

Targeting different transthyretin binding sites with unusual natural compounds

Dr. Gabriella Ortore*, Prof. Elisabetta Orlandini, Prof. Alessandra Braca, Dr. Lidia Ciccone, Prof. Armando Rossello, Prof. Adriano Martinelli, Dr. Susanna Nencetti*

Dr. G. Ortore, Prof. E. Orlandini, Prof. A. Braca, Dr. L. Ciccone, Prof. A. Rossello, Prof. A. Martinelli, Dr. S. Nencetti**

Dipartimento di Farmacia

Università di Pisa, V. Bonanno 6, 56126 Pisa

*E-mail: gabriella.ortore@unipi.it
susanna.nencetti@unipi.it*

Abstract

Some nutraceuticals, such as flavonoids and natural polyphenols, have recently been investigated as modulators of the self-assembly process of transthyretin (TTR), but they suffer from generally limited bioavailability. To discover new innovative and more bioavailable natural compounds able to inhibit the TTR amyloids, the TTR crystallographic structure through a docking study was explored. The computational strategy was projected *ad-hoc* to inspect the possible relationship between binding site location and modulation of the assembly process: the interaction with the until now unexplored Epigallocatechin gallate (EGCG) sites and with the Thyroxine (T4) pocket were simultaneously analyzed. All the compounds studied seem to prefer the traditional T4 binding site, but some interesting results emerged from the screening of an *in-house* database, used for validating the computational protocol, and of the “Herbal Ingredients’ Targets” (HIT) catalogue available on the ZINC database.

Introduction

Amyloidosis is a group of diseases sharing common pathogenic mechanisms, depending on the deposition of characteristic rigid fibrillar material (amyloids) in various organs and tissues. As these deposits build up, they begin to damage the structure and affect organ function. Systemic amyloidosis has been classified into three very different major types: primary (AL), caused by the accumulation of monoclonal immunoglobulin light chains as amyloid fibrils; secondary (AA), associated with chronic inflammatory diseases (e.g. rheumatoid arthritis); TTR amyloidosis (ATTR), non-hereditary due to fibrilization of wild-type TTR and hereditary, due to a mutation in the gene encoding the plasma TTR protein. ^[1]

TTR is a homo-tetrameric protein, made up of four 127-aminoacid subunits rich in β -sheets, which is synthesized mainly in the liver and the choroid plexus of the brain, and is present both in human plasma and in cerebrospinal fluid (CSF). TTR is involved in the transport of thyroxine (T4) in CSF as well as being a carrier of T4 and retinol in plasma with the assistance of the retinol binding protein. TTR circulates normally as an innocuous soluble protein but in some individuals it polymerizes to form amyloid fibrils, inducing many diseases ^[2,3] like senile systemic amyloidosis (SSA), familial amyloid polyneuropathies (FAP), familial amyloid cardiomyopathy (FAC) and central nervous system selective amyloidosis (CNSA). These pathologies induce severe, life-threatening conditions, which can produce lethal consequences. In 1996 it was observed that amyloid fibril formation is prevented by the T4 binding; ^[4] unfortunately hormonal activity of thyroxine or their analogues makes its application an undesirable therapy.

The mechanism of TTR amyloid fibril formation is not completely clarified, but the amyloidogenic potential of the protein might be related to its extensive β -sheet structure. Several observations and studies ^[5] seem to indicate that the first step consists of a conformational change, which induces the

dissociation of the tetramer, followed by the formation of an intermediate monomer, that at the end leads to self-assembly into insoluble amyloids.

In the last decade some authors ^[6-9] discovered that several structurally different families of small molecules stabilize the tetramer of TTR, and more than 200 crystal structures of TTR complexes with more than 100 different ligands were deposited on the PDB database. Among these molecules the most interesting include several nonsteroidal anti-inflammatory drugs (NSAIDs), estrogen analogues, natural products or synthetic carboxylates that inhibit the TTR fibril formation. All these compounds bind the T4 binding site, at the TTR dimers interface. Only one crystal structure, the one co-crystallized in 2010 with epigallocatechine gallate (EGCG), highlighted a second bind site location. In spite of their potency against TTR amyloidosis, NSAIDs and estrogens have significant undesired effects due to long-term administrations. Only two TTR tetramer stabilizers are, at present, available for TTR amyloidosis therapy: tafamidis and diflunisal. Tafamidis has been approved for the treatment of FAP in European countries and Japan. Although tafamidis binds the TTR tetramer more tightly, diflunisal overcomes weaker TTR binding coefficients with high serum drug concentrations. Gene therapies with antisense oligonucleotides and small interfering RNAs are promising strategies to block TTR synthesis and are currently in clinical trials. In addition, some alternative mechanisms of amyloid inhibition are under study. ^[10] Recently, a strategy was suggested thanks to the TauRX Therapeutics Ltd studies for the Rember® trials in Alzheimer's Disease: the acceleration of the transition from oligomers to fibrils, which are less toxic than oligomers and protofibrils. ^[11] A second alternative strategy is to disrupt the self-association process itself, modulating the assembly process into the formation of nontoxic structures; the first evidence of this additional mechanism in TTR inhibition was referred to epigallocatechin. ^[12] Some authors ^[13] suggest that there could be a correlation between the different binding sites of EGCG in TTR and its alternative mechanism of fibrils inhibition. The alternative binding sites might be involved in the oligomerization of TTR tetramers,

leading to the formation of nontoxic, off-pathway aggregates. In addition to EGCG, some nutraceuticals such as flavonoids, natural polyphenols and caffeic acid, can modulate the self-assembly process,^[10,14] but suffer from generally limited bioavailability. To discover new innovative and more bioavailable natural compounds able to inhibit the TTR amyloids, a computational study using the TTR three-dimensional structure was performed. Given the possible relationship between the binding site location and modulation of the assembly process, the interaction with the until now unexplored EGCG sites and with the T4 pocket was explored, simultaneously. All the potential ligands here studied seem to prefer the traditional T4 binding site, but some interesting results due to the variety and novelty of the given natural scaffolds emerged from the screening of an *in-house* database, used for validating the computational protocol, and of the “Herbal Ingredients’ Targets” (HIT) catalogue available on the ZINC database.^[15]

Results

Docking

The small *in house* database of hydroxylated natural compounds reported in Chart 1 was added to the HIT database in order to explore the TTR binding sites.

The docking results provide two different kinds of information about the interactions with TTR of the natural compounds here studied. Firstly, the disposition of the potential new inhibitors in the protein, targeting different binding sites: the traditional T4 pocket and the new EGCG sites (Figure 1).

The *blind docking* procedure allowed us to explore the whole protein in a range of 30 Å from Ser23, a residue rather central within the known binding sites. The preferred docking position for all compounds was automatically calculated as described in the Experimental Section, and the results are here summarized. Only 92 within the 802 HIT compounds lay outside the T4 binding pocket, and within them, from a graphical analysis, 23 lay in the primary EGCG binding site. Among them, none interacted as EGCG between two adjacent monomers, as reported in Figure 2a, with three “anchors” pointing to Asp18 (right), TyrD114 (up) and SerA85 (left).

Some occupied the monomers’ interface but could not reach Asp18 and TyrD114 (Figure 2b); some reached both Asp18 and TyrD114, but exposed lipophilic groups towards Asp18 (Figure 2c); others left empty the Asp18 and SerA85 pockets, reaching TyrD114 through lipophilic or aromatic moieties (Figure 2d); and in some cases compounds lay on the surface of only one monomer, scarcely fitting both the up and down interfaces (Figure 2e). Only three compounds filled all the pockets (Figure 2f), but engaged weaker interactions with respect to EGCG. Also within the *in-house* database only 2 compounds seemed to prefer the primary EGCG binding site, bilobetin **10** and isoginkgetin **11**, without any interaction with TyrD114. Compound **10**, through its distal hydroxyl group, interacted

with Asp18, but the whole molecule inserted itself in the same monomer, having only a phenol in common with EGCG disposition (Figure 3a). All the other compounds lay in the T4 binding pocket.

As regards the efficacy of the disposition in each binding site, the docking results were analyzed in view of the interaction with key residues, as described in the Experimental Section. These compounds were further analyzed on the basis of their docking score; in Table 1 are reported the values relative to the *in-house* database.

As reported in Figure 4, some compounds showed interesting interactions, in spite of their structural diversity with respect to the usual inhibitors.

In Figure 4a are reported compounds which showed the classical docking pose of the flavones: the poses of 7,4'-dihydroxyflavone **6**, **7** (ZINC03871576: apigenin), **8** (ZINC03869685: quercetin), **9** (ZINC03874317: myricetin), **15** (liquiritigenin) and **16** (ZINC00156701: naringenin) were corroborated by a docking score higher than 22. Also the flavanols **12** (ZINC00119983: catechin) and **14** (fisetidinol) allowed a good interaction with Lys15 and Ser117 (see Figure 4b), with scores lower than but comparable to the flavones ones, in spite of the lack of the carbonyl group. Different with regard to the scaffolds but similar for some pharmacophoric portions, alloperiplogenin **21** (Figure 4c) and pinoresinol **23** (Figure 4d) permitted polar interactions with Lys15 and Ser117, less directional with respect to flavones but compensated through good steric relations with the binding pocket, which incremented the docking score. The hinge structures of chlorogenic acid **17**, neochlorogenic acid **18** (Figure 4e) and 5-*O*-galloylquinic acid **19** (Figure 4f), were able to fit into the T4 binding pocket with lower docking scores, but higher significance because of their unusual scaffolds. In particular compound **19** seemed to interact strongly with Lys15 through its quinic acid, and with Ser117 thanks to its gallic moiety. More complex, and very uncommon as a potential fibrils inhibitor, was the iridoid **24**, 6-*O*-(*p*-hydroxybenzoyl)-6-epiaucubin (Figure 4g). It was very interesting for the presence of a glucose moiety, that is frequently related to a decrease of TTR fibrils inhibition due to the

correspondent aglycones; ^[21] furthermore, it was very remarkable that the sugar portion lay in the inner part of the T4 binding site, in contrast with the flavonoid glycosides present in the HIT database and docked in the same way (data not shown). The hydroxybenzoyl moiety of **24** bound both monomers through Lys15, and the glucose one *via* Ser117; however, also in this case, the docking scores were lower than the flavones ones. Unfortunately, the aglycone of compound **24** was not available for a comparison. The evaluation of the sugar moiety influence was possible for the two neolignans urolignoside **3** and glochidiobioside **4** of the *in-house* database (Figure 4h). Also in this case, the glucose of **4** entered into the inner part of the binding site interacting with Ser117, and its aglycone portion was superimposed on the docking pose of compound **3**, occupying the region towards Lys15. The presence of glucose seemed to ameliorate the interaction with Ser117, but the docking scores were very similar, specifically compound **4** was slightly worse.

A separate description was due to **25** (rosmarinic acid), present in the HIT database as ZINC00899870, and already known as an inhibitor of V30M mutant TTR fibril formation. ^[20] As reported in Figure 4i, the docking pose calculated in the *wt*-TTR (sky-blue) was very similar to the experimental one (green, from PDB complex of V30M-TTR/ROA, code 4PWI). The good interaction was supported through a very high docking score.

Bilobetin **10**, present in the HIT database as ZINC03979028, as already described preferred the EGCG binding site. But the score difference between the EGCG-pose and the T4-pose was less than 0.5. In Figure 3 a comparison between the two binding modes was reported. Although it had a good docking score, the fitting with the EGCG binding site was not superposed on the experimental one: it seemed to occupy a border region. So, the docking into the T4 binding site was also analyzed: compound **10** inserted only the apigenine moiety into the pocket, slightly distorted by the bulky 8-substitution which lay outside the pocket.

Virtual Screening

The docking results were ranked as described in the Experimental Section. Few compounds seemed to prefer binding with the EGCG site, and none seemed to fit effectively this region. So, the Virtual Screening (VS) was performed targeting the T4 binding pocket. The 710 HIT compounds preferring the T4 binding site were ranked on the basis of the key interactions with Lys15 and Ser117. They included apigenin **7**, quercetin **8**, myricetin **9**, catechin **12**, naringenin **16** and rosmarinic acid **25**, available also in the *in house* database, and our compounds **1-6**, **10-11**, **13-15**, **17-24**. Bilobetin **10** was included despite its best score in EGCG because of its doubting behaviour. This enriched database included 14 actives and 5 inactives (see experimental section for details). Among the HIT compounds, 293 interacted with both the lysines A15 and C15, and only a set of 119 compounds with at least one serine 117. All the *in house* compounds engaged interactions with the key residues. The 138 total surviving compounds were ranked on the basis of the docking score, and only scores higher than 20 were accepted. In spite of the claimed imperfections and approximations of the scoring functions,^[22] the use of the docking scores together with other computational methods, as in this case the interactions with key residues, allows good results in the structure-based design.^[23] The ROC curve (Figure 5) shows that all the actives survived filtering and only 1 inactive compound was able to pass with the 88 hit compounds. In this VS, Recall was = 1, Precision = 0.93 and EF = 8.74; all these values indicated a good predictivity.

For clarity, a schematic representation of the VS is reported in Figure 6. The surviving *in-house* database compounds are ranked in Table 2. Among these 15 compounds, 6 were already known as TTR fibril inhibitors. Their potency was not comparable because of the different methods for testing the effect on amyloid production. In some cases, such as for rosmarinic acid, the biological test was not performed on the wt-TTR. Within the 9 new potential inhibitors, some compounds were selected to test their ability to inhibit the fibril formation through a turbidimetric assay. The criterion used for

the selection was novelty in the molecule structure, thus liquiritigenin **15** was discarded for its similarity to the known naringenin. Between the enantiomers periplogenin **20** and alloperiplogenin **21** was selected the best scored **21** and, between **12** and **14**, the catechin **12**. Furthermore, pinoresinol **23**, compounds **3** and **4**, very interesting for a comparison between an aglycone and its glycoside, and bilobetin **10**, with its ambiguous and intriguing localization in the protein, were designated for the turbidimetric test. Rosmarinic acid **25**, very good for interactions and docking score, was also tested to confirm its activity in non-amyloidogenic TTR.

In spite of the lower docking scores, also compounds **19** and **24**, for their structure novelty and significance of their docking pose, were selected for testing.

Biological assay

The natural compounds selected from the small *in house* library were tested for the inhibition of TTR fibril formation in an acid mediated turbidimetric assay. Figure 7 shows the percentage of TTR fibril formation (FF) evaluated by a doublet concentration of selected natural compounds to determine their potential as TTR fibril formation inhibitors. The concentration used in the tests (7.2 μM) represents twice the TTR concentration in plasma (3.6 μM). Diflunisal (9% FF) was used as the reference drug (positive control) while quercetin (11%), a known TTR inhibitor, as the reference natural compound. TTR without inhibitor under acid conditions (pH 4.2) was used as a negative control to determine the 100% FF mark. ^[17]

Analysis of the turbidimetric assay-results indicates that all the tested natural compounds were able to inhibit the fibrillogenesis with a percentage of fibril formation ranging from 15% to 73%; in particular, compound **25** with only 15% of fibril formation is the best inhibitor. The other natural compounds, **3**, **4**, **10**, **12**, **19**, **21** and **23**, showed a similar trend with an inhibitory activity around a value of 50% FF.

Discussion

The possible role of natural compounds in the TTR amyloidoses is investigated by many research groups in view of their ability to impair TTR aggregation through different modes of action. Different compounds could suppress TTR tetramer dissociation, promote TTR oligomerization into small “off-pathway” nontoxic aggregates, or reduce the amount of TTR aggregation. ^[19] In this context, we theoretically explored the *wt*-TTR structure to test novel natural potential inhibitors, targeting different sites of interaction and, therefore, presumable different modes of action. The docking procedure here described allowed us to discriminate which is the preferred site of interaction, for 820 natural compounds, in a range of 30 Å from a strategic centre of all the possible reported TTR binding sites. The searching of docking sites alternative to the T4 ones was successful only for 92 compounds, mostly concentrated near the primary EGCG site. Considering that this site lies on the surface of the protein, allowing a high level of freedom in the docking disposition, only 23 compounds fit into the RX ligand region and only three of them interact with the three EGCG “anchors”. However, it is interesting to highlight that compounds such as podophyllotoxin ZINC01532024, ginkgolides (ZINC04245650, ZINC08552016, ZINC08552017, ZINC95098816, ZINC95098818, ZINC95098819, ZINC95098830) and schisandrin ZINC02000967 (see Figure 2), are known or under study as protectors against A β toxicity. ^[24-27] Further studies are needed to examine in depth the role of ZINC05369365 analogues, and some assays are already in progress.

The ambiguous bilobetin **10**, which slightly prefers an alternative docking on the TTR surface interacting with Asp18, but with a comparable score interacts in the T4 pocket, allowed the reduction of the fibril formation to about 50%. For this compound, the mechanism of inhibition could be not clear, given that the preferred binding site is unclear and the turbidimetric test could not give information about the inhibition process.

The docking results revealed, indeed, that the preferred binding site for the about 730 virtual screened compounds was the T4 pocket. This is compatible with the crystallographic trend: within the about 120 ligands co-crystallized with TTR, only one (EGCG) shows a binding site different from the T4 ones. The results will be discussed from two different points of view: the efficacy of the computational procedure and its effect on the compounds available in the *in-house* database.

The VS was structured for emphasizing the role of the key interactions in the inhibitory activity. This is not a really essential requirement, because many active compounds have not polar interactions with Lys15 or Ser117 but, in particular within the natural molecules, all the compounds engaging hydrogen bonds with these residues show a strong reduction in TTR fibril formation (see Figure 8).

We can consider this choice a good method for VS if the survived compounds enclose the known actives and discard the inactives. In our case, the HIT database comprises 14 actives and 5 inactives; our VS yields 88 compounds (15 of them available in the *in-house* database), which included the entire set of active compounds and only one inactive. In this context, we have to specify that genistin, and also the luteolin 7-glycosides, pass the screening, too late in the ROC curve. Unfortunately, this is an inescapable limit, due to the real evidence that flavonoid derivatives as glucuronides, considered mimics of glycosides, can bind effectively TTR without preventing the toxic response.^[21] The crystallographic structure of some glucuronides^[28] showed a good fit of these derivatives with the T4 binding site, inserting the flavonoid into the pocket and placing the sugar outer, near Lys15, exactly as calculated by GOLD for the HIT analogues. Despite the crystallographic binding, they did not significantly compete with radiolabeled T4-TTR,^[28] and had no effect on TTR toxicity.^[21]

Indeed, the VS protocol was able to discriminate known inhibitors and to reject the inactives, with the exception of the flavonoid glycosides which are not compatible with this method. This is validated by good enrichment factor values, and encourage us to further study the 15 hits available in our *in-house* database. Among them, 7,4'-dihydroxyflavone **6**, apigenin **7**, naringenin **16** and myricetin **9** have been

reported by many authors, and their activity on TTR fibril formation was already detected in *wt*-TTR by turbidity measurements. ^[16-17,28] In contrast, rosmarinic acid **25** has been firstly reported as not stabilizing TTR in whole plasma and not competing with [125I]T4 for its binding to TTR; ^[19] later it has been reported as a potent inhibitor of acid-mediated aggregation on the amyloidogenic variant V30M-TTR. ^[20] There was no evidence of *wt*-TTR fibril inhibition: our test verified the high activity of rosmarinic acid reducing the *wt*-TTR fibril formation to 15%. Furthermore, the docking pose suggested by GOLD was superimposed on the experimental one, validating the calculations both from a structural and activity perspective.

As regards catechin **12**, no data were available about its activity on TTR until August 2015; in our turbidimetric test catechin allowed a fibril formation inhibition higher than 50%, coherently with the immunodetection of TTR aggregates published on L55P variant during the course of this study. ^[18]

Also alloperiplogenin **21** and pinoresinol **23** halves the TTR fibril formation, very similar to catechin for their docking pose and score. The only steroid analogue already known as a TTR ligand was 16 α -Br-estradiol (4PM1 PDB code); ^[29] compound **21**, having several hydroxyl groups and an α,β -unsaturated lactone ring, docked in a reversed pose in the T4 binding site and seems to slightly better inhibit TTR fibril formation. Pinoresinol **23** was the first lignan tested as TTR fibril formation inhibitor, but the lignan benign functions, such as their antioxidant activity and oestrogenic functions, as well as the protection against certain chronic diseases, are known. ^[30] Lignans are found in a wide range of food such as fruits, vegetables, flaxseed and beverages such as coffee, tea and wine. Less common, urolignoside **3** and its glucoside glochidiobioside **4** are very interesting, also for comparing the effect of the sugar on the aglycone activity. All the cases already reported in literature have shown a loss of TTR fibril inhibition with the insertion of glucose on flavonoids. ^[21,31] Their crystallographic structures show the insertion of flavonoid into the T4 pocket, positioning the sugar outside, towards and over Lys15. ^[21] In our docking, the glucose moiety of **4** interacted with Ser117 in the inner part of

the T4 pocket; the inhibition of TTR fibrils was maintained also transient from aglycone **3** to glucoside **4**. This is the first active glycoside in TTR amyloid inhibition. To further analyze the sugar effect, also 6-*O*-(*p*-hydroxybenzoyl)-6-epiaucubin **24**, whose docking seemed to engage good interactions with key residues despite the low score, being tested. The inhibitory activity was lesser, confirming the score significance in the filtering process of our VS. Finally, 5-*O*-galloylquinic acid **19**, ranked under the VS threshold, was tested as a sample of the quinic derivatives set (**17-19**), all interesting and pharmacophorically related to the oleuropein aglycone very recently reported as an inhibitor of TTR amyloid aggregate cytotoxicity ^[32]. The gallate **19** reduces the TTR fibril formation to 50%, and could be considered a false negative.

All the compounds selected through this VS, indeed, show they inhibit the deposit of TTR fibrils. The potency of new scaffolded inhibitors was not comparable with the flavonoids one, but it is an excellent starting point for lead optimization. Furthermore, the docking orientations reproduce the available crystallographic poses and give further information or suggestions about new possible binding modes of innovative scaffolds.

Conclusions

Starting from the recent results about some nutraceuticals able to modulate the self-assembly process in amyloidosis, the *wt*-TTR structure to search for novel natural potential inhibitors within three-dimensional databases was studied through a computational approach. We targeted, through a focused procedure, different TTR binding sites, with the aim of presumably exploring alternative modes of action suggested by different behaviours of some natural compounds, in inhibiting the TTR amyloids.

^[10,19] The screening of innovative, different scaffolded compounds, could suggest novel and more bioavailable natural compounds as TTR fibril formation inhibitors. The computational procedure was validated through a VS, with good values of enrichment factors. Also the docking poses calculated for

some HIT database compounds, ^[15] already co-crystallized into TTR, were in agreement with the experimental ones. A preliminary test for corroborating our VS was performed on the hit compounds; the turbidimetric results confirm the predictivity of our theoretical study, showing a noticeable inhibitory activity on the fibril formation for all the tested compounds. The structural diversity of the natural compounds here described opens a new perspective in the discovery of nutraceuticals, both in the lead optimization and in the further exploration of their biological role in TTR amyloidosis.

Experimental Section

Chemistry

Compounds **4**, **10**, ^[33] **12**, ^[34] **19**, ^[35] **22**, ^[36] **24** ^[37] and **25** ^[38] were isolated from different plant extracts, as previously reported. Quercetin **8** was obtained from Sigma-Aldrich. Compounds **3** and **23** were purified from chloroform extract of *Periploca graeca* L. stems as following: the CHCl₃ extract was purified by flash silica gel column chromatography as previously reported; ^[36] fraction 5 was purified by RP-HPLC on a C₁₈ μ -Bondapak column (30 cm x 7.8 mm, flow rate 2.0 mL min⁻¹) using MeOH-H₂O 3:2 to yield pure compounds **3** and **23**.

Docking

Automated docking into TTR of the HIT database's ligands, ^[15] enriched with 18 compounds not enclosed in the HIT database but belonging to an *in-house* natural database, reported in Chart 1, was carried out by means of the GOLD 5.1 program. ^[39] The 18 ligands were built using the Maestro program, ^[40] and subjected to a Conformational Search (CS) of 1000 steps in an implicit water solvation model using the MacroModel program. ^[41] The Monte Carlo algorithm was used with the MMFFs forcefield, performing 100 step for rotatable bond with an energy window of 5 Kcal/mol, and eliminating the redundant conformers within a 0,5 Å of cutoff. An automatic setup of the torsion sampling options was performed during the calculation. The ligands were then minimized using the

Conjugated Gradient method to a convergence value of 0.05 kcal/Å·mol, using the same forcefield and parameters as for the CS.

The crystal structure of TTR was selected through a cross-docking that will be described in detail elsewhere, as the structure capable of better reproducing the co-crystal poses of wild-type TTR ligands deposited in Protein Data Bank (PDB). The region of interest of this complex, the 1Y1D PDB coded, ^[42] was defined in three different ways for three different and parallel docking procedures: a) for detecting the traditional T4 binding pocket, b) for exploring the EGCG one's, and c) for extending the docking region over both the binding sites. In the a) case the pocket was defined within 10 Å from the 1Y1D ligand iododiflunisal, allowing the flexibility of Ser117; in b) the EGCG coordinates were derived through the superposition of the 3NG5 coded crystal structure (EGCG-TTR complex) ^[43] on the 1Y1D ones, and the binding site was defined within 10 Å from the translated EGCG. The Arg21 side chain was allowed to adopt both the conformations of 1Y1D and 3NG5 crystal structures during the docking, using the EXTRA PARAMETER option of GOLD. Finally, in the c) procedure the binding site definition was independent from the ligands: it contains all the residues within 30 Å from the C alpha of the residue Ser23. Also in this case the Arg21 side chain was allowed to adopt both the conformations of 1Y1D and 3NG5 crystal structures during the docking. This protocol allows us to perform a blind docking considering all the cavities of the protein, using as a center of the docking cavity a point in the middle between the T4 and EGCG binding sites. Thus, the docking is performed simultaneously in all the known TTR binding sites. The 'allow early termination' command was deactivated in all a), b) and c) protocols. All the ligands were submitted to 40 Genetic Algorithm runs using the Chemscore fitness function, resulted the best fitness function in the previous cross-docking study, and clustering the output orientations on the basis of a RMSD distance of 1.5 Å. The default GOLD parameters were used for all the variables, except for the side chains rotamers. The top scored docking pose for each ligand was then used for further studies.

Analysis of the docking results and ranking of compounds

The graphical analysis of the docking results was performed by Chimera.^[44] They were analyzed from two points of view: the binding site localization (T4 and EGCG binding sites) and the docking goodness in terms of scores and interactions with the key residues of each binding site. The RMSD between the best docking pose produced by blind docking (bd-pose) and the best pose resulting from the docking in the T4 pocket (T4-pose) or in EGCG binding site (EGCG-pose) were evaluated through the *rms analysis* tool in the GOLD utilities. From a preliminary sample analysis was defined a threshold of RMSD between bd/T4-pose values lower than 12Å, which characterized the bd-poses inside the T4 binding site, and lower than 14Å for the RMSD between bd/EGCG-pose, which is typical for bd-poses preferring the EGCG binding site. In addition to the binding site localization, the interaction with both the lysines 15 and almost one of the serines 117 of the T4 binding site, and Asp18 or TyrD114 of EGCG binding site, were investigated. The hydrogen bonds calculation was performed using Chimera, and the results were elaborated through Excel, ranking compounds firstly on the basis of the presence of these interactions. Compounds engaging h-bonds with the key residues were further ranked on the basis of their docking score.

Virtual Screening of HIT + *inhouse* database

The HIT database was analyzed to check known actives as TTR fibrils inhibitors within the 802 compounds. In view of the available data, some EC50 or turbidimetric results are known for 14 actives (7,4'-dihydroxyflavone, myricetin,^[16] catechin, gallic acid, epigallocatechin, epigallocatechin gallate,^[18] curcumin,^[45] genistein, daidzein, apigenin^[28-31], luteolin^[21] and rosmarinic acid^[19]) and 5 inactives (caffeic acid,^[20] estriol, genistin,^[31] luteolin 7-*O*-galactoside,^[21] mangiferin^[46]). Thus, it is possible to virtual screen the database validating the calculation through the enrichment factor method. The threshold value for selecting the possible active compounds was defined analyzing the rate of actives and inactives in the ROC curve (Figure 5): a score value higher than 20 allowed to filter all the

actives and only 1 inactive compound. The VS results were evaluated using Recall, Precision and Enrichment. The Recall value for positives describes the ratio of correctly classified members of a data set, and is defined by the equation $\text{Recall} = \text{tp}/(\text{tp}+\text{fn})$, where tp is the number of active compounds not rejected (true positives) and fn is the number of active compounds rejected during the VS filtering (false negatives). The Precision gives the percentage of observed positives that are correctly predicted, and is defined by the equation $\text{Precision} = \text{tp}/(\text{tp}+\text{fp})$, where fp is the number of inactive compounds not rejected (false positives). The Enrichment Factor (EF) measures the enrichment of the method compared with random selection, and is defined by equation $\text{EF} = \text{Precision} * \text{NC}_{\text{tot}} / \text{NC}$, where NC_{tot} is the total number of molecules of the database ($\text{NC}_{\text{tot}}=802 \text{ HIT}+18 \text{ in-house}$) and NC is the total number of compounds obtained at the end of the VS protocol (88). Among the *in-house* database compounds showing a docking score higher than 20, therefore enclosed within the hits of the VS, were selected some structures to test, favouring the aglycones and maximizing the structural variability.

Turbidimetric assay

Wild-type TTR fibrillogenesis was evaluated in vitro by turbidimetric UV-vis assay at moderately acid pH (4.2). To ensure that the turbidity depended of TTR amyloid fibrils, the absence of absorbance at 400 nm was checked before assaying. 7.2 μM wt-TTR was preincubated (30 min) at neutral pH (10 mM phosphate buffer, 100 mM KCl, 1.7 mM EDTA, pH 7.6) with 2 molar equivalents of natural compounds dissolved in DMSO, or DMSO and 2 molar equivalents of diflunisal for negative and positive controls, respectively. 200 mM acetate buffer, 100 mM KCl and 1.7 mM EDTA, was added to yield final pH 4.2. The microplate wells was then incubated (72h) at 37 °C to promote the fibrillogenesis. The percentage of fibril formation was determined by observing the increase in turbidity at 400 or 450 nm (for compound **8**) in the absence and in the presence of inhibitors.

Acknowledgements

This work was partially supported by the Italian Ministero dell'Istruzione, dell'Universita` e della Ricerca (PRIN 20109MXHMR_007).

Keywords: Docking, Fibril formation, Natural products, Transthyretin, Virtual screening

References

- [1] J. A. Hamilton, M. D. Benson, *Cell. Mol. Life Sci.* **2001**, 58, 1491-1521.
- [2] M. J. Saraiva, J. Magalhaes, N. Ferreira, M. R. Almeida, *Curr. Med. Chem.* **2012**, 19, 2304-2311.
- [3] X. Hou, M. I. Aguilar, D. H. Small, *FEBS J* **2007**, 274, 1637-50.
- [4] G. J. Miroy, Z. Lai, H. Lashuel, S. A. Peterson, C. Strang, J. W. Kelly, *Proc. Natl. Acad. Sci.* **1996**, 93, 15051-15056.
- [5] G. Ortore, A. Martinelli, *Curr. Med. Chem.* **2012**, 19, 2380-2387.
- [6] T. Klabunde, H. M. Petrassi, V. B. Oza, P. Raman, J. W. Kelly, J. C. Sacchettini, *Nat. Struct. Biol.* **2000**, 7, 312-321.
- [7] S. A. Peterson, T. Klabunde, H. A. Lashuel, H. E. Purkey, J. C. Sacchettini, J. W. Kelly, *Proc. Natl. Acad. Sci.* **1998**, 95, 12956-12960.
- [8] S. Connelly, S. Choi, S. M. Johnson, J. W. Kelly, I. A. Wilson, *Curr. Opin. Struct. Biol.* **2010**, 20, 54-62.
- [9] S. Nencetti, E. Orlandini, *Curr. Med. Chem.* **2012**, 19, 2356-2379.
- [10] T. Liu, G. Bitan, *Chem.Med.Chem* **2012**, 7, 359-374.
- [11] S. Taniguchi, N. Suzuki, M. Masuda, S. Hisanaga, T. Iwatsubo, M. Goedert, M. Hasegawa, *J. Biol. Chem.* **2004**, 280, 7614-7623.
- [12] N. Ferreira, I. Cardoso, M. Domingues, R. Vitorino, M. Bastos, G. Bai, Saraiva, M.J., M. Almeida, *FEBS Lett.* **2009**, 583, 3569-3576.

- [13] M. Almeida, M. Saraiva, *FEBS Letters* **2012**, *586*, 2891–2896.
- [14] A. Attar, F. Rahimi, G. Bitan, *Transl. Neurosci.* **2013**, *4*, 385-409 .
- [15] <http://zinc.docking.org/catalogs/hitnp>, [Online].
- [16] P. Baures, S. Peterson, J.W. Kelly, *Bioorg Med Chem* **1998**, *6*, 1389-1401.
- [17] D. Trivella, C. Dos Reis, L. Lima, D. Foguel, I. Polikarpov, *J. Struct. Biology* **2012**, *180*, 143-153.
- [18] N. Ferreira, A. Pereira-Henriquesa, M. Almeida, *Biochem Biophys Rep* **2015**, *3*, 123–133.
- [19] N. Ferreira, M. Saraiva, M. Almeida, *FEBS Letters* **2011**, *585*, 2424–2430.
- [20] T. Yokoyama, Y. Kosaka, M. Mizuguchi, *J. Med. Chem.* **2014**, *57*, 8928–8935.
- [21] I. Iakovleva, A. Begum, M. Pokrzywa, M. Walfridsson, A. Sauer-Eriksson, A. Olofsson, *PlosONE* **2015**, e0128222
- [22] T. Tuccinardi, E. Nuti, G. Ortore, A. Rossello, S. Avramova, A. Martinelli, *Bioorg. Med. Chem.* **2008**, *16*, 7749-58.
- [23] D. Kitchen, H. Decornez, J. Furr, J. Bajorath, *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.
- [24] L. Bohlin, B. Rosin, *Drug Discovery Today* **1996**, *1*, 343-351.
- [25] Y. Wu, Z. Wu, P. Butko, Y. Christen, M. Lambert, W. Klein, C. Link, Y. Luo, *J. Neurosci.* **2006**, *26*, 13102-13113.
- [26] S. Bastianetto, C. Charles Ramassamy, S. Doré, Y. Christen, J. Poirier, R. Quirion, *Eur. J. Neurosci.* **2001**, *12*, 1882-1890.
- [27] J. Song, X. Lin, RN. Wong, SC. Sze, Y. Tong, P. Shaw, Y. Zhang, *Phytother Res* **2011**, *25*, 435-443.
- [28] P. Florio, C. Folli, M. Cianci, D. Del Rio, G. Zanotti, R. Berni, *J. Biol. Chem.* **2015**, *290*, 29769-29780.
- [29] L. Ciccone, L. Tepshi, S. Nencetti, E. Stura, *N. Biotechnol.* **2015**, *32*, 54-64.
- [30] J. Landete, *Food Res. Int.* **2012**, *46*, 410–424.
- [31] N. Green, T. Foss, J. Kelly, *Proc. Natl. Acad. Sci.* **2005**, *102*, 14545–14550.
- [32] M. Leri, D. Nosi, A. Natalello, R. Porcari, M. Ramazzotti, F. Chiti, V. Bellotti, S. Doglia, M. Stefani, M. Bucciantini, *J. Nutr. Biochem* **2016**, Available online 12 January.
- [33] L. Faiella, A. Temraz, N. De Tommasi, A. Braca, *Phytochemistry* **2012**, *76*, 172-177.

- [34] M. Bruzual De Abreu, A. Temraz, N. Malafronte, F. Gonzalez-Mujica, S. Duque, A. Braca, *Chem. Biodivers.* **2011**, *8*, 2126-2134.
- [35] A. Pawlowska, M. De Leo, A. Braca, *J. Agric. Food Chem.* **2006**, *54*, 10234-10238.
- [36] D. Spera, T. Siciliano, N. De Tommasi, A. Braca, A. Vessières, *Planta Med.* **2007**, *73*, 384-387.
- [37] F. Dal Piaz, A. Vassallo, A. Temraz, R. Cotugno, M. A. Belisario, G. Bifulco, M. Chini, C. Pisano, N. De Tommasi, A. Braca, *J. Med. Chem.* **2013**, *56*, 1583-1595.
- [38] M. Vera Saltos, B. Naranjo Puente, N. Malafronte, A. Braca, *J. Brazil Chem. Soc.* **2014**, *25*, 2121-2124.
- [39] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins* **2003**, *52*, 609-623.
- [40] Maestro, version 9.0, Schrödinger Inc., Portland, OR (USA), **2009**.
- [41] Macromodel, version 9.7, Schrödinger Inc., Portland, OR (USA), **2009**.
- [42] L. Gales, S. A. G. Macedo-Ribeiro, G. Valencia, M. Saraiva, A. Damas, *Biochem J.* **2005**, *388*, 615-621.
- [43] M. Miyata, T. Sato, M. Kugimiya, M. Sho, T. Nakamura, S. Ikemizu, M. Chirifu, M. Mizuguchi, Y. Nabeshima, Y. Suwa, H. Morioka, T. Arimori, M. Suico, T. Shuto, Y. Sako, M. Momohara, T. Koga, S. Morino-Koga, Y. Yamagata, H. Kai, *Biochemistry* **2010**, *49*, 6104-6114.
- [44] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin. *J Comput. Chem.* **2004**, *25*, 1605-1612.
- [45] R. Maheshwari, A. Singh, J. Gaddipati, R. Srimal, *Life Sci.* **2006**, *78*, 2081-2087.
- [46] T. Yokoyama, M. Ueda, Y. Ando, M. Mizuguchi, *Sci Rep.* **2015**, *5*, srep13570.

Chart 1: Natural compounds of our *in-house* database

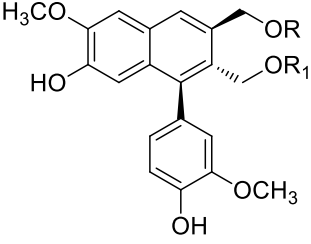
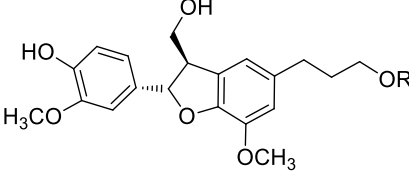
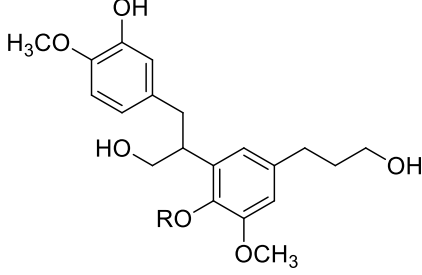
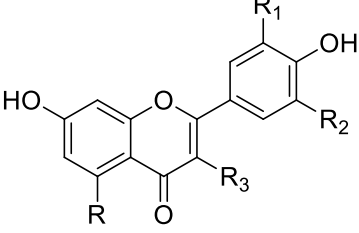
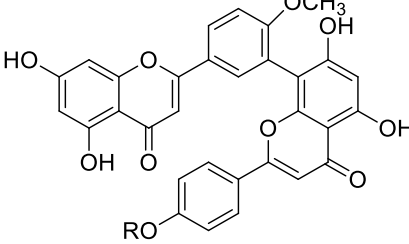
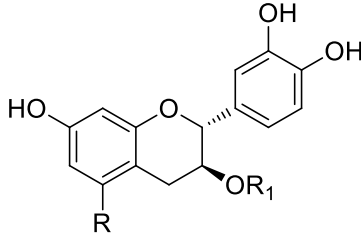
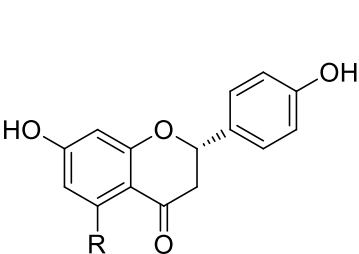
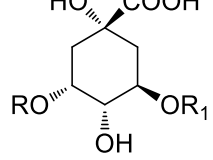
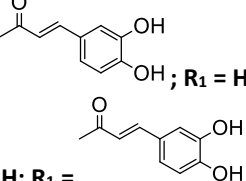
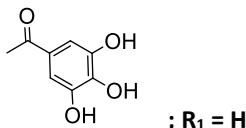
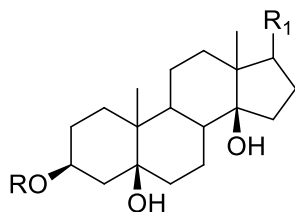
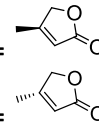
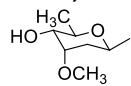
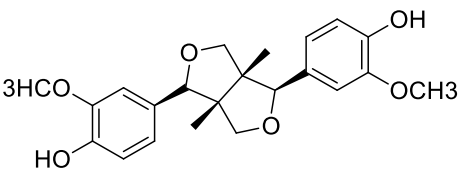
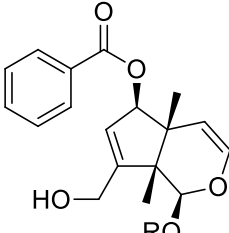
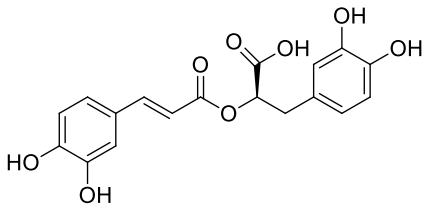
 <p>1: R = β-Glc; R₁ = H 2: R = H; R₁ = β-Glc</p>	 <p>3: R = H 4: R = β-Glc</p>	 <p>5: R = β-Glc</p>
 <p>6: R = H; R₁ = H; R₂ = H; R₃ = H 7: R = OH; R₁ = H; R₂ = H; R₃ = H 8: R = OH; R₁ = OH; R₂ = H; R₃ = OH 9: R = OH; R₁ = OH; R₂ = OH; R₃ = OH</p>	 <p>10: R = H 11: R = CH₃</p>	 <p>12: R = OH; R₁ = H 13: R = OH; R₁ = α-Rha 14: R = H; R₁ = H</p>
 <p>15: R = H 16: R = OH</p>	 <p>17: R =  18: R = H; R₁ =</p> <p>19: R =  19: R = ; R₁ = H</p>	 <p>20: R = H; R₁ =  21: R = H; R₁ =</p> <p>22: R =  22: R = ; R₁ =</p>
 <p>23</p>	 <p>24: R = β-Glc</p>	 <p>25</p>

Table 1: Docking score of compounds **1-15**

Compound	Docking Score
1	14.7
2	17.0
3	21.7
4	20.3
5	14.7
*6[a]	25.7
*7 (Apigenin) ^[b] ZINC03871576	26.4
*8 (Quercetin) ^[b] ZINC03869685	23.6
9 (Myricetin) ^[a] ZINC03874317	23.2
10 (Bilobetin) ZINC03979028	19.6
11	18.9
*12 (Catechin) ^[c] ZINC00119983	23.5
13	13.7
14 (Fisetidinol)	23.4
15 (Liquiritigenin)	24.9
*16 (Naringenin) ^[b] ZINC00156701	23.5
17	11.8
18	13.8
19	12.7
20	23.8
21	26.3
22	19.2
23 (Pinoresinol)	23.5
24	9.5
*25 (Rosmarinic Acid) ^[d,e] ZINC00899870	25.2

*Compounds already known as TTR fibrils inhibitors. [a]: Ref. [16]; [b]: Ref. [17]; [c]: Ref. [18]; [d]: Ref. [19]; [e]: Ref. [20]

Table 2 : Top scored *in-house* compounds ranking

Compounds
*7 (Apigenin) [a]
21
*6 [b]
*25 (Rosmarinic Acid) [c,d]
15 (Liquiritigenin)
20
*8 (Quercetin) [a]
*12 (Catechin) [e]
*16 (Naringenin) [a]
23 (Pinoresinol)
14 (Fisetidinol)
*9 (Myricetin) [b]
3
4
10 (Bilobetin)

*Compounds already known as TTR fibrils inhibitors. [a]: Ref. [17]; [b]: Ref. [16]; [c]: Ref. [19]; [d]: Ref. [20]; [e]: Ref. [18]

Figure 1: TTR structure highlighting the traditional T4 pocket and the new EGCG sites. In magenta, the regions detected by GOLD for blind docking

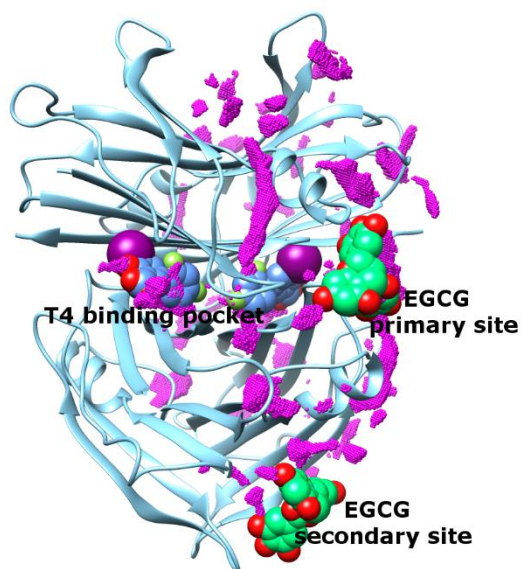


Figure 2: Docking poses of the HIT compounds preferring the EGCG binding site, compared to the EGCG RX pose (a). Docking of ZINC01532024, ZINC04245650, ZINC08552016, ZINC13816216, ZINC31156201, ZINC95098816, ZINC950988168, ZINC95098819 (b); docking of ZINC02000967, ZINC43450320, ZINC95098766 (c); docking of ZINC30726694, ZINC95098792, ZINC95098793 (d); docking of ZINC05369365, ZINC05369366, ZINC05369367 (e).

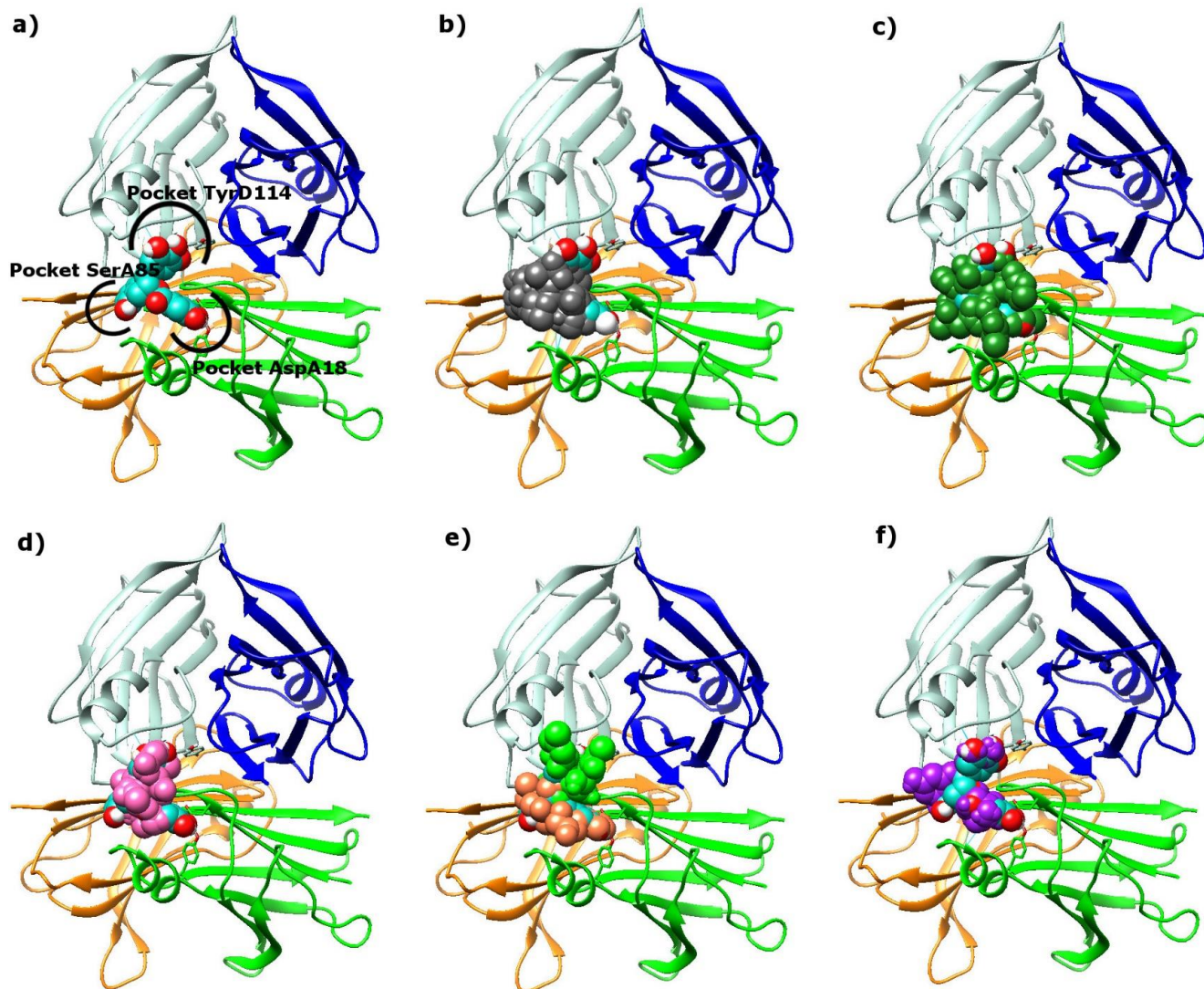


Figure 3: (a) Docking of **10** (purple) into EGCG binding site, compared with the experimental EGCG pose (cyan); (b) docking of **10** into the T4 binding pocket.

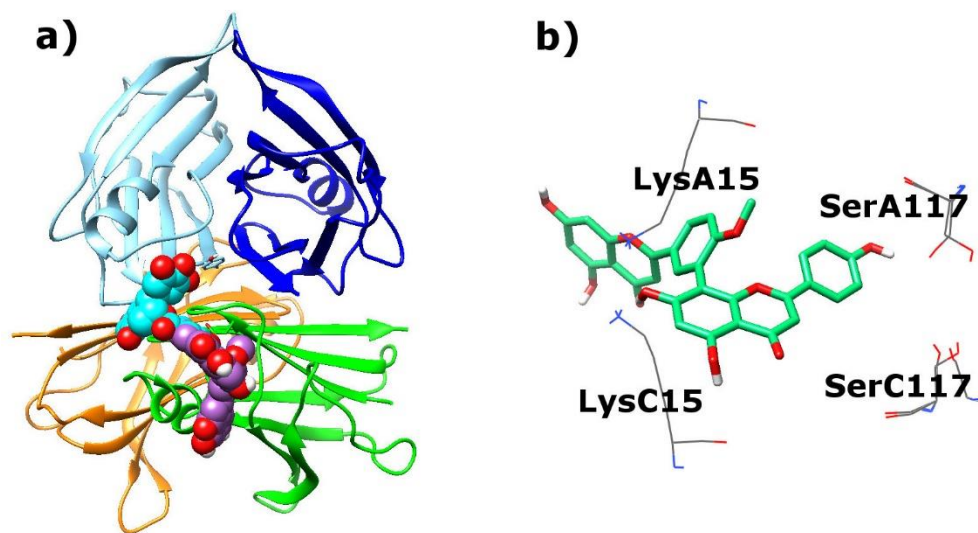


Figure 4: Docking of some *in-house* compounds in the T4 binding site: a) **6** (gold), **7** (magenta), **8** (yellow), **9** (green), **15** (blue) and **16** (violet); b) **12** (grey) and **14** (violet); c) **21**; d) **23**; e) **17** (magenta) and **18** (green); f) **19**; g) **24**; h) **3** (gold) and **4** (green); i) **25** (blue sky) compared to the experimental one (green).

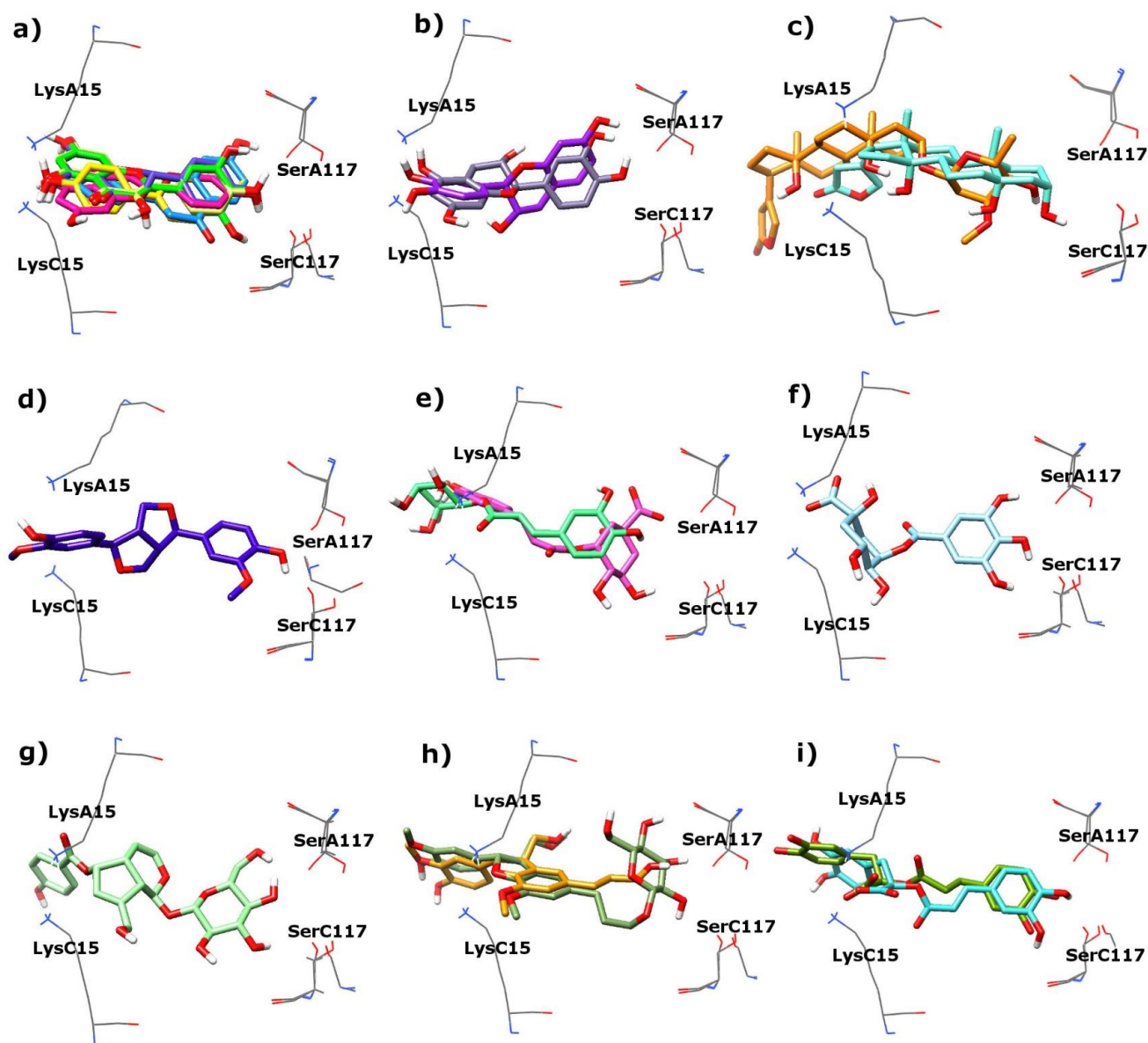


Figure 5: ROC curve of the VS. Serie 1: actives; serie 2: inactives. Yellow marked the *in house* database compounds.

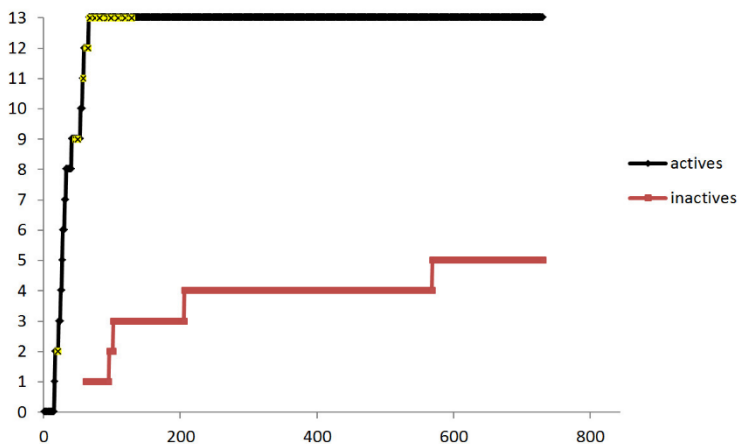


Figure 6: Scheme of our VS workflow

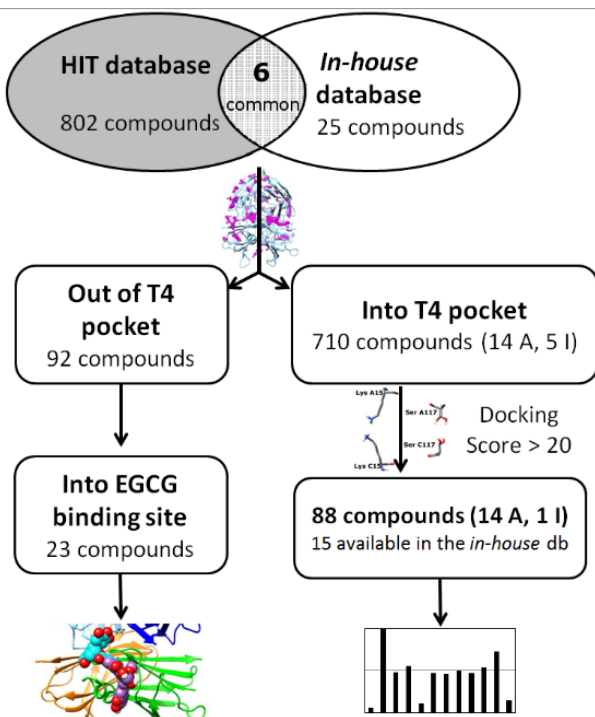


Figure 7: In vitro acid-mediated *wt*-TTR (7.2 μ M) percentage of fibril formation in presence of the hit compounds. For each compound, the formation is reported as the mean \pm 5 (standard error), from three independent determinations.

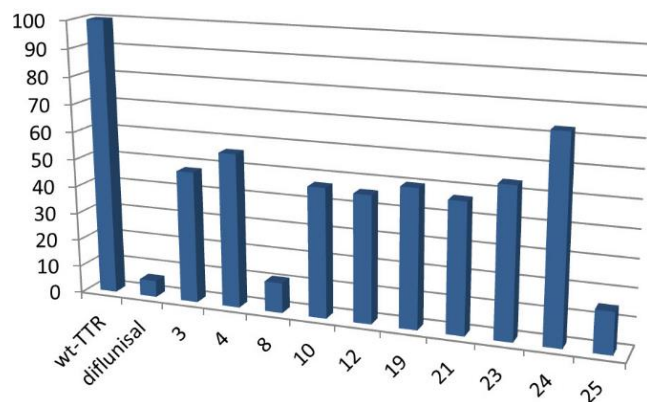
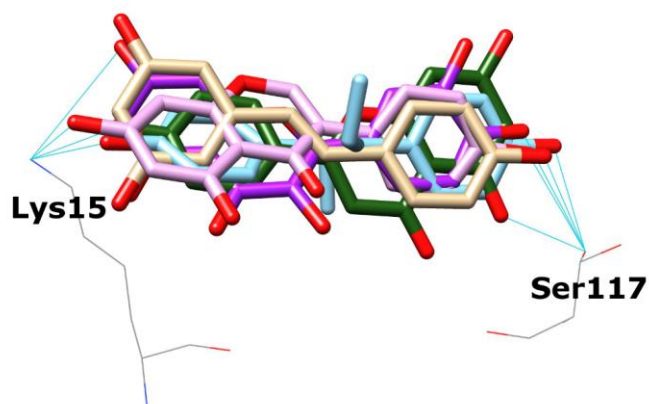


Figure 8: Superimposition of 4WO0, 1TT6, 3FGU, 4WNJ and 1DVS crystallographic structures, showing the key interactions of apigenine (green), diethylstilbestrol (cyan), genisteine (pink), quercetine (puple) and resveratrol (gold).



TOC: Looking for a help from the plant kingdom for transthyretin amyloidosis by exploring different TTR binding sites through an *in-house* library of natural compounds

