

*Supporting Information for*

## **Crystal structure of thioflavin-T and its binding to amyloid fibrils: insights and implications at the molecular level**

*Cristina Rodríguez-Rodríguez, Albert Rimola, Luis Rodríguez-Santiago, Piero Ugliengo, Ángel Álvarez-Larena, Hugo Gutiérrez-de-Terán, Mariona Sodupe,\* Pilar González-Duarte\**

### **1. Experimental details**

All reagents were purchased from Aldrich, the commercial ThT(Cl) was recrystallized in acetonitrile. The slow addition of an aqueous saturated solution of KI (5 ml) to an aqueous solution (20 ml) of (ThT)Cl (0.600 g, 1.88 mmol) resulted in the formation of a yellow solid that was filtered off, washed with cold water (10 ml) and dried *in vacuo*. MS(+ESI-MS) showed the presence of ThT<sup>+</sup> exclusively. *m/z* (relative intensity): 283.1 ([ThT]<sup>+</sup>, 100).

A saturated CHCl<sub>3</sub> solution of this solid was allowed to evaporate slowly at room temperature under open atmosphere. Reddish crystals, (ThT)<sub>2</sub>I<sub>4</sub>·2CHCl<sub>3</sub> (**1**), were obtained after 24 h and yellow crystals, (ThT)I·CHCl<sub>3</sub> (**2**), after ca. one week, both types of crystals being suitable for X-ray diffraction.

### **2. X-ray diffraction**

Single-crystal X-ray diffraction analyses were performed on a Bruker SMART-APEX CCD area-detector diffractometer at room temperature with graphite-monochromated MoK $\alpha$  radiation. Lorentz-polarization and absorption corrections were applied using Bruker SAINT<sup>[1]</sup> and SADABS<sup>[2]</sup> software. Structures were solved by direct methods and refined by full-matrix least-squares on F<sup>2</sup> for all reflections using SHELXTL.<sup>[3]</sup> In **1**, the chloroform molecule is highly disordered. Four geometrically constrained CHCl<sub>3</sub> groups have been considered and refined with isotropic displacement parameters. The rest of non-hydrogen atoms in the structures **1** and **2** were refined with anisotropic displacement parameters. Hydrogen atoms were included with riding model constraints and isotropic displacement parameters 1.2 times U<sub>eq</sub> value of the corresponding carbons. The two crystal structures have been deposited at the Cambridge Crystallographic Data Centre under the following deposition numbers: (ThT)<sub>2</sub>I<sub>4</sub>·2CHCl<sub>3</sub> (CCDC 728298) and (ThT)I·CHCl<sub>3</sub> (CCDC 728299).

### **3. Computational details**

#### **3.1. DFT calculations**

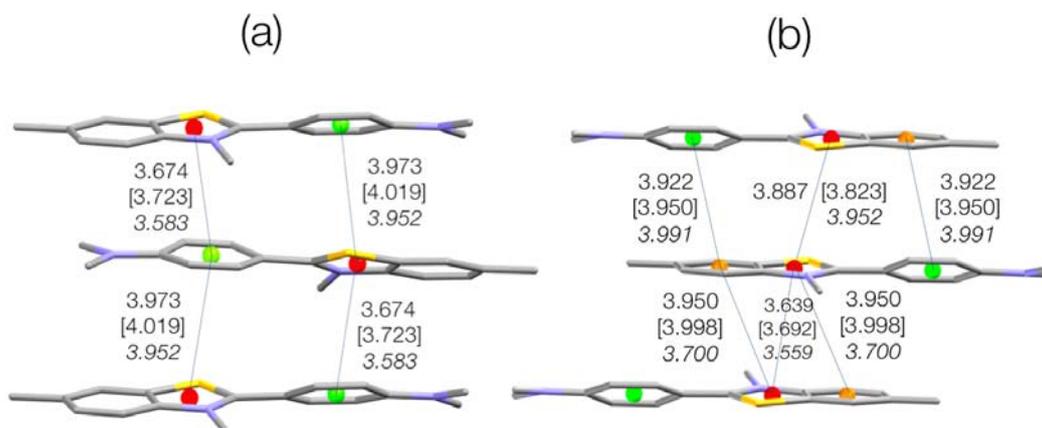
Gas phase calculations for (ThT)<sup>+</sup> have been carried out at the B3LYP<sup>[4, 5]</sup> level of theory with the 6-31++G(d,p) basis set. In the case of ThT<sup>+</sup> interacting with a model of the  $\beta$ -sheet structure optimizations have been performed at the B3LYP-D level; that is, adding an empirical correction for dispersion of the form  $-C_6 \cdot R^{-6}$  ( $s_6=1.05$ )<sup>[6]</sup> to the B3LYP energy. In this case, and due to the size of the system, we used the 6-31G(d) basis set for geometry optimizations and the 6-31+G(d,p) basis for single point energy calculations. This procedure was calibrated for one of the minimum located ( $\varphi = 21$ ) and found to be very

effective. All gas phase calculations were carried out with the Gaussian03 program package.<sup>[7]</sup> Grimme's dispersion term and gradients were programmed in an external driver.

Calculations addressed to compute the crystal structures **1** and **2** were carried out using the latest version of the periodic ab-initio code CRYSTAL06.<sup>[8]</sup> This code adopts a local Gaussian basis set, in which the outer shell for each atom belonging to the present system are explicitly given, in Bohr<sup>-2</sup> (more details Details of the adopted Gaussian basis set are available on the CRYSTAL web site<sup>[9]</sup>): H, 31G\* ( $\alpha_s = 0.16$ ,  $\alpha_p = 1.1$ ); C, 6-31G\* ( $\alpha_{sp} = 0.17$ ,  $\alpha_d = 0.80$ ); N, 6-31G\* ( $\alpha_{sp} = 0.21$ ,  $\alpha_d = 0.80$ ); S, 86-311G\* ( $\alpha_{sp} = 0.106$ ,  $\alpha_d = 0.383$ ); Cl, 86-311G ( $\alpha_{sp} = 0.323$ ,  $\alpha_d = 0.125$ ); I, seven-valence-electron pseudopotentials derived from the scalar-relativistic energy-consistent variety to represent the innermost electrons, whereas an uncontracted [4s4p] atom-optimized basis set ( $\alpha_s = 0.122$ ,  $\alpha_p = 0.101$ ) for the 7 outermost electrons and valence orbitals. The hybrid B3YLP<sup>[4, 5]</sup> density functional was used for all periodic calculations. The Hamiltonian matrix was diagonalized in 10k points, corresponding to a shrinking factor of the reciprocal space of 3.<sup>[10]</sup> Values of  $10^{-6}$ ,  $10^{-6}$ ,  $10^{-6}$ ,  $10^{-6}$ ,  $10^{-14}$  for the tolerances controlling Coulomb and exchange series were adopted for all calculations. The condition for the SCF convergence was set to  $10^{-7}$  on the energy difference between two subsequent cycles.

Geometry optimizations were performed at the B3LYP level by means of a quasi-Newton algorithm, combining the quadratic step (BFGS Hessian updating scheme) with a linear one (parabolic fit), as proposed by Schlegel.<sup>[10]</sup> It should be noted that only the internal degrees of freedom were allowed to relax, the unit cell parameters for structures (**1**) and (**2**) being kept fixed at the experimental values. In principle, the B3LYP-D level of calculation (i.e. including dispersive interactions) should have been adopted to fully optimize both the cell parameters and the internal coordinates for both crystal structures. This approach has, however, a number of flaws as described recently for molecular crystals.<sup>[11]</sup> This is particularly noticeably in the present systems due to the complexity of the intermolecular potential energy surface arising from packing forces of different nature that need to be very well balanced and described by the level of theory used. On the other hand, both crystal structures are characterized by a large number of independent atoms and by low crystal symmetries, which entails a huge computational effort. In addition, the size of the system forced us to adopt a gaussian basis set of moderate flexibility, which means that both structures (and in particular the cell size controlling the crystal packing) and interaction energies can be seriously affected by basis set superposition error (BSSE). When this latter is not properly accounted for it tends to behave as a spurious attractive term, mimicking in some way the missed dispersive component of a pure B3LYP calculation. The role of dispersion in modulating the intramolecular degrees of freedom can be evaluated by performing a B3LYP-D optimization limited to the internal degrees of freedom and keeping the cell parameters fixed at the experimental values. However, this strategy does not always provide the best results as compared to the experimental data due to limitations on the basis sets used in periodic calculations. To illustrate this point, in the following Figure S0 some key intermolecular distances between stacked ThT<sup>+</sup> as resulting from B3LYP and B3LYP-D internal coordinates optimization are reported along with the experimental values. Clearly, B3LYP values are in better agreement with experiment than B3LYP-D ones, this latter giving too short inter-ring distances. This is again understandable in terms of the BSSE which tends to add an extra spurious attractive term to the B3LYP-D energy bringing molecules too close to each other. For the above reasons,

the B3LYP structures, optimized keeping the cell parameters fixed to that derived from diffraction data, are the ones reported and discussed in the main text.



**Figure S0.** Distances between centroids (in Å) obtained from X-ray experimental data, periodic B3LYP [in brackets] and periodic B3LYP-D in *italics*.

DFT calculations addressed to compute the binding of  $\text{ThT}^+$  with  $\beta$ -amyloid were carried out assuming five strands of the model peptide  $\text{CH}_3(\text{NHCOCH}_2)_3\text{NHCOCH}_3$ . Starting geometry was taken from docking results, by replacing Met35 and Val33 residues were replaced by Gly to simulate only the interaction with the polypeptide chain and thus, the model used represents both external poses (see below). In order to avoid artificial fibril distortions due to the small size of the model, the terminal carbon atoms in both ends of each strand were kept fixed during the optimization process. Calculations were done both at the B3LYP and B3LYP-D levels of theory to analyze the role of dispersion in the bound complex. The basis set used in these calculations is 6-31+G(d,p).

### 3.2. Docking protocol

Each of the 10 protein models of the  $\beta$ -amiloid peptide deposited under the PDB code 2BEG was used as a rigid protein model for an independent blind docking analysis. We increased default Autodock parameters to obtain 2.5 million Lamarckian Genetic Algorithm (LGA) evaluations, with a population size of 150, as recommended for blind docking.<sup>[12]</sup> A box was defined to cover all the observed protein binding crevices, with dimensions 92x70x60 Å, centered on the center of mass of the protein and using a 0.375 Å grid step. 100 docking runs were conducted on each of the 10 protein models, thus leading to a final pool of 1000 possible solutions. In a first analysis, we identified the most reliable solutions based on scoring and population criteria for each protein model independently. In a second step, all solutions (1000) were reclustered attending a 2 Å RMSD criteria. Each cluster with a minimum population of 50 individuals was represented by the best protein-ligand complex, according to the scoring function and further refined by energy minimizations in NAMD<sup>[13]</sup> using the OPLS force field<sup>[14]</sup> with manual assignment for the missing parameters. Explicit solvent was considered with a sphere of TIP3 water molecules that ensured enough solvation for the protein.

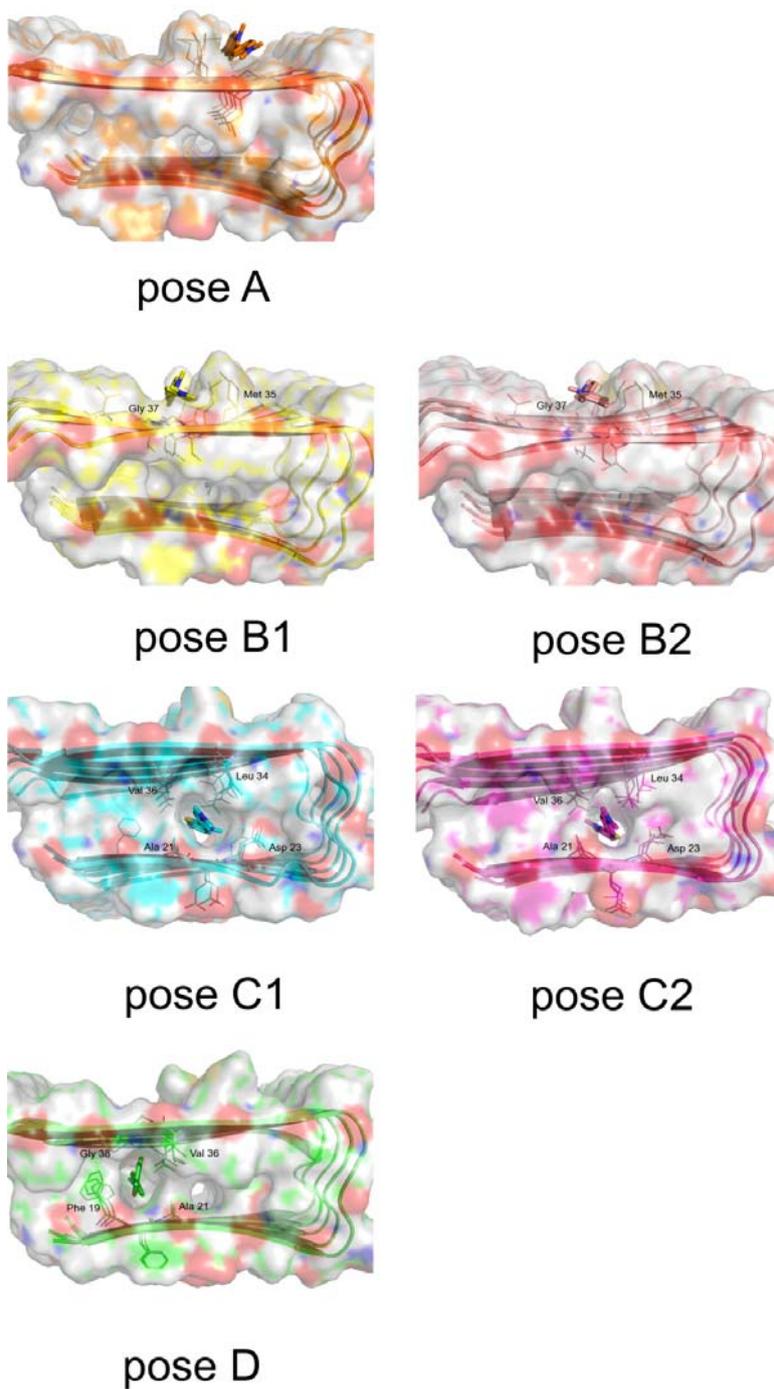
### 3.3. Molecular Dynamics and binding free energy calculations

Relative binding affinities for the binding poses selected from the docking protocol were calculated using the linear interaction energy (LIE) method, described in detail elsewhere.<sup>[15, 16]</sup> Originally this approach estimates the absolute ligand free energy of binding from the difference in the ligand – surrounding interaction energies in both its bound and free state. In our case, since only relative binding affinities between different docking poses were attempted, we will substitute the “free state” by a “reference bound state” (e.g. the state with lowest free energy as determined with LIE calculations using the free state in water as the reference state). The relationship between the ligand intermolecular interaction energies and the (relative) free energy of binding is given by the equation:

$$\Delta G_{bind} = \alpha \Delta \langle V_{l-s}^{vdW} \rangle + \beta \Delta \langle V_{l-s}^{el} \rangle + \gamma$$

where  $V_{l-s}^{vdW}$  and  $V_{l-s}^{el}$  denote, respectively, the Lennard-Jones and electrostatic interactions between the ligand and its surroundings ( $l-s$ ). These interactions are evaluated as energy averages (denoted by the broken brackets) from separate MD simulations of the two bound states to be compared. The difference ( $\Delta$ ) between such averages for each type of potential is scaled by different coefficients<sup>[16]</sup> giving the polar and non-polar contributions to the binding free energy. For the non-polar contribution, we used the empirical value of  $\alpha = 0.181$ , while for the polar contribution, the scaling factor follows the linear response approximation ( $\beta = 0.5$ ) since the ligand holds a net positive charge.

MD simulations were done using the program Q<sup>[17]</sup> and the OPLS force field there implemented.<sup>[14]</sup> Each system was solvated with a simulation sphere of TIP3P waters<sup>[18]</sup> of radius 20 Å, centered on the central atom of the ligand. The water surface of this sphere was subjected to radial and polarization restraints<sup>[19]</sup> in order to mimic bulk water at the sphere boundary. Ionizable residues in the inner solvation sphere were modeled as charged, while residues close to the boundary of this sphere were considered in their neutral form, except if they form salt bridges. The total charge of the sphere was in all cases equal to the net charge of the ligand (+1). The electrostatic interaction energy between the ligand and the neglected charges of the protein was approximated from the initial structure by Coulomb's law with a high dielectric constant. Non-bonded interaction energies were calculated up to a 10 Å cutoff, except for the ligand atoms for which no cutoff was applied. Beyond the cutoff, long-range electrostatics were treated with the local reaction field (LRF) multipole expansion method.<sup>[20]</sup> Protein atoms outside the simulation sphere were restrained to their initial positions, and only interacted with the system through bonds, angles and torsions. A heating and equilibration procedure was applied before the data collection phase. The equilibration protocol started with 1000 steps MD using very short time step (0.2 fs) at 1 K temperature, coupled to a strong bath (0.2 fs bath coupling) with positional restraints on heavy atoms. Then the system was gradually heated up to 300 K, relaxing the bath coupling to 100 fs and increasing the timestep to 1 fs, while the positional restraints were smoothly released. This equilibration phase (100 ps) was followed by 500 ps of unrestrained MD before data collection, collecting energies at regular intervals of 15 fs. Energy averaging was performed on the energetically stable phase of the collection period, never shorter than 250 ps, where stability was addressed by comparing the average binding free energy values of the first and second halves of the data collection period.



**Figure S1.** Docking poses found for ThT<sup>+</sup> after interaction with the fibrillar core structure of A $\beta$ <sub>1-42</sub> peptide (PDB 2BEG) according to the automated docking exploration described in the text.

## References

- [1] SAINT v.6.22, Bruker AXS Inc., Madison, WI, USA, **2001**.
- [2] G. M. Sheldrick, SADABS v. 2.03, University of Göttingen, Germany, **2002**.
- [3] G. M. Sheldrick, SHELXTL v.6.10, Bruker AXS Inc., Madison, WI, USA, **2000**.
- [4] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648.
- [5] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B.* **1988**, *37*, 785.
- [6] S. Grimme, *J. Comp. Chem.* **2006**, *27*, 1787.
- [7] Gaussian03. Rev. E,01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. V. Jr., K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian Inc., Wallingford CT, **2004**.
- [8] R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D'Arco, M. Llunell, University of Torino, Torino, **2006**.
- [9] [http://www.crystal.unito.it/Basis\\_Sets/Ptable.html](http://www.crystal.unito.it/Basis_Sets/Ptable.html), Torino, **2005**.
- [10] H. B. Schlegel, *J. Comp. Chem.* **1982**, *3*, 214.
- [11] B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano, P. Ugliengo, *CrystEngComm* **2008**, *10*, 405.
- [12] C. Hetenyi, D. van der Spoel, *FEBS Lett.* **2006**, *580*, 1447.
- [13] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
- [14] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- [15] J. Aqvist, C. Medina, J. E. Samuelsson, *Protein Eng.* **1994**, *7*, 385.
- [16] T. Hansson, J. Marelus, J. Aqvist, *J. Comput. Aided Mol. Des.* **1998**, *12*, 27.
- [17] J. Marelus, K. Kolmodin, I. Feierberg, J. Aqvist, *J. Mol. Graph. Model.* **1998**, *16*, 213.
- [18] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [19] G. King, A. Warshel, *J. Chem. Phys.* **1989**, *91*, 3647.
- [20] F. S. Lee, A. Warshel, *J. Chem. Phys.* **1992**, *97*, 3100.