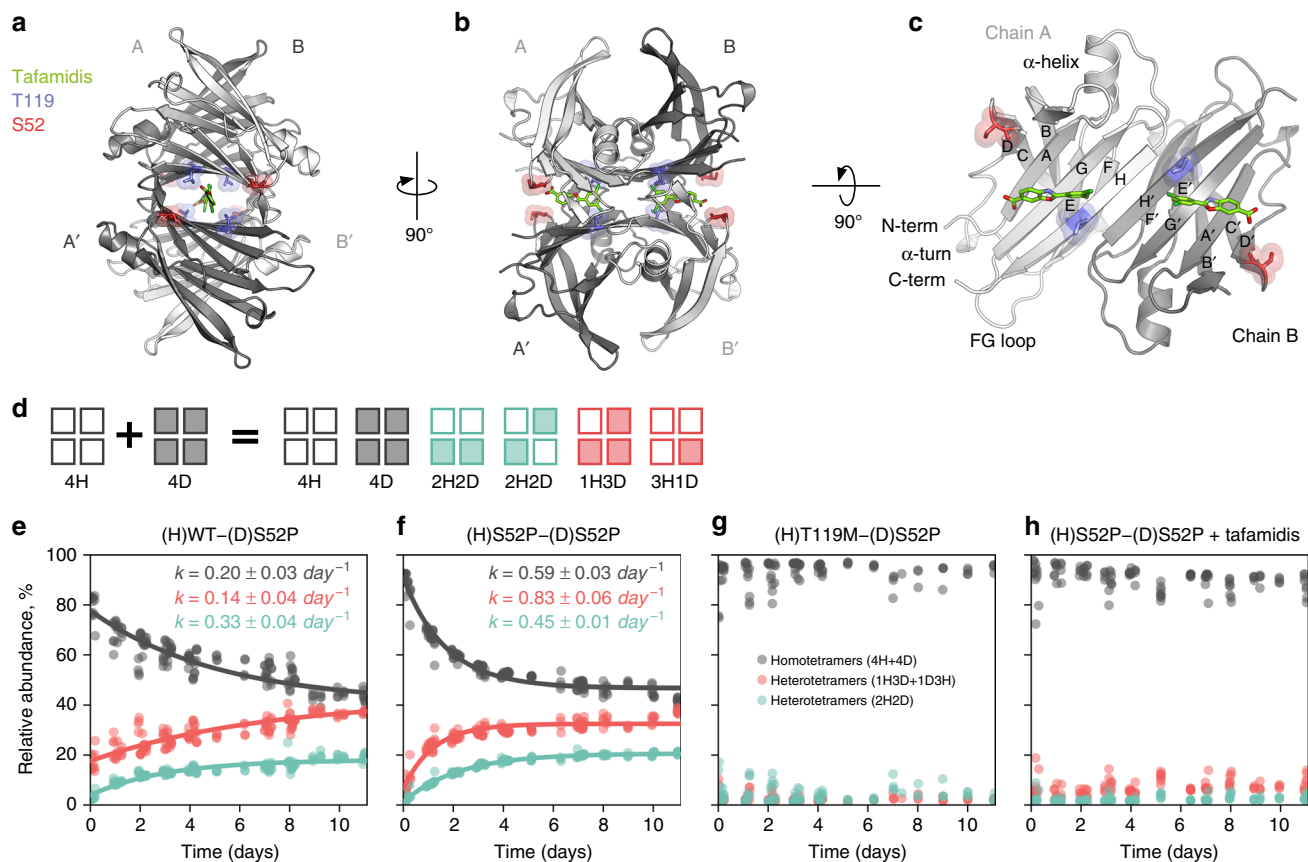
The dissociation rate of homo-tetramers in the (H)WT-(D)S52P experiment (Fig. 1e) was  $0.20 \pm 0.03 \text{ day}^{-1}$ . Among the hetero-tetramer species, those containing identical dimers (2H2D; two hydrogenated and two deuterated TTR monomers) were formed at a higher rate ( $0.33 \pm 0.04 \text{ day}^{-1}$ ) than those formed by two different dimers (3H1D and 1H3D;  $0.14 \pm 0.03 \text{ day}^{-1}$ ). The observed rate of dissociation of homo-tetramers in the (H)S52P-(D)S52P experiment (Fig. 1f) was about three times larger than in the (H)WT-(D)S52P ( $0.59 \pm 0.03 \text{ day}^{-1}$ ). In this case, however, the 3H1D/1H3D species were formed faster than the 2H2D ones:  $0.83 \pm 0.06 \text{ day}^{-1}$  and  $0.45 \pm 0.01 \text{ day}^{-1}$ , respectively. In the (H)T119M-(D)S52P experiment (Fig. 1g), almost no tetramer dissociation was observed over the course of eleven days—a timescale over which essentially all of the TTR would be recycled in a physiological context (the biological half-life of TTR is 1–2 days<sup>23</sup>). A final experiment was carried out with (H)S52P and (D)S52P in the presence of tafamidis at a molar ratio of 1:1 (Fig. 1h). In this case, tetramer dissociation was almost eliminated by the presence of the drug. This also demonstrated that one molecule of tafamidis per tetramer was sufficient for effective stabilisation.

Overall, the MS experiments revealed that the S52P mutation increases the rate of dissociation of TTR tetramers by about a factor of three as compared with the WT (reduced tetramer stability was also observed as a function of pH—see Supplementary Figure 2a–b). In strong contrast, both the T119M mutation and tafamidis effectively abolished tetramer dissociation.

**Effect of mutations on TTR equilibria.** Free energy (G) calculations were used to study the effects of the T119M and S52P mutations on the thermodynamic stability of TTR tetramers, dimers and monomers. These calculations were based on MD simulations<sup>24</sup> and the thermodynamic cycle is shown in Fig. 2. Control over the potential energy function used allowed specific residues to be changed into other ones, so that the free energy change associated with the modifications could be calculated for the protein in its tetrameric ( $\Delta G_{\text{tetr}}$ ), dimeric ( $\Delta G_{\text{dime}}$ ) or monomeric ( $\Delta G_{\text{mono}}$ ) states. Differences between these  $\Delta G$  values provide an estimate of the stability of the mutant tetramers, dimers and monomers relative to the WT:  $\Delta\Delta G_{\text{tetr}}$ ,  $\Delta\Delta G_{\text{dime}}$ , and  $\Delta\Delta G_{\text{mono}}$  respectively (Fig. 2).

The calculations suggested that the S52P mutation would not have a significant effect on the stability of the tetramers and dimers for their dissociation to dimers and monomers, respectively ( $\Delta\Delta G_{\text{tetr}}$  and  $\Delta\Delta G_{\text{dime}}$ ). However, a large and statistically significant  $\Delta\Delta G_{\text{mono}}$  ( $+2.1 \text{ kcal mol}^{-1}$ ) indicated that the introduction of the S52P mutation had a destabilising effect on the monomer with respect to the WT-TTR, corresponding to a shift towards unfolded TTR by a factor of  $\sim 34$ . The T119M mutation had no effect on the stability of the dimers. However, in contrast to the case for S52P, a large stabilising effect on the tetramers was observed ( $\Delta\Delta G_{\text{tetr}} = -8.9 \text{ kcal mol}^{-1}$ ; equilibrium constant fold-change of  $\sim 3 \cdot 10^6$ ). In addition, the T119M monomers were calculated to be more stable than WT monomers by  $1 \text{ kcal mol}^{-1}$  ( $\sim 5$ -fold increase in stability), a small but statistically significant effect.

When considering the overall process of tetramer dissociation and unfolding, the S52P mutation was estimated to shift the equilibrium toward the unfolded state by  $8.1 \text{ kcal mol}^{-1}$  (a change of  $\sim 8 \cdot 10^5$ ), whereas the T119M mutation was estimated to have the opposite effect of stabilising the tetramers by  $12.4 \text{ kcal mol}^{-1}$  (a change of  $\sim 10^9$ ). Given that the two mutation sites are not in close contact, it is reasonable to assume additivity for their associated energetic effects and on this basis the data suggest that in a double S52P/T119M-TTR mutant, the T119M mutation



**Fig. 1** Structure of human TTR and experimental studies of TTR stability. **a–c** Crystallographic structure of WT-TTR (PDB-ID 4PVM). Tafamidis is shown in green (by superimposition of PDB-ID 6FFT), the S52 amino acid in red, and T119 in blue. **d** Scheme of the native MS experiment: 4H (hydrogenated) and 4D (deuterated) tetramers are mixed in equal parts; following dissociation, four hybrid tetrameric species may form by the assembly of monomeric and dimeric species. **e–h** MS result for the subunit exchange between deuterated S52P and hydrogenated WT, S52P, T119M, and S52P bound to tafamidis, respectively. Changes in relative abundance of homo- and hetero-tetrameric species over the course of 11 days are shown, along with an estimate of the of their association/dissociation rates (uncertainties are one standard deviation of the posterior densities)

would be able to balance the destabilising effect of the S52P mutation; this is in agreement with the observed *in vivo* protective effect of the T119M mutation<sup>25,26</sup>.

In summary, the free energy calculations show that the S52P mutation results in a TTR protein having a less stable fold. However, this appears to have little direct impact on tetramer and dimer dissociation. In contrast, the T119M mutation directly affects tetramer dissociation to dimers by stabilising the tetramer species and also has the effect of stabilising the protein fold.

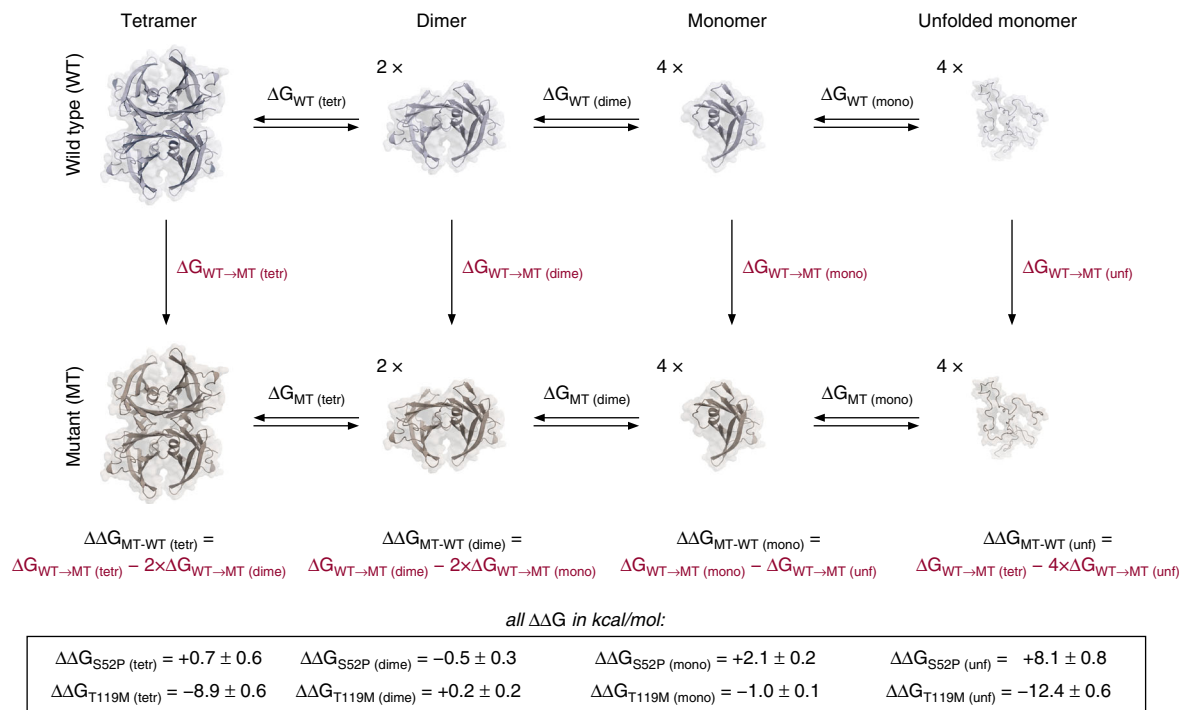
**Neutron diffraction reveals the basis of TTR instability.** In order to study the intramolecular interactions of S52P- and T119M-TTR in atomic detail, room-temperature neutron crystallography studies were carried out. The crystal structures of the S52P mutant, the T119M mutant, and the S52P mutant in complex with tafamidis, were determined at resolutions of 1.80 Å, 1.85 Å, and 2.00 Å, respectively. Most of the interaction network at the monomer–monomer and dimer–dimer interfaces (Supplementary Figure 3), as well as the energetics of bridging water molecules (Supplementary Figure 4), were found to be conserved across the WT and mutant structures. However, crucial differences are identified as the source of the different kinetic and thermodynamic stabilities observed.

In the T119M structure, the longer Met119 side chain extends across the thyroxine-binding channel into a hydrophobic pocket

surrounded by residues Leu17, Ala19, Leu110 and Val121 (Supplementary Figure 5). These interactions, as also noted in previous studies, enhance the association of two dimers and hence the overall stability of the tetramer<sup>27</sup>.

The amino acid at position 52 plays a crucial role. In the WT and the T119M mutant structures, the amide-D and the side-chain hydroxyl of the Ser52 residue form hydrogen bonds with O<sub>γ</sub> of Ser50, resulting in a stable CD loop (Fig. 3a). The highly amyloidogenic S52P mutant has a proline residue at position 52 in place of the serine present in the WT. The lack of a main-chain amide-H and a side-chain hydroxyl group in the proline residue prevents the formation of two hydrogen bonds and causes a loss of stability of the protein (Fig. 3b). The distance between amide-N of residue 52 and C<sub>α</sub> of Ser50 in the WT and T119M structures is 4.2 Å (Fig. 3a). The same distance is 0.4 Å wider in the S52P mutant, implying a looser CD loop (Fig. 3b). This destabilises the β-turn where the residue is located as well as the associated C and D strands.

In the S52P/tafamidis complex, the CD loop is not stabilised by tafamidis binding and the distance between Ser52-N and Ser50-C<sub>α</sub> remains as 4.6 Å. In contrast to the findings of Bulawa et al.<sup>28</sup>, no water molecules were found bridging the protein and tafamidis in this complex. Instead, tafamidis induced a 180° flip to the Thr119 side chain (Fig. 3c), with the hydroxyl moiety now forming a hydrogen bond to a water molecule that bridges the two TTR dimers via Asp18; this was observed in both binding sites. Interestingly, this water molecule fills the same space that is



**Fig. 2** Relative stability of mutant tetramers, dimers, and monomers, with respect to WT-TTR.  $\Delta G_{(\text{tetr})}$  is the stability of the mutant tetramers with respect to WT-TTR for the process of dissociation to dimers;  $\Delta\Delta G_{(\text{dime})}$  represents the relative stability of the mutant dimers when they dissociate to monomers;  $\Delta\Delta G_{(\text{mono})}$  indicates the relative stability of the mutant monomers for unfolding;  $\Delta\Delta G_{(\text{unf})}$  is the relative stability of the mutant tetramers for the overall processes of tetramer dissociation and unfolding. The uncertainties shown are one standard error of the mean, obtained from ten repeated calculations

occupied by the Met119 side chain in the non-pathogenic T119M mutant (Fig. 3d). Grand canonical Monte Carlo calculations<sup>29</sup> suggested that the stabilisation of this water molecule upon tafamidis binding may substantially contribute toward the affinity of tafamidis, by about  $-2.8 \text{ kcal mol}^{-1}$  (Supplementary Figure 4).

Collectively, the neutron structures of T119M-TTR, and S52P-TTR in *apo* and *holo* forms, have revealed that most interactions between the  $\beta$ -strands forming the core of the protein are conserved. However, the  $\beta$ -turn between the C and D strands is loosened in S52P-TTR, suggesting that this location is crucial to the reduced stability of this mutant. The appearance of a bridging water molecule in the S52P/tafamidis complex results in a conformation that resembles that of Met119 in T119M-TTR in this region of the protein, and may provide insight to the mechanism by which the drug stabilises the TTR tetramer.

#### Coupling between quaternary and tertiary structure stability.

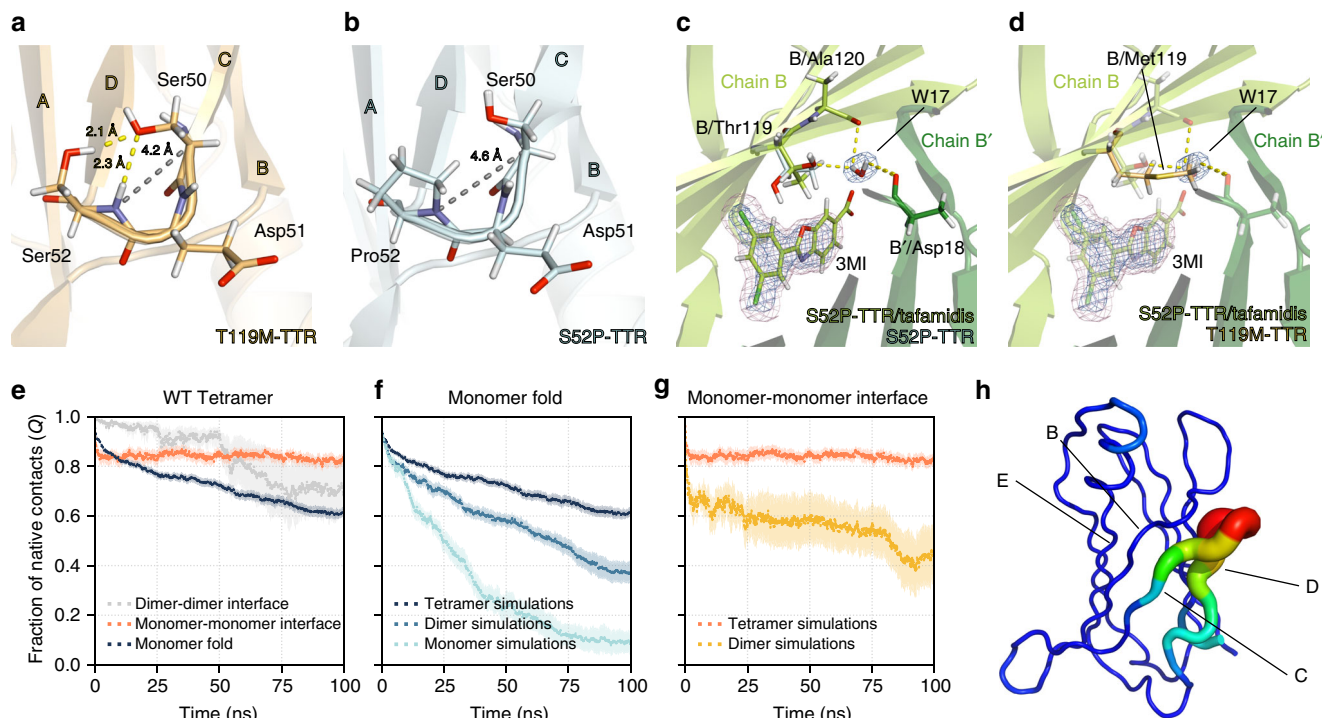
High-temperature MD simulations were used to study the effect of TTR quaternary structure on the kinetics of TTR unfolding. Simulations were carried out at 598 K for the tetramers, dimers and monomers of WT-TTR with two modern force fields (Amber99sb\*-ILDNP<sup>30,31</sup> and Charmm36<sup>32</sup>). As a measure of unfolding, the fraction of native contacts  $Q$  was employed<sup>33</sup>, where 1 indicates a folded protein and 0 an unfolded protein. A similar metric was used to define the fraction of monomer–monomer and dimer–dimer interface contacts retained during the high-temperature simulations (Fig. 3e; Charmm36 results in Supplementary Figure 6). At the temperature of 298 K, all monomers and interfaces were stable with  $Q$  close to one for simulations of up to 1  $\mu\text{s}$  (Supplementary Figure 7).

In Fig. 3f, the degree of unfolding of TTR monomers over the course of 100 ns simulations is shown for simulations of the tetramer, dimer and monomer in solution. The monomer in

solution was found to unfold faster than the monomers that were part of a dimer, which in turn unfolded faster than the monomers that were part of the tetramer. Thus, it appeared that the quaternary structure of TTR had an effect on the stabilisation of the individual monomeric chains. It is conceivable that the opposite would hold true as well: i.e., that highly stable monomeric chains (as in T119M) may enhance the stability of the tetramer, whereas unstable monomeric chains (as in S52P) are likely to have a detrimental effect on the stability of the tetramers. The same effect was observed for the stability of the monomer–monomer interface. This interface was more resistant to disruption than the dimer–dimer interface (Fig. 3e). However, when simulating the TTR dimer, the same interface became much more easily disrupted within the timeframe of the simulations (Fig. 3g).

Partial-least squares (PLS) functional mode analysis (FMA)<sup>34,35</sup> was used to identify regions of the TTR protein that contributed most to its unfolding while being part of a tetrameric unit. In Fig. 3h, a WT-TTR monomer is shown and colour-coded according to the root mean square fluctuations of the maximally correlated mode. The figure shows how the largest loss of native contacts observed was due to large motions in the C and D strands (Charmm36 results in Supplementary Figure 8). Similar observations were also made when calculating the fraction of native contacts by strand (Supplementary Figure 9).

These MD simulations showed that TTR tertiary structure is stabilised by the formation of tetrameric and dimeric units. Similarly, the monomer–monomer interface is stabilised by the association of TTR dimers into tetramers. Taken together, the observations explain why tetramer dissociation is the rate-limiting step in the formation of amyloid. In addition, in agreement with previous NMR data<sup>16,36</sup>, PLS-FMA identify the C and D strands as the regions of TTR tetramers most prone to unfolding.



**Fig. 3** Neutron structures and molecular dynamics results. **a** In the T119M mutant (as well as in the WT; not shown), the Ser52 residue forms two hydrogen bonds (yellow dash lines) with the Ser50 residue. The distance between Ser50-C $\alpha$  and the amide-N of Ser52 is 4.2 Å (grey), and it is the same in the WT (not shown). **b** In the S52P mutant, due to the absence of the hydrogen bonds between Pro52 and Ser50, the distances between Ser50-C $\alpha$  and the amide-N of Pro52 is longer (4.6 Å), creating a looser CD loop. **c** Binding of tafamidis (3MI) results in a change of orientation for residue Thr119 (cyan, before binding; green, after binding). In the S52P/tafamidis complex, the hydroxyl side chain of Thr119 forms a hydrogen bond to a water molecule. **d** The water molecule in the S52P/tafamidis complex occupies the same position as the side chain of Met119 residue (light orange) in the non-pathogenic T119M mutant. The binding site between chain B and B' is shown. The blue and magenta mesh show the  $2F_o - F_c$  neutron scattering length density map and the  $2F_o - F_c$  X-ray electron density map, respectively. Maps were contoured at  $1.5\sigma$ . **e-g** Loss of native fold and contacts in WT-TTR during high-temperature MD simulations with the Amber99sb\*-ILDNP force field; shown are the mean and its standard error from 10 simulation repeats (results for Charmm36 are in Supplementary Figure 6). **e** Degree of disruption of the monomer fold, the monomer-monomer interface, and the dimer-dimer interface, when simulating the WT tetramer over the course of 100 ns. **f** Degree of unfolding of TTR monomers as part of a tetramer, dimer, or monomer in solution. **g** Degree of disruption of the monomer-monomer interface when simulating the dimer and the tetramer. **h** Protein regions in the PLS-FMA mode contributing most to the change in the fraction of native contacts for the high-temperature simulations

## Discussion

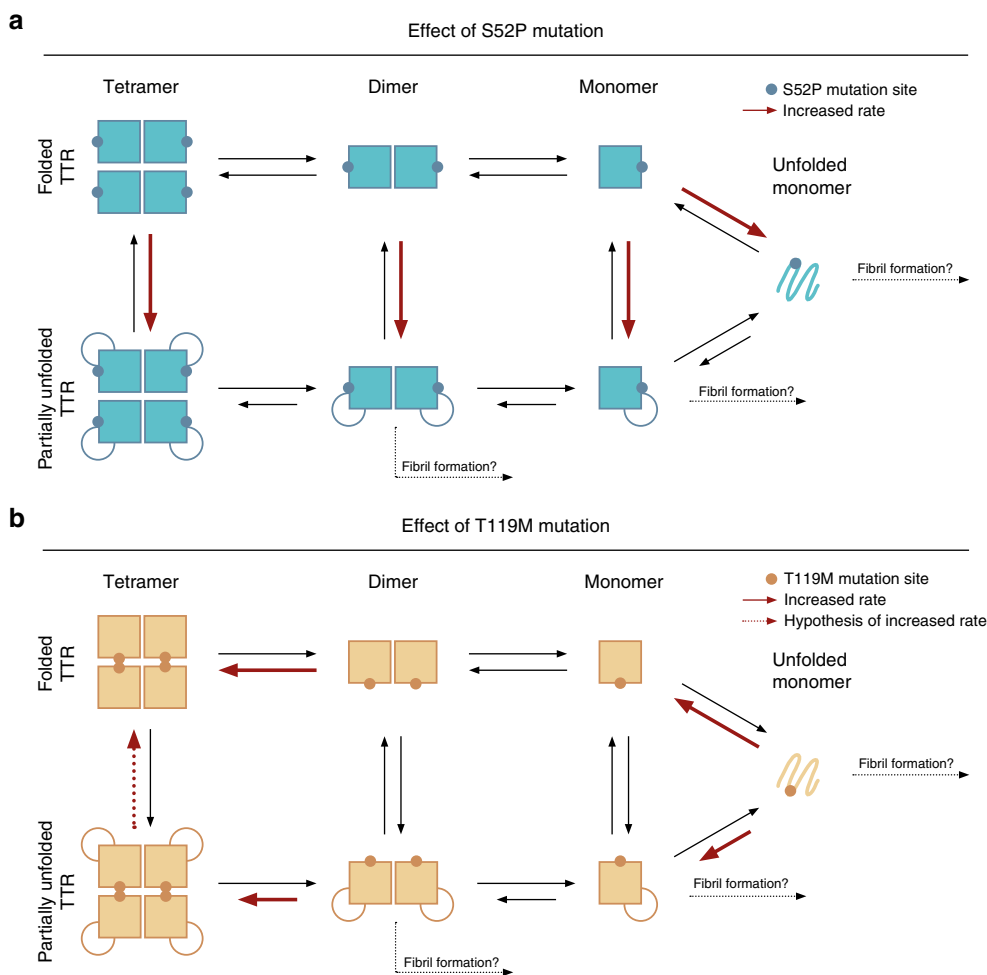
Based on the data presented, a molecular mechanism that explains the different behaviours of the T119M and S52P mutants is proposed (Fig. 4). It is suggested that partial unfolding events originating at the CD strands lead to a parallel equilibrium of folded and partially unfolded TTR states (tetramers, dimers and monomers) that allows the effects of the S52P and T119M mutations to be rationalised in the context of the current and previous experimental observations<sup>13,16,36,37</sup>. Given that a number of other pathogenic mutations are located on the C and D strands (e.g. L55P, S50R, E54G), the mechanism proposed here may apply to other mutations in this region of the protein<sup>8,13,38,39</sup>.

S52P-TTR tetramers were observed to dissociate faster than the WT in solution. However, no thermodynamic or structural cause for a direct destabilisation of S52P tetramers could be established from these observations or from those published previously. In contrast, the free energy calculations described here provide evidence that the tertiary structure of S52P-TTR is thermodynamically less stable than that of the WT. This, together with structural (neutron diffraction) and dynamical (simulation) evidence of a loose CD loop, suggests that unfolding events at the CD strands are likely. As it is expected that a partially unfolded TTR would proceed towards dissociated species and amyloid formation faster than the corresponding fully folded protein (Fig. 4a), this parallel equilibrium can explain the high amount of

amyloid fibrillation of S52P-TTR (Supplementary Figure 2c) as well as the faster tetramer dissociation rates observed.

T119M-TTR tetramers were found to be highly resistant to dissociation. The free energy calculations estimated a stabilising effect of 8.9 kcal mol<sup>-1</sup> on the tetramer, which corresponds to a change in equilibrium constant of  $\sim 3 \cdot 10^6$ . This effect can be understood in terms of the increased contact area between the two dimers, which stabilises the dimer-dimer interface (Supplementary Figure 5). In addition, the free energy calculation results suggest a tertiary fold in T119M-TTR that is more stable than in the WT by about 1 kcal mol<sup>-1</sup>. As illustrated in Fig. 4b, tetramer stabilisation directly reduces the population of amyloidogenic building blocks. Furthermore, a stable tertiary structure can protect against the unfolding of monomeric TTR. Finally, the coupling between TTR quaternary structure and the stability of the TTR fold suggested by the simulations indicates that the highly stable T119M-TTR tetramer reduces the probability of the CD strands unfolding (Fig. 4b).

Stabilisation of the native tetrameric structure by small molecule binding to the T4 binding sites of TTR is a well-cited rationale for the inhibition of TTR amyloidogenesis<sup>40</sup>. MS kinetic experiments have shown how tafamidis, like the T119M mutation, strongly inhibits tetramer dissociation. The binding affinity of tafamidis to TTR quantifies the stabilising effect of the drug on TTR tetramers, and was measured to be  $\sim 12$  kcal mol<sup>-1</sup> ( $K_{d1} = 2$



**Fig. 4** Proposed model of mutational effect on TTR stability. Shorter arrows indicate a lower propensity to association/folding. Red arrows reflect the qualitative change in reaction rates for the mutants with respect to WT-TTR. **a** The S52P mutation increases the likelihood of partial and full unfolding of TTR monomeric units. This, in turn, also leads to a lower stability of TTR multimeric assemblies; hence partially unfolded dimers and monomers, and fully unfolded monomers, are generated at a higher rate. Some or all of these species are then removed from solution via the formation of amyloid fibrils. **b** In contrast to S52P, the T119M mutation stabilises the folded monomeric forms of TTR. Most importantly, this mutation has a large stabilising effect on TTR tetramers, which are believed not to form fibrils. Furthermore, it is proposed that the higher stability of the tetramer also has the effect of further stabilising the monomer fold

nM)<sup>28</sup>. The stabilising effect of T119M on the tetramer (with respect to the dimer) was found to be about 9 kcal/mol. However, considering the further gains in TTR fold stability conferred by the mutation ( $\Delta\Delta G_{T119M}(\text{mono})$ ), the overall thermodynamic effect is very similar to that of the tafamidis-S52P complex, with  $\Delta\Delta G_{T119M}(\text{unf}) \approx -12 \text{ kcal mol}^{-1}$ . Tafamidis therefore stabilises native TTR thermodynamically and kinetically in a manner that is comparable to T119M. However, it is difficult to predict the extent to which the two separate effects of tetramer and fold stabilisation may impact on the in vivo protection against amyloidogenesis. Structurally, the binding of tafamidis results in the reorientation of Thr119 and the emergence of a water molecule in the binding pocket. This water molecule is located in the same position as the side chain of Met119 in the T119M mutant.

The results presented also suggest that TTR tetramers have a stabilising effect on the monomer–monomer interface, which becomes weakened when TTR is in its dimeric form (Fig. 3g). Thus, even though the monomer–monomer interface is more stable than the dimer–dimer interface<sup>16,41</sup> when the tetramer is intact, the loss of native quaternary structure can destabilise the monomer–monomer interface such that TTR dimers are able to quickly proceed to further dissociation into monomers

(in agreement with the data described here and elsewhere<sup>22,42</sup>). This observation further explains why tetramer dissociation is the rate-limiting step in the formation of amyloid. It is noted that dissociation might not be the only factor causing the formation of TTR fibrils (e.g. seeding with ex vivo fibrils has been found to promote the formation of fibrils in vitro) but it is a necessary one<sup>43</sup>.

Finally, a number of studies<sup>44–47</sup> have identified a TTR fragment (residues 49–127) in ex vivo amyloid fibrils. This fragment is formed by proteolytic cleavage of the peptide bond between Lys48 and Thr49, with the S52P mutant being particularly susceptible to it<sup>45,47</sup>. Within the mechanism set out in Fig. 4, the higher rate of proteolytic cleavage in S52P-TTR can be explained by an increased accessibility to Lys48 due to a higher propensity to unfold at the CD strands. Cleavage of the unfolded structure causes an irreversible transition to a partially unfolded state, thus preventing refolding and ultimately enhancing the rate of fibril formation.

In summary, the results described here provide novel structural and dynamical insights into the opposing effects of the S52P and T119M mutations in TTR, as well as the effects of tafamidis binding on the stability of TTR. The results provide molecular

level detail of direct relevance to TTR amyloidogenicity, and have provided a framework for further investigation into the effects of residue mutation on TTR's states and their equilibria, and for the development of novel targeted therapies for FAP and SSA.

## Methods

**Protein preparation and crystallization.** The cDNA corresponding to the gene coding for the 127 amino acids of human transthyretin (TTR) protein was cloned into a pET-M11 vector encoding an N-terminal His<sub>6</sub> tag and a TEV cleavage site (EMBL Protein and Purification Facility, Germany) and then expressed in *E. coli* BL21 (DE3) cells (Invitrogen). The S52P and T119M mutants were generated using QuikChange Lightning Multi Site-Directed Mutagenesis kit (Stratagene). WT-TTR cDNA sequence and the sequences of primers used for S52P and T119M mutations are shown in Supplementary Table 1. Deuterated protein was produced in the Deuteration Laboratory (D-Lab) platform within ILL's Life Sciences Group<sup>48</sup>. Cells were adapted to perdeuterated Enfors minimal medium and grown in a fed-batch fermenter culture at 30 °C using *d*<sub>8</sub>-glycerol (99% deuterium; Euriso-top) as the only carbon source. H-TTRs and D-TTRs were purified in an identical manner. Cell paste was resuspended homogeneously in lysis buffer (20 mM Tris pH 8, 250 mM NaCl, 3 mM imidazole) in the presence of EDTA-free protease inhibitor cocktail (Roche). The cells were lysed by ultrasonication on ice at 50% amplitude and 25 sec pulses. The lysate was cleared by centrifugation at 18,000 rpm at 4 °C for 30 min. The supernatant containing crude protein extracts was recovered and purified using benchtop gravity-flow chromatography. Nickel-nitriloacetic acid (Ni<sup>2+</sup>-NTA) resin was pre-washed with lysis buffer and left in incubation with the protein at 4 °C for at least an hour to increase binding efficiency. The protein/resin mix was then loaded into 10 ml disposable column. The column was washed three times with wash buffer (20 mM Tris pH 8, 15 mM imidazole, NaCl at 500 mM, 1 M, and 250 mM, respectively) before the protein was eluted (elution buffer: 20 mM Tris pH 8, 250 mM NaCl, 250 mM imidazole). Fractions containing protein were then pooled together for TEV protease treatment to remove His<sub>6</sub>-tag and dialysed into gel filtration buffer (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM DTT) overnight at 4 °C. Protein was separated from the cleaved poly-histidine tail using the same gravity-flow column described above. Flowthrough from the column was concentrated, filtered (with 0.22 µm membrane pore size) and loaded onto Superdex 75 HiLoad 16/600 gel filtration column (GE Healthcare) running at 1 ml/min at room temperature. Peak fractions were pooled and concentrated using Amicon Ultra centrifugal filter units (Millipore) and stored at -20 °C. Analysis was carried out on 12% Tris-Tricine gel after each step of purification to monitor the purity of the protein. The details of the expression and purification have been described previously<sup>18,49</sup>. While the deuterated TTR was expressed under conditions where only deuterium atoms were present, protiated (<sup>1</sup>H-based) solutions were used during purification. Prior to protein crystallization, the purified protein was buffer-exchanged using a deuterated solution so that the labile proton atoms acquired during the purification steps that involved hydrogenated solutions were replaced by deuterium. Tafamidis (HPLC purity: ≥ 98%; Carbosynth Limited, U.K.) was dissolved in 100% deuterated dimethyl sulfoxide (D-DMSO) for crystallization. All crystals were grown by sitting-drop vapour diffusion at 18 °C using deuterated protein and deuterated solution. The TTR S52P mutant crystal (~0.60 mm<sup>3</sup>) were grown in 1.9 M malonate pD 6.4 with a protein concentration of 25 mg/ml (drop volume 50 µl; protein-to-buffer ratio of 7:5), while the T119M mutant crystal (~0.80 mm<sup>3</sup>) were in 1.9 M malonate pD 5.9 with 40 mg/ml protein concentration (drop volume 50 µl; protein-to-buffer ratio of 1:1). For the crystal of the S52P/tafamidis complex (~0.11 mm<sup>3</sup>), a protein concentration of 20 mg/ml and tafamidis concentration of 678 µM in 10% D-DMSO (protein-to-ligand molar ratio of 1:2) were used (drop volume 40 µl; protein-to-buffer ratio of 1:1). For neutron data collection, the crystals were mounted in quartz capillaries and surrounded by a small amount of mother liquor from the crystallization well. The capillaries were sealed tightly using wax to eliminate the diffusion of gas and atmospheric water.

**Neutron data collection and processing.** Details of the neutron data collection for WT-TTR have been reported by Haupt et al.<sup>49</sup>. Neutron quasi-Laue diffraction data from the crystals of the S52P mutant and the S52P/tafamidis complex were collected at room temperature using the LADI-III diffractometer<sup>50</sup> at the Institut Laue-Langevin (ILL), Grenoble. For the S52P mutant, a neutron wavelength range ( $\Delta\lambda/\lambda = 30\%$ ) of 2.9–3.9 Å was used, with data extending to 1.8 Å resolution. As is typical for a Laue experiment, the crystal was held stationary at different  $\phi$  (vertical rotation axis) settings for each exposure. A total of 26 images (exposure time of 12 h per image) were collected from two different crystal orientations. For the S52P/tafamidis complex, a neutron wavelength range of 2.7–3.6 Å was used with data extending to 2.0 Å resolution. In total, 13 images were collected (exposure time of 8 h per image) from three different crystal orientations. The neutron diffraction images were indexed and integrated using the LAUE suite program LAUEGEN<sup>51</sup>. The program LSCALE<sup>52</sup> was used to determine the wavelength-normalisation curve using the intensity of symmetry-equivalent reflections measured at different wavelengths and to apply wavelength-normalisation calculations to the observed data. The data were then merged in SCALA<sup>53</sup>. Relevant data collection statistics are summarised in Supplementary Table 2.

Neutron monochromatic diffraction data for the T119M mutant crystal were collected at room temperature using the instrument BIODIFF<sup>54</sup> operated by the Forschungsreaktor München II (FRM II) and the Jülich Centre for Neutron Science (JCNS) at Garching. For data collection, the wavelength was set at 2.67 Å ( $\Delta\lambda/\lambda = 2.9\%$ ). A total of 202 frames were recorded with a rotation range of 0.35° and an exposure time of 53 min per frame. The diffraction data were indexed, integrated and scaled using HKL2000<sup>55</sup> to a resolution of 1.85 Å. The output file of HKL2000 in SCA format was converted into MTZ format by using *scalepack2mtz* program in the CCP4 suite<sup>56</sup>. Relevant data collection statistics are summarised in Supplementary Table 2.

**X-ray data collection and processing.** X-ray diffraction data of the crystals of all three variants were recorded on beamline ID30B<sup>57</sup> at the European Synchrotron Research Facility (ESRF), Grenoble using a heavily attenuated X-ray beam (1.96% for S52P; 19.17% for T119M; 9.83% for S52P/tafamidis) of wavelength 0.9763 Å. For S52P and S52P/tafamidis complex, the same crystals used for data collection at LADI-III was used; whereas for T119M, a crystal from the same crystallization drop as the one used at BIODIFF was used. Data were recorded at room temperature on capillary-mounted crystals. Data were processed to the same maximum resolution of the corresponding neutron data with XDS<sup>58</sup>, scaled and merged with SCALA<sup>53</sup>, and converted to structure factors using TRUNCATE in the CCP4 suite<sup>56</sup>. Relevant data collection statistics are summarised in Supplementary Table 2.

**Joint neutron and X-ray structure refinement.** The PDB structures 5CLX (refined X-ray models of TTR S52P; data collection at 100 K) was used as starting models for joint X-ray and neutron refinement against the datasets of S52P and S52P/tafamidis; and 5CM1 (refined X-ray models of TTR T119M; data collection at 100 K) was used for the datasets of T119M. The *phenix.refine* program<sup>59</sup> in the PHENIX package<sup>60</sup> was used for refinement. The preparation of the starting model, the refinement settings and workflow, and the modelling of solvent molecules were as detailed in Haupt et al.<sup>41</sup>. Using the *ReadySet* option in PHENIX<sup>60</sup>, exchangeable hydrogen and deuterium atoms were placed at appropriate sites of the protein models with deuteriums elsewhere. For positions where both hydrogen and deuterium atoms were modelled, the occupancies of both were set to 0.5 and then their occupancies were refined, with their total occupancy constrained to 1. D<sub>2</sub>O molecules were added using *ReadySet* based on the positive neutron scattering length density in  $F_o - F_c$  maps. *Coot*<sup>61</sup> was used for model modifications, such as addition of solvent molecules, and rotamer- and torsion-angle adjustments, according to positive and negative nuclear scattering length density in both  $2F_o - F_c$  and  $F_o - F_c$  maps. The final refinement statistics are summarised in Supplementary Table 3.

**Monitoring subunit composition of tetrameric TTR by native MS.** The details of the instrumental arrangements and sample preparation protocols have been described previously<sup>18</sup>. MS analyses were carried out on a quadrupole time-of-flight mass spectrometer (Q-TOF Ultima, Waters Corporation, Manchester, U.K.) that was modified for the detection of high masses<sup>62,63</sup>. As has been tested previously<sup>18</sup>, experiments carried out at 4 °C in static (non-shaking) conditions provided well-resolved MS spectra, hence these conditions were used for the work described here. Prior to native MS analysis, proteins were buffer-exchanged into 250 mM ammonium acetate pH 7. All the labile sites (i.e. N- and O-bound deuterium atoms) were allowed to exchange from D to H. The mass difference between (H)TTR and (D)TTR was contributed by carbon-bound Ds that were incorporated into the amino acid chain during synthesis and were non-exchangeable. Subunit exchange was initiated by mixing a solution containing one unlabelled protein variant and one with deuterium-labelled protein variant at 1:1 molar ratio. A concentration of 3 µM for each protein tetramer solution was used for all experiments. The relative abundance of the tetramers was calculated from the peak area of the 13 + to 15 + charge states and expressed as a percentage of the total area of the peaks assigned to the tetrameric TTR. For the experiment of mixing HS52P and DS52P in the presence of tafamidis, tafamidis was in 0.5% DMSO at a TTR:tafamidis molar ratio of 1:1. In this case, the relative abundance was calculated from the peak area of the 12 + to 14 + charge states. The presence of DMSO resulted in broader peaks on the MS spectra and less charge on the protein particles. The MS data were fitted by Bayesian regression using a first order kinetics model.

**Native MS data fitting and parameter estimation.** Bayesian regression of the native MS data was performed in *python* using the *PyMC3* library<sup>64</sup>. The following one-phase exponential decay model was used:

$$y(t) = c + (y_0 - c)e^{-kt} \quad (1)$$

where  $t$  is the time (in minutes),  $y(t)$  is the relative abundance of each species (in percentage),  $y_0$  is the abundance at  $t = 0$ ,  $c$  is the abundance at  $t = \infty$  and  $k$  the rate constant (in inverse minutes);  $c < y_0$  for dissociating species, and  $c > y_0$  for associating species. For  $y_0$ , a normally distributed prior was used, with mean 50% and standard deviation of 15% for dissociation, and with mean of 0% and standard deviation of 5% for association. For  $c$ , a uniform prior between 0% and  $y_0$  was used for dissociation, and between  $y_0$  and 100% for association. For  $k$ , a half normal prior with mean of 0 min<sup>-1</sup> and standard deviation of 10 min<sup>-1</sup> was used. White



noise was modelled with a half normal prior with zero mean and standard deviation of 10%. The posterior estimate of the parameters was obtained by drawing 10,000 Markov chain Monte Carlo (MCMC) samples with the No-U-Turn Sampler (NUTS) algorithm. Five thousand tuning steps were carried out and discarded before collecting the posterior samples used for the estimate. The maximum a posteriori estimate (MAP) of the parameter space was passed to the sampler as the starting value. The mean and standard deviation of the rate constants ( $k$ ) from their posterior distributions are reported.

**High-temperature molecular dynamics simulations.** Molecular dynamics (MD) simulations were performed with Gromacs 2016<sup>65</sup>. The starting structure for the simulations was the WT (PDB-ID 4PVM) variant of TTR protonated at pH 7 with *pdb2gmx*. The dimeric and monomeric models were obtained from the same neutron crystal structure: for the dimers, chains A and B were used, and for the monomers chain A was used. The protein was modelled with the Amber99SB\*-ILDNP<sup>30,31</sup> and the Charmm36 (Nov. 2016)<sup>32</sup> force field, and water molecules with the TIP3P model<sup>66</sup>. The protein was solvated in a dodecahedral box with periodic boundary conditions and a minimum distance between the solute and the box of 12 Å. Sodium and chloride ions were added to neutralise the systems at the concentration of 0.15 M.

Energy minimisation was performed using a steepest descent algorithm with force tolerance of 10 kJ mol<sup>-1</sup> nm<sup>-1</sup> for a maximum of 10,000 steps. 1 ns NVT and then NPT equilibrating simulations were performed with all solute heavy atoms restrained with a force constant of 1000 kJ mol<sup>-1</sup> nm<sup>-2</sup>. In these simulations, the temperature was coupled with the stochastic v-rescale thermostat by Bussi et al.<sup>67</sup>, at the target temperature of 298 K; the pressure was controlled with the Berendsen weak coupling algorithm<sup>68</sup> at a target pressure of 1 bar. A leap-frog integrator was used with a time-step of 4 fs and virtual sites<sup>69</sup> were used for all solute hydrogens. All bonds were constrained with the P-LINCS algorithm<sup>70</sup>. The particle mesh Ewald (PME) algorithm<sup>71</sup> was used for electrostatic interactions with a real space cut-off of 12 Å. The Verlet cut-off scheme was used with a Lennard-Jones interaction cut-off of 12 Å and a buffer tolerance of 0.005 kJ mol<sup>-1</sup> ps<sup>-1</sup>. For the production simulations at 298 K, a single unbiased 1 μs simulation in the NPT ensemble was performed using the Parrinello-Rahman pressure coupling scheme<sup>72</sup>. For the simulations at 598 K, ten simulations of 100 ns were performed in the NVT ensemble. Coordinates were saved every 50 ps.

The fraction of native contacts ( $Q$ ), as described by Best et al.<sup>33</sup>, was employed as a measure of the degree of protein (un)folded and native-likeness of protein-protein interfaces. When used to describe protein folding, the list of native contacts was built by taking all pairs of heavy atoms  $i$  and  $j$  that are within 4.5 Å of each other and are on two different residues separated by at least three other residues. When describing protein-protein interfaces, the original definition of the authors was adapted so that the list of native contacts included all pairs of heavy atoms ( $i, j$ ) that are within 4.5 Å of each other and are on two different protein chains. Once the set of atom pairs is defined,  $Q$  is then calculated as follows:

$$Q(x) = \frac{1}{N} \sum_{(i,j) \in N} \frac{1}{1 + \exp[\beta(r_{ij}(x) - \lambda r_{ij}^0)]} \quad (2)$$

where  $N$  is the number of atom pairs ( $i, j$ ) forming the native contacts,  $r_{ij}(x)$  is the distance between atom  $i$  and atom  $j$  in the configuration  $x$ ,  $r_{ij}^0$  is the distance between  $i$  and  $j$  is the native state (i.e. the starting neutron crystal structure used),  $\beta$  is a smoothing parameters with value of 5 Å<sup>-1</sup> and  $\lambda$  is a factor that accounts for fluctuations that takes the value of 1.8 for atomistic simulations. MD trajectories were analysed and  $Q$  calculated with scripts written in *python* and using the *mdtraj* library<sup>73</sup>.

Partial Least-Squares Functional Mode Analysis (PLS-FMA)<sup>34,35</sup> was performed on the main chain of the proteins and using the fraction of native contacts as the functional property of interest. In particular, we analysed with PLS-FMA the trajectories of the tetramer simulations, using the fraction of native contacts of each monomeric unit as the measure of unfolding. About 75% of simulation snapshots were used for training and 25% for validation. The model was built using 20 PLS components, which resulted in a Pearson correlation coefficient between the data and the model  $\geq 0.97$  for both the training and the validation sets, for both force fields. We refer to the ensemble weighted maximally correlated mode (ewMCM) as described Krivobokova et al.<sup>35</sup> simply as the PLS-FMA mode.

**Free energy calculations of protein mutation.** Alchemical free energy calculations were performed using Gromacs 2016<sup>65</sup> and a custom version of Gromacs 4.6 that implements the soft-core potential proposed by Gapsys et al.<sup>74</sup>. The input hybrid topologies were generated using the *pmx* python tool<sup>75</sup>. The starting structures for the calculations were the WT-TTR (PDB-ID 4PVM), S52P-TTR (PDB-ID 5NFW) and T119M-TTR (PDB-ID 5NFE) protonated at pH 7 with *pdb2gmx*. The dimeric and monomeric starting structures were obtained from the same neutron structures; for the dimers, chains A and B were used, and for the monomers chain A was used. The unfolded proteins were modelled as capped tripeptides, where the central amino acid was the one being mutated, and the other two were the adjacent amino acids in protein sequence. Proteins were modelled with the Amber99SB\*-ILDNP force field<sup>30,31</sup> and water molecules with the TIP3P model<sup>66</sup>. The proteins were solvated in a dodecahedral box with periodic boundary conditions and a minimum

distance between the solute and the box of 12 Å. Sodium and chloride ions were added to neutralise the systems at the concentration of 0.15 M.

Energy minimisation was performed using a steepest descent algorithm for a maximum of 10,000 steps. Temperature and pressure were equilibrated with ten independent 0.5 ns simulations in the NPT ensemble, with all solute heavy atoms restrained with a force constant of 1000 kJ mol<sup>-1</sup> nm<sup>-2</sup>. A leap-frog stochastic dynamics integrator<sup>76</sup> was used with time-step of 2 fs; temperature was coupled using Langevin dynamics at the target temperature of 298 K, and pressure with the Berendsen weak coupling algorithm<sup>68</sup> at a target pressure of 1 bar. The particle mesh Ewald (PME) algorithm<sup>71</sup> was used for electrostatic interactions with a real space cut-off of 11 Å, a spline order of 4, a relative tolerance of 10<sup>-5</sup> and a Fourier spacing of 1.2 Å. The Verlet cut-off scheme was used with a Van der Waals interaction cut-off of 11 Å and a buffer tolerance of 0.005 kJ mol<sup>-1</sup> ps<sup>-1</sup>. Bonds involving hydrogens were constrained with the P-LINCS algorithm<sup>70</sup>. Ten equilibrium simulations of 10 ns duration were then initiated from the last frame of the short (0.5 ns) equilibration simulations. From each of the ten equilibrium simulations, a non-equilibrium trajectory was spawned every 0.2 ns, for a total 50 trajectories per equilibrium simulation and 500 trajectories overall. Forward and reverse transformations were performed for both mutants, with 500 non-equilibrium trajectories per transformation: WT→S52P, S52P→WT, WT→T119M, T119M→WT. Coordinates for hybrid residues were built with *pmx*<sup>75</sup> after extracting the starting configurations of the systems. Energy minimisation was performed on the dummy atoms only, before equilibrating velocities with a 10 ps simulation. Then, the non-equilibrium alchemical transformation was performed over 100 ps. The non-equilibrium simulations were performed using a custom version of Gromacs 4.6 that implements the soft-core potential described in Gapsys et al.<sup>74</sup>, where both vdW and Coulombic interactions were soft-cored. Free energy differences were estimated using the Bennet's Acceptance Ratio (BAR)<sup>77</sup> as implemented in *pmx*<sup>75</sup>. Uncertainties in the  $\Delta G$  values were calculated by taking the standard error of the BAR estimate from the ten independent equilibrium simulations and related non-equilibrium trajectories.

**Grand canonical monte carlo calculations.** The simulation package ProtoMS 3.3 was used for Grand Canonical Monte Carlo (GCMC) calculations and data analysis<sup>29,78</sup>. The starting tetramer structures for the calculations were the WT-TTR (PDB-ID 4PVM), S52P-TTR (PDB-ID 5NFW), S52P-TTR/tafamidis (PDB-ID 6FFT) and T119M-TTR (PDB-ID 5NFE) with protonation states as resolved experimentally. Proteins were modelled with the Amber ff14SB force field<sup>79</sup>, water with the TIP3P model, and tafamidis with the GAFF/AM1-BCC force field<sup>66,80</sup>.

For the calculation of the stability of W17, a GC box was created around the water molecule by adding 2 Å padding in all three dimensions and used for both the calculations in S52P-TTR and S52P-TTR/tafamidis. For the calculation of the monomer-monomer interface hydration free energy, the grand canonical box was defined so to encompass the interface containing the conserved water molecules of interest; the same box was used for all proteins, which had been previously aligned. Because the hydration free energy results might be sensitive to the protonation state of the histidine residues in proximity of these conserved water molecules, from each experimental structure we generated the other two proteins by mutating the amino acids at positions 52 and 119. Thus, for each TTR variant, we performed three sets of calculations, based on each neutron structure and its respective protonation states of His31, His56, His88 and His90.

Protein residues that were further than between 16 and 20 Å away from the grand canonical region were removed, with the exact distance chosen to retain whole residues. The systems were then solvated up to a radius of 30 Å around the grand canonical region. All simulations were carried out at 298 K, and a 10 Å cut off was applied to the non-bonded interactions. Before initiating the GCMC simulations, the systems were equilibrated using 50 million (M) solvent-only moves in the canonical ensemble, so to equilibrate the water around the proteins. Water molecules present in the pre-defined GCMC box were then removed; this set of coordinates represented the starting point of the GCMC simulations.

For the calculation of the monomer-monomer interface hydration, a set of 32 simulations were performed at a range of Adams values from -29 to +2 at unit increments. For the calculations of the stability of W17, Adams values between -18 and 0 were used instead. For each window, 15 M equilibration moves were performed only on the grand canonical solute, with insertion, deletion and translation/rotation moves generated at the same ratio. An additional 5 M equilibration moves followed where the protein and solvent molecules were also sampled, before starting the production simulation of 50 M moves. Half of the MC moves were dedicated to the grand canonical water molecules, and the other half was split between protein residues and solvent in proportion to the number of solvent molecules and protein residues, according to the ratio of 1:5. Hamiltonian exchange was employed<sup>78</sup>, with exchanges being performed every 0.2 M steps. Data for analysis, energies, number of GC solutes present, and coordinates were saved to file every 0.1 M moves.

The above procedure was repeated three times for the monomer-monomer interface hydration calculation (i.e. 9 calculations for each TTR variant, three repeated calculations per starting structure) and ten times for the W17 stability calculation. The hydration free energy of the whole grand canonical region, and the binding free energy of individual water molecules, were computed using *grand canonical integration* (GCI) as described in Ross et al.<sup>29</sup>. GCI was performed using

the *calc\_gci.py* script that is part of the ProtoMS 3.3 tools, with the data from all repeats being analysed together. The amount of data discarded as equilibration was determined using the equilibration detection tool available in *calc\_series.py*. The same amount of data was discarded from all windows after determining the average number of moves needed for equilibration.

### Data availability

Atomic coordinates and diffraction data have been deposited in the Protein Data Bank (accession codes 5NFW for S52P-TTR, 5NFE for T119M-TTR, and 6FFT for S52P-TTR/tafamidis complex). Other data are available from the corresponding author upon reasonable request.

Received: 3 August 2018 Accepted: 14 January 2019

Published online: 25 February 2019

### References

- Cornwell, G. G. 3rd, Murdoch, W. L., Kyle, R. A., Westermark, P. & Pitkanen, P. Frequency and distribution of senile cardiovascular amyloid. A clinicopathologic correlation. *Am. J. Med.* **75**, 618–623 (1983).
- Westermark, P., Sletten, K., Johansson, B. & Cornwell, G. G. Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Proc. Natl Acad. Sci. USA* **87**, 2843–2845 (1990).
- Planté-Bordeneuve, V. & Said, G. Familial amyloid polyneuropathy. *Lancet Neurol.* **10**, 1086–1097 (2011).
- Planté-Bordeneuve, V. Update in the diagnosis and management of transthyretin familial amyloid polyneuropathy. *J. Neurol.* **261**, 1227–1233 (2014).
- Ando, Y., Nakamura, M. & Araki, S. Transthyretin-related familial amyloidotic polyneuropathy. *Arch. Neurol.* **62**, 1057–1062 (2005).
- Alves, I. L. et al. Thyroxine binding in a TTR Met 119 kindred. *J. Clin. Endocrinol. Metab.* **77**, 484–488 (1993).
- Harrison, H. H., Gordon, E. D., Nichols, W. C. & Benson, M. D. Biochemical and clinical characterization of prealbuminCHICAGO: an apparently benign variant of serum prealbumin (transthyretin) discovered with high-resolution two-dimensional electrophoresis. *Am. J. Med. Genet.* **39**, 442–452 (1991).
- González-Duarte, A., Lem-Carrillo, M. & Cárdenas-Soto, K. Description of transthyretin S50A, S52P and G47A mutations in familial amyloidosis polyneuropathy. *Amyloid* **20**, 221–225 (2013).
- Stangou, A. J. et al. Progressive cardiac amyloidosis following liver transplantation for familial amyloid polyneuropathy: implications for amyloid fibrillogenesis. *Transplantation* **66**, 229–233 (1998).
- European Medicines Agency (EMA). *European Public Assessment Report: Annex I. EMA website* (European Medicines Agency (EMA), 2011). <https://www.ema.europa.eu/en/medicines/human/EPAR/vyndaqel>
- Hörnberg, A., Eneqvist, T., Olofsson, A., Lundgren, E. & Sauer-Eriksson, A. E. A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *J. Mol. Biol.* **302**, 649–669 (2000).
- Palaninathan, S. K. Nearly 200 X-ray crystal structures of transthyretin: what do they tell us about this protein and the design of drugs for TTR amyloidoses. *Curr. Med. Chem.* **19**, 2324–2342 (2012).
- Serpell, L. C., Goldsteins, G., Dacklin, I., Lundgren, E. & Blake, C. C. F. The ‘edge strand’ hypothesis: prediction and test of a mutational ‘hot-spot’ on the transthyretin molecule associated with FAP amyloidogenesis. *Amyloid Int. J. Exp. Clin. Investig.* **3**, 75–85 (1996).
- Fitzpatrick, A. W. P. et al. Atomic structure and hierarchical assembly of a cross- $\beta$  amyloid fibril. *Proc. Natl Acad. Sci. USA* **110**, 5468–5473 (2013).
- Kim, J. H., Oroz, J. & Zweckstetter, M. Structure of monomeric transthyretin carrying the clinically important T119M mutation. *Angew. Chem. Int. Ed.* **55**, 1–5 (2016).
- Olofsson, A., Ippel, J. H., Wijmenga, S. S., Lundgren, E. & Öhman, A. Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. *J. Biol. Chem.* **279**, 5699–5707 (2004).
- Gronning, M., Campos, R. I., Hirschberg, D., Hammarström, P. & Vestergaard, B. Considerably unfolded transthyretin monomers precede and exchange with dynamically structured amyloid protofibrils. *Sci. Rep.* **5**, 11443 (2015).
- Yee, A. W. et al. Impact of deuteration on the assembly kinetics of transthyretin monitored by native mass spectrometry and implications for amyloidosis. *Angew. Chem. Int. Ed. Engl.* **55**, 9292–9296 (2016).
- Johnson, S. M. et al. Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidosis introduction to the transthyretin amyloidosis. *Acc. Chem. Res.* **38**, 911–921 (2005).
- Schneider, F., Hammarström, P. E. R., Kelly, J. W., Hammarström, P. & Kelly, J. W. Transthyretin slowly exchanges subunits under physiological conditions: A convenient chromatographic method to study subunit exchange in oligomeric proteins. *Protein Sci.* **10**, 1606–1613 (2001).
- Lai, Z., Colón, W. & Kelly, J. W. The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry* **35**, 6470–6482 (1996).
- Colon, W. & Kelly, J. W. Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **31**, 8654–8660 (1992).
- Oppenheimer, J. H., Surks, M. I., Bernstein, G. & Smith, J. C. Metabolism of iodine-131-labeled thyroxine-binding prealbumin in man. *Science* **149**, 748–750 (1965).
- Gapsys, V., Michielssens, S., Seeliger, D. & de Groot, B. L. Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. *Angew. Chem. Int. Ed. Engl.* **55**, 7364–7368 (2016).
- Alves, I. L., Hays, M. T. & Saraiva, M. J. Comparative stability and clearance of [Met30]transthyretin and [Met119]transthyretin. *Eur. J. Biochem.* **249**, 662–668 (1997).
- Coelho, T. Familial amyloid polyneuropathy: new developments in genetics and treatment. *Curr. Opin. Neurol.* **9**, 355–359 (1996).
- Sebastião, M. P., Lamzin, V., Saraiva, M. J. & Damas, A. M. Transthyretin stability as a key factor in amyloidogenesis: X-ray analysis at atomic resolution. *J. Mol. Biol.* **306**, 733–744 (2001).
- Bulawa, C. E. et al. Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proc. Natl Acad. Sci. USA* **109**, 9629–9634 (2012).
- Ross, G. A., Bodnarchuk, M. S. & Essex, J. W. Water sites, networks, and free energies with Grand Canonical Monte Carlo. *J. Am. Chem. Soc.* **137**, 14930–14943 (2015).
- Lindorff-Larsen, K. et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).
- Aliev, A. E. et al. Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics. *Proteins Struct. Funct. Bioinforma.* **82**, 195–215 (2014).
- Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., & MacKerell Jr., A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi1 and chi2 dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
- Best, R. B., Hummer, G. & Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl Acad. Sci. USA* **110**, 17874–17879 (2013).
- Hub, J. S. & de Groot, B. L. Detection of functional modes in protein dynamics. *PLoS Comput. Biol.* **5**, e1000480 (2009).
- Krivobokova, T., Briones, R., Hub, J. S., Munk, A. & de Groot, B. L. Partial least-squares functional mode analysis: application to the membrane proteins AQP1, Aqy1, and CLC-ec1. *Biophys. J.* **103**, 786–796 (2012).
- Liu, K., Cho, H. S., Lashuel, H. A., Kelly, J. W. & Wemmer, D. E. A glimpse of a possible amyloidogenic intermediate of transthyretin. *Nat. Struct. Biol.* **7**, 754–757 (2000).
- Marcoux, J. et al. A novel mechano-enzymatic cleavage mechanism underlies transthyretin amyloidogenesis. *EMBO Mol. Med.* **7**, 1337–1349 (2015).
- Miyata, M. et al. Role of the glutamic acid 54 residue in transthyretin stability and thyroxine binding. *Biochemistry* **49**, 114–123 (2010).
- Sebastião, M. P. et al. The crystal structure of amyloidogenic Leu55 -> Pro transthyretin variant reveals a possible pathway for transthyretin polymerization into amyloid fibrils. *J. Biol. Chem.* **273**, 24715–24722 (1998).
- Foss, T. R., Kelker, M. S., Wiseman, R. L., Wilson, I. A. & Kelly, J. W. Kinetic stabilization of the native state by protein engineering: implications for inhibition of transthyretin amyloidogenesis. *J. Mol. Biol.* **347**, 841–854 (2005).
- Haupt, M. et al. Binding site asymmetry in human transthyretin: Insights from a joint neutron and X-ray crystallographic analysis using perdeuterated protein. *IUCr* **1**, 429–438 (2014).
- Nettleton, E. J. et al. Protein subunit interactions and structural integrity of amyloidogenic transthyretins: evidence from electrospray mass spectrometry. *J. Mol. Biol.* **281**, 553–564 (1998).
- Saelices, L. et al. Amyloid seeding of transthyretin by ex vivo cardiac fibrils and its inhibition. *Proc. Natl Acad. Sci. USA* **115**, E6741–E6750 (2018).
- Ihse, E. et al. Amyloid fibrils containing fragmented ATTR may be the standard fibril composition in ATTR amyloidosis. *Amyloid* **20**, 142–150 (2013).
- Thylén, C. et al. Modifications of transthyretin in amyloid fibrils: analysis of amyloid from homozygous and heterozygous individuals with the Met30 mutation. *EMBO J.* **12**, 743–748 (1993).
- Bergström, J. et al. Amyloid deposits in transthyretin-derived amyloidosis: cleaved transthyretin is associated with distinct amyloid morphology. *J. Pathol.* **206**, 224–232 (2005).

47. Mangione, P. P. et al. Proteolytic cleavage of Ser52Pro variant transthyretin triggers its amyloid fibrillogenesis. *Proc. Natl Acad. Sci. USA* **111**, 1539–1544 (2014).
48. Haertlein, M., Moulin, M., Devos, J. M., Laux, V., Dunne, O., & Forsyth, V. T. (2016). Biomolecular deuteration for neutron structural biology and dynamics. *Methods Enzymol.* **566**, 113–157 (2016).
49. Haupt, M. et al. Preliminary neutron crystallographic study of human transthyretin. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **67**, 1428–1431 (2011).
50. Blakeley, M. P. et al. Neutron macromolecular crystallography with LADI-III. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 1198–1205 (2010).
51. Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D. & Wilkinson, C. LAUEGEN version 6.0 and INTLDM. *J. Appl. Crystallogr.* **31**, 496–502 (1998).
52. Artz, S., Campbell, J. W., Harding, M. M., Hao, Q. & Helliwell, J. R. LSCALE – the new normalization, scaling and absorption correction program in the Daresbury Laue software suite. *J. Appl. Crystallogr.* **32**, 554–562 (1999).
53. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **62**, 72–82 (2006).
54. Heinz Maier-Leibnitz Zentrum, Technische Universität München & Forschungszentrum Jülich. BIODIFF: Diffractometer for large unit cells. *J. large-scale Res. Facil.* **1**, 1–4 (2015).
55. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
56. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **67**, 235–242 (2011).
57. McCarthy, A.A. et al. ID30B – a versatile beamline for macromolecular crystallography experiments at the ESRF. *J. Synchr. Rad.* **25**, 1249 (2018).
58. Kabsch, W. Xds. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 125–132 (2010).
59. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **68**, 352–367 (2012).
60. Adams, P. D. et al. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 213–221 (2010).
61. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **66**, 486–501 (2010).
62. Sobott, F., Hernández, H., McCammon, M. G., Tito, M. A. & Robinson, C. V. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. *Anal. Chem.* **74**, 1402–1407 (2002).
63. van den Heuvel, R. H. H. et al. Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78**, 7473–7483 (2006).
64. Salvatier, J., Wiecki, T. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55 (2016).
65. Abraham, M. J. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).
66. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
67. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
68. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
69. Feenstra, K. A., Hess, B. & Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* **20**, 786–798 (1999).
70. Hess, B. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
71. Essmann, U. et al. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
72. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
73. McGibbon, R. T. et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
74. Gapsys, V., Seeliger, D. & De Groot, B. L. New soft-core potential function for molecular dynamics based alchemical free energy calculations. *J. Chem. Theory Comput.* **8**, 2373–2382 (2012).
75. Gapsys, V., Michielsens, S., Seeliger, D. & de Groot, B. L. pmx: automated protein structure and topology generation for alchemical perturbations. *J. Comput. Chem.* **36**, 348–354 (2015).
76. Goga, N., Rzepiela, A. J., de Vries, A. H., Marrink, S. J. & Berendsen, H. J. C. Efficient algorithms for Langevin and DPD dynamics. *J. Chem. Theory Comput.* **8**, 3637–3649 (2012).
77. Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).
78. Ross, G. A., Bruce Macdonald, H. E., Cave-Ayland, C., Cabedo Martinez, A. I. & Essex, J. W. Replica-exchange and standard state binding free energies with Grand Canonical Monte Carlo. *J. Chem. Theory Comput.* **13**, 6373–6381 (2017).
79. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
80. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

## Acknowledgements

V.T.F. acknowledges support from the EPSRC under grant numbers GR/R99393/01 and EP/C015452/1 which funded the creation of the Deuteration Laboratory (D-Lab) in the Life Sciences group of the ILL. The authors thank the ILL for the provision of beamtime (proposals 8-01-428, 8-01-440, 8-01-463) on the LADI-III diffractometer. This work used the MS platform of the Grenoble Instruct Centre (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with the support from FRISBI (grant number ANR-10-INSB-05-02) and GRAL (grant number ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). A.W.Y. was supported by a PhD studentship from ILL and Keele University. M.A. is supported by a Research Fellowship of the Alexander von Humboldt Foundation. M.A. thanks Vytautas Gapsys for providing technical support on the free energy calculations.

## Author contributions

A.W.Y. conducted all biochemical, neutron and X-ray crystallography, and native MS studies, including protein purification, sample preparation, data analysis and interpretation. M.A. designed, performed, analysed and interpreted all computer simulations, and performed the fitting of the MS data. E.B.E. performed native MS data collection. M.P.B. and A.O. performed neutron diffraction data collection and processing. M.W.B., C.M.-D. and D.d.S. performed X-ray diffraction data collection. M.M. and P.J.M. assisted in the initial stage of this project. E.P.M., M.H., V.T.F. and B.L.d.G. designed and supervised the execution of experiments in this manuscript and interpreted the results. A.W.Y. and M.A. wrote the manuscript; all authors reviewed the final manuscript.

## Additional information


**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-08609-z>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks Andrey Kovalevsky and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Crown 2019