J Mol Model (2013) 19:151–162
DOI 10.1007/s00894-012-1532-y

ORIGINAL PAPER

Molecular modeling to investigate the binding of Congo red toward GNNQQNY protofibril and in silico virtual screening for the identification of new aggregation inhibitors

Jian-Hua Zhao · Hsuan-Liang Liu · Pavadai Elumalai · Wei-Hsi Chen · Lee-Chung Men · Kung-Tien Liu

Received: 16 February 2012 / Accepted: 8 July 2012 / Published online: 27 July 2012

Abstract Understanding the nature of the recognition between amyloid protofibrils and dye molecules at the molecular level is essential to improving instructive guides for designing novel molecular probes or new inhibitors. However, the atomic details of the binding between dyes and amyloid fibrils are still not fully understood. In this study, molecular docking, consensus scoring, molecular dynamics (MD), and molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) analyses were integrated to investigate the binding between Congo red (CR) and the GNNQQNY protofibril from yeast prion protein Sup35 and to further evaluate their binding stabilities and affinities. Our results reveal that there are four CR binding sites located on GNNQQNY protofibril surface. These four CR binding sites adopt dual binding modes by which CR binding with its long axis parallel and perpendicular to the long axis of the protofibril. In addition, CR was also found to bind to the edge of the protofibril via hydrophobic/aromatic and hydrogen-bonding interactions, which is inferred as the possible inhibition mechanism to prevent the elongation of the protofibril from the addition of incoming peptides. Virtual screening from National Cancer Institute (NCI) database obtained three hit compounds with higher binding affinity than CR to the edge of the protofibril due to the fact that the central parts of these compounds are able to form additional hydrogen bonds with the protofibril. The results of the study could be useful for the development of new molecular probes or inhibitors for clinical applications.

Keywords Amyloid protofibrils · Congo red · MM-PBSA · Molecular docking · Molecular dynamics · Virtual screening

Electronic supplementary material The online version of this article (doi:10.1007/s00894-012-1532-y) contains supplementary material, which is available to authorized users.

J.-H. Zhao · W.-H. Chen (✉) · L.-C. Men
Chemistry Division, Institute of Nuclear Energy Research,
1000, Wunhua Rd., Longtan Township,
Taoyuan County 32546, Taiwan
e-mail: whchen@iner.gov.tw

H.-L. Liu · P. Elumalai
Graduate Institute of Biotechnology,
National Taipei University of Technology,
1 Sec. 3 ZhongXiao E. Rd.,
Taipei 10608, Taiwan

H.-L. Liu (✉)
Department of Chemical Engineering and Biotechnology,
National Taipei University of Technology,
1 Sec. 3 ZhongXiao E. Rd.,
Taipei 10608, Taiwan
e-mail: f10894@ntut.edu.tw

K.-T. Liu
Everlight Chemical Industrial Corporation,
6th Fl, 77, Tun Hua South Road, Sec.2,
Taipei 106, Taiwan

Introduction

Amyloid fibrils, which can be stained with dyes that preferentially interact with β-sheets, such as Congo red (CR), Thioflavin T (ThT) and others, a fact that can be exploited for neuropathological diagnosis, are insoluble, ordered assemblies of normally soluble proteins. So far, more than 30 different unrelated proteins and peptides are known to cause diseases due to their misfolding and assembly into amyloid fibrils [1]. It is believed that the protein/peptide fibril formation and deposition share common pathogenic mechanisms. Experimental evidence has shown the presence of strong reflections at 4.8-Å and at 10- to 11-Å resolution in the fiber diffraction patterns of these aggregates, leading to the commonly accepted notion that a cross-β-spine

represents the basic structural motif of amyloid fibers [2, 3]. Although several models, essentially variations on a common theme (cross-β-spine), have been proposed [4], the complete structural definition of these aggregates is still highly debated.

To date, it is still difficult to obtain atomic-resolution structures with traditional experimental methods of structural biology due to the non-crystalline and insoluble nature of amyloid fibrils. However, some short peptides are capable of forming both fibrils and microcrystals that are sometimes found together in solution [5]. Recent crystallographic studies have provided atomic-level insight into amyloid-like microcrystals formed by several short (four to seven residues) peptides [5], revealing a common "steric zipper" motif where pairs of β-sheets form a dry interface with an interdigitation of side chains [5, 6]. One of these peptides is the heptapeptide fragment (GNNQQNY) from yeast prion protein Sup-35. This short fragment has many of the common characteristics of full-length Sup-35, such as cooperative kinetics of aggregation, formation of fibrils, and binding of CR [5]. A recent theoretical study of pairs of β-sheets containing three to 50 strands of the heptapeptide GNNQQNY using molecular dynamics (MD) simulations reported that the β-sheets are twisted in the absence of crystal packing while the cross-β spine architecture is preserved [7].

Staining with the CR dye has been used to indentify and detect amyloid fibrils since the beginning of the 1920s [8]. CR is a symmetrical linear molecule (Fig. 1a) with the hydrophobic center (central part) containing two phenyl rings linked through diazo bonds to two charged terminal naphthalene moieties. These terminal parts of CR contain two negatively charged sulfate groups (denoted as SO_3^- group) and two amine groups. It has been reported that CR can reduce the neurotoxicity of amyloid-β peptide (Aβ) in AD patients by binding to preformed fibrils [9]. The importance of these two SO_3^- groups on CR for the binding has been shown by removing these groups and thereby dramatically reducing the affinity [10]. However, the role of the CR SO_3^- groups on amyloid fibril inhibition still remains unknown.

Amyloid fibril-related diseases have affected a large human population worldwide without effective treatments. The identification of amyloid fibrils and, more likely, protofibrillar species as the primary sources of toxicity [9] has raised the hope that effective therapeutic agents could be developed in the near future. However, the binding mode(s) of the molecular probes or the inhibition mechanisms of inhibitors for amyloid fibrils are still not fully understood. Although the binding modes of CR and ThT have been proposed, high-resolution characterizations of these dyes with the steric zippers proposed to resemble the structure of the cross-β spine [5, 6] have not yet succeeded. The lack of high-resolution models complicates the examination of the binding mode to amyloid fibrils, and existing models are both incomplete and conflicting. These limitations hinder not only our understanding of binding mechanisms but also the design of better dyes or inhibitors for clinical purposes.

Although molecular dynamics (MD) simulations have been widely used to gain atomic insights into the binding interactions between small molecules and amyloid fibrils [11-13], it is still a time-consuming task to recognize or search for the specific binding sites on a given receptor using this approach due to the long kinetic process. In this study, a new strategy combining molecular docking, consensus scoring, MD simulations, and molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) analysis was developed to investigate the binding modes of CR toward GNNQQNY protofibrils and to evaluate their binding stabilities and affinities. Regarding the binding sites and modes of CR on GNNQQNY protofibrils, a possible inhibition mechanism that has also been inferred by previous theoretical studies [11-13] was further discussed. Subsequently, a virtual screening technique was employed to identify new inhibitors with higher affinity from the National Cancer Institute (NCI) compound database. The results of this study provide useful information for the development of new molecular probes and inhibitors for clinical applications.

Methods

Identification of CR binding sites and modes for GNNQQNY protofibrils

Global search for the CR binding sites on the GNNQQNY protofibril surface

Esposito and coworker performed several MD simulations for various GNNQQNY aggregates, such as SH2-ST10 (two β-sheet layers with ten β-strand for each layer) and SH2-ST50 (two β-sheet layers with 50 β-strand for each layer) [7]. Most importantly, they observed a significant twist of β-strands (Fig. 1b and c), suggesting that twisted β-sheet structures as a basic motif of amyloid fibrils. This structure of GNNQQNY protofibril with twisting β-strands represents a good model for studying the binding interactions between dye molecules and amyloid protofibril. In this study, the cross-β-spine of the GNNQQNY protofibril (Fig. 1b and c), exhibiting the so-called "steric zipper" model consisting of two β-sheet layers with ten twisting β-strands for each layer, was used as the receptor for the subsequent simulations performed with the modeling program Discovery Studio 2.1 (Accelrys Inc., San Diego, USA) under the CHARMm force field. The 3D structure of CR was built and optimized by the "*visualization*" tool and "*minimization*" protocol, respectively.

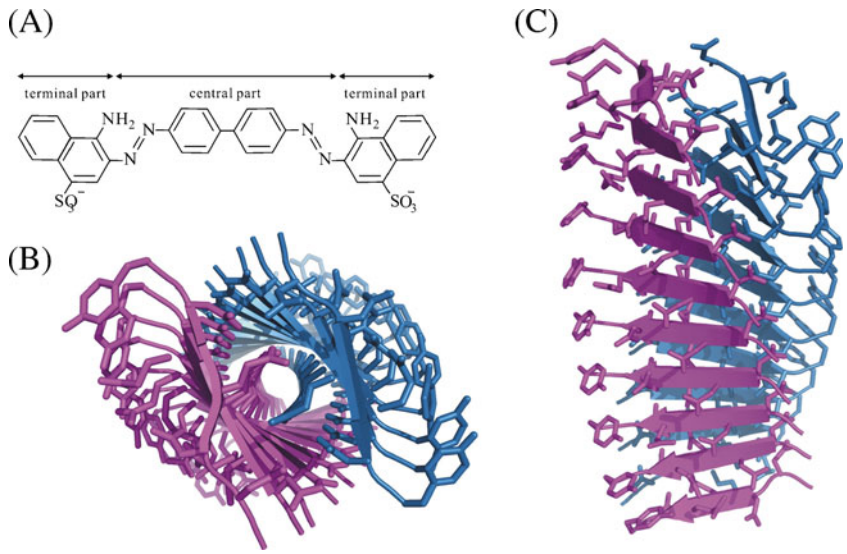

Fig. 1 **a** Chemical structure of CR. The 3-D structures of (**b**) top and (**c**) side view of GNNQQNY protofibril. The central and terminal parts of CR molecule are indicated. The heavy atoms of side chains of GNNQQNY protofibril are shown in stick. β-sheet configuration of the secondary structure elements for each β-strand is displayed as arrows. β-strands in β-sheet layer 1 and 2 are colored in magenta and blue, respectively. Each β-sheet layer contains 10 β-strands. This same figure representation used for all figures of GNNQQNY protofibril and all figures were generated with the program PyMol

Prior to performing a global search of the binding sites on the GNNQQNY protofibril surface, "*diverse comformation generation*" and "*analyze ligand poses*" protocols were used to diversify low-energy conformers of CR molecule and compute all-against-all root-mean-square deviation (RMSD) between pairwise CR conformers, respectively, so that two CR conformers with pairwise heavy atom RMSD value below 2 Å were assigned to the same cluster. The representative CR conformer from each of the clusters was used as a probe to screen and detect the possible binding sites on the GNNQQNY protofibril surface using rigid-body docking calculations based on fast Fourier transform (FFT) correlation method [14]. This calculation was made by Pipeline Pilot (Accelrys Inc., San Diego, USA). At this stage, rigid-body docking produces billions of putative complexes; thus, pairwise shape complementarity was used for scoring functions to retain 5000 predictions [15]. The algorithm searches exhaustively the entire rotational and translational space of the rigid CR with respect to the fixed GNNQQNY protofibril. The rotational search was performed by explicitly rotating the CR around each of its three Cartesian angles by a certain increment, 6 ° in this case. For every rotation, the translational space was rapidly scanned using the FFT method. The cluster analysis was further performed, and the RMSD (maximal ligand interface RMSD from the cluster center for a pose to be included in a cluster) and interface cutoff (cutoff distance used to define the interface region between receptor and ligand) were both set to 10 Å. The result indicates 47 clusters which could be further classified into four binding sites according to the binding locations of CR on GNNQQNY protofibril surface, as shown in supporting information (Fig. S1). The four binding sites were further used as inputs for the subsequent molecular docking calculation to predict the precise binding orientations of CR.

Docking and consensus scoring

The docking procedure was carried out using the "*CDOCKER"* protocol, which is a grid-based molecular docking method that employs CHARMm force field [16]. A sphere radius of 12 Å from the center of the ligand-binding site was specified for each site. Conformational search for CR was performed using a simulated annealing MD approach, in which ten conformations of CR were produced by heating to a temperature of 1000 K in 1000 dynamics steps. Further productions of random orientations were generated from the ten conformations by translating the center of the ligand to a specific binding site and performing a series of random rotations. Each orientation was then heated to a temperature of 700 K in 2000 steps and then annealed to 300 K in 5000 steps. The grid extension was set to 8 Å. One hundred binding poses of CR for each binding site were thus generated. Finally, the binding site was held rigid while the ligand was allowed to relax into their proper geometry during the minimization step. Ten scoring functions, LigScore1, LigScore2 [17], PLP1, PLP2 [18], Jain [19], PMF, PMF04 [20, 21], Ludi1, Ludi2, and Ludi3 [22], implemented in the Discovery Studio 2.1 program, were applied to obtain a single consensus score value for each docked pose of CR using the "*consensus score"* protocol. Thus, the best docked pose of CR for each of the four binding sites can be determined.

Molecular dynamics simulations and MM-PBSA binding free energy analysis

The CR-GNNQQNY complexes for each binding site were solvated in a rectangular box of water molecules with solvent layers spanning 10 Å between the box edges and the solute surface. The system was neutralized by adding

sodium counterions using the "*solvation*" protocol. The particle mesh Ewald (PME) method [23] was used to treat long-range electrostatic interactions, with cutoff distance value for counting nonbonded interaction pairs set at 12.0 Å. All covalent bonds to hydrogen atoms were constrained using the *SHAKE* algorithm [24]. The "*minimization"* and "*dynamics*" (*heating or cooling, equilibration, and production*) protocols were used to process 500-ps MD simulations for each CR-GNNQQNY complex. Each complex was energy-minimized to remove possible steric stress and bad contacts by the steepest descent method for the first 2500 steps and the conjugated gradient method for the subsequent 5000 steps. Then, gradual heating was performed from 0 to 300 K in 50 ps, followed by equilibration for 50 ps at 300 K using the NVT ensemble. During the heating procedure, 100 kcal mol^{-1} $Å^{-2}$ was specified to keep the protofibril constrained. Finally, periodic boundary dynamics simulations of 500 ps were carried out for the production step in an NPT ensemble at 1 atm and 300 K. The output trajectory files were saved every 1 ps for subsequent analysis.

To evaluate the binding affinity between CR and GNNQQNY protofibrils for each binding site, the binding free energies (ΔG_{bind}) of the CR-GNNQQNY complexes were calculated by the molecular mechanics-Poisson-Boltzmann surface area (MM-PBSA) method [25] using the "*calculate binding energy*" protocol. The binding free energies (ΔG_{bind}) were computed as:

$$\Delta G_{bind} = G_{complex} - \left(G_{receptor} + G_{ligand}\right). \quad (1)$$

The binding free energy is composed of an enthalpic and entropic contribution:

$$\Delta G_{bind} = \Delta H - T\Delta S. \quad (2)$$

The enthalpy of binding ΔH contains ΔE_{MM}, the change in the molecular mechanics free energy upon complex formation, ΔG_{solv}, the solvated free energy contribution, and-$T\Delta S$ represents the entropy term. The molecular mechanics free energy is calculated as follows:

$$\Delta E_{MM} = \Delta E_{\text{int}}^{ele} + \Delta E_{\text{int}}^{vdw}. \quad (3)$$

$\Delta E_{\text{int}}^{ele}$ and $\Delta E_{\text{int}}^{vdw}$ represent the electrostatic and van der Waals interaction energies between a ligand and protein, respectively. The solvation free energy can be divided onto two components.

$$\Delta G_{sol} = \Delta G_{sol}^{ele} + \Delta G_{sol}^{nonpol} \quad (4)$$

ΔG_{sol}^{ele} is the polar contribution to salvation and could be obtained by solving the Poisson–Boltzmann equation for MM-PBSA method. The dielectric values applied for water and protein were set to 80 and 1, respectively. ΔG_{sol}^{nonpol} is the nonpolar solvation term and was determined using:

$$\Delta G_{sol}^{nonpol} = \gamma \cdot \Delta SASA + b \quad (5)$$

where γ, representing the surface tension, and b, being a constant, were set to 0.00542 kcal/(mol $Å^2$) and 0.92, respectively. SASA is the solvent accessible surface area ($Å^2$) that was estimated with probe radius of 1.4 Å.

The ΔG_{bind} was calculated as the average of the binding free energy of the last 50 ps of MD simulations, taken at 1-ps intervals from the single trajectory of the CR-GNNQQNY complex simulation. The single trajectory approach was applied to estimate the energies and was proven to be successful in a previous study [26]. The entropic change of the solute molecule is generally estimated using normal mode analysis. However, this requires high computational demands and tends to have a large margin of error that introduces significant uncertainty in the result [27-29]. Because the system contains only one ligand, one can assume that the entropic differences among the CR-GNNQQNY complex systems are quite similar, and therefore the contribution of the entropy is negligible.

Binding stability and affinity of CR analog toward the edge of GNNQQNY protofibrils

To further understand the importance of the SO_3^- group of CR on the binding stability and affinity toward the edge of protofibrils (binding site D), two SO_3^- groups in complex D were replaced by two methoxyl groups (denoted as CR-OCH_3) to mimic the CR analog from a previous experimental study [10] using the "*visualization*" tool. The CR analog (CR-OCH_3) was then minimized with the fixed GNNQQNY protofibril and soaked in a rectangular box of water molecules with solvent layers spanning 10 Å between the box edges and the solute surface using the "*minimization*" and "*solvation*" protocols, respectively. Subsequently, 500-ps MD simulations for the CR-OCH_3-GNNQQNY complex were conducted to evaluate the binding stability using the same parameters as described above (Method section).

Virtual screening

The NCI compound database, consisting of 260,071 compounds, was screened for the identification of hit compounds toward the edge of the protofibril (binding site D). The 2-D ligand-based searching in terms of Lipinski rules was used as a first query to restrict drug-like compounds [30] using the "*ligand prepare*" protocol in DS 2.1. The

compounds which fit in with Lipinski rules were retrieved and were subjected to screening by pharmacophore-based virtual screening. In this stage, a 3D pharmacophore query was manually defined based on the structural features of complex D by using the "*pharmacophore*" tool as implemented in DS 2.1. The features considered were two H-bond donors and two hydrophobic or aromatic features. Then, the 3D pharmacophore screening was performed using the "*ligand pharmacophore mapping*" protocol with flexible method. A conformational set was generated for each molecule using the "best-quality" option, based on the CHARMm force field; in such a way, a conformational model consisting of a maximum of 255 conformers within 20 kcal mol^{-1} in energy from the global minimum was generated for each compound, so as to reproduce the flexibility of molecules during the database search. All other parameters were kept at their default settings. The molecules associated with their conformational models were mapped onto the pharmacophore model using the "best fit" option to obtain the bioactive conformation of each molecule. Only compounds that matched all the features of the pharmacophore model were retrieved as hits.

The hits obtained from the 3D pharmacophore search were docked into binding site D using the "*dock ligands* (*pharmacophore restrains and CHARMm*)" protocol, in which any 3D pharmacophore containing location constraints can be used in the initial docking of the compounds. Harmonic restraints were then created between the location constraints and the mapped atoms. Standard parameters were used during docking simulations. This protocol performs the following steps: (i) ligand conformations were generated for the ligands using the "*diverse conformation generation*" protocol; (ii) the conformations were then mapped to the pharmacophore using the "*ligand pharmacophore mapping*" protocol; (iii) restraints were set up between the location constraints in the pharmacophore and the mapped atoms in the ligand; (iv) simulated annealing and minimization steps from the "*CDOCKER*" protocol was then used to optimize the ligand pose in the active site; and (v) an energy filter was applied to remove poses that still clash with the protein. This pharmacophore-based docking was analyzed by fit values of pharmacophore mapping and visual inspection, only compounds that matched all the features of the pharmacophore mode were retrieved as hits. Finally, these compound-GNNQQNY complexes were soaked in a rectangular box of water molecules with solvent layers spanning 10 Å between the box edges and the solute surface using the "*solvation*" protocol and were subjected to 500-ps MD simulations and MM-PBSA analysis to evaluate their binding stability and affinity, using the same parameters as described above (Method section).

Results and discussion

CR binding sites and modes on GNNQQNY protofibril

Four CR binding sites

Four binding sites were detected through rigid-body docking calculations based on the FFT method. Then, molecular docking and consensus scoring calculations were further applied to determine the best docked pose of CR for each site as shown in Fig. 2. It shows that CR molecule bound to (i) the G1-N2-N3 groove (complex A), (ii) the aromatic ladder of Y7 (complex B), (iii) the repeating shallow grooves between the two β-sheet layers (complex C) with its long axis parallel to the long axis of the fibrils, and (iv) the edge of the protofibril (complex D) with its long axis perpendicular to the β-sheet extension direction. Several 500-ps MD simulations and MM-PBSA analyses were conducted to evaluate the stabilities and affinities of these CR-GNNQQNY complexes, respectively. In Fig. 3, the time series of atom contacts between CR and the protofibril was computed to indicate the stabilities of CR binding toward these four binding sites. It is obvious that CR can stably bind to these sites during the entire simulation courses, as reflected by the fact that all- and heavy-atom contacts between CR and the protofibril for these complexes are well preserved during the 500-ps MD simulations (Fig. 3a–d). The numbers of hydrogen bonds between the SO_3^- group and the protofibril or water molecules in these complexes were computed to indicate its role in binding (Fig. 3e–h). The results indicate that they exhibit similar characterizations, in which the two SO_3^- groups of CR tend to be exposed to solvent and form more hydrogen bonds with water molecules (8 to 11 hydrogen bonds) than with the protofibril (1 to 4 hydrogen bonds), which is consistent with the results from a recent MD study [31]. Furthermore, the averaged hydrogen bonds between the SO_3^- groups of CR and the protofibril for complexes A, B, C and D were 2.9±1.1, 1.4±1.0, 1.3±0.9, and 3.9±0.9, respectively, suggesting that hydrogen-bonding interactions play a role in binding, particularly for complex D (see below). Moreover, the MM-PBSA analysis also indicates that the binding affinities of CR toward the GNNQQNY protofibril are in the order of A>B>C>D (Table 1), suggesting that CR is able to bind to multiple sites with various binding affinities.

Moreover, to demonstrate that our statistics are converged, three additional MD simulations (run 1, 2, and 3) with different random number seed for each of CR-GNNQQNY complexes were conducted and the results were shown in supporting information (Fig. S2). In Fig. S2, it is obvious that the potential energy of system are converged, suggesting that our results are statistically reliable.

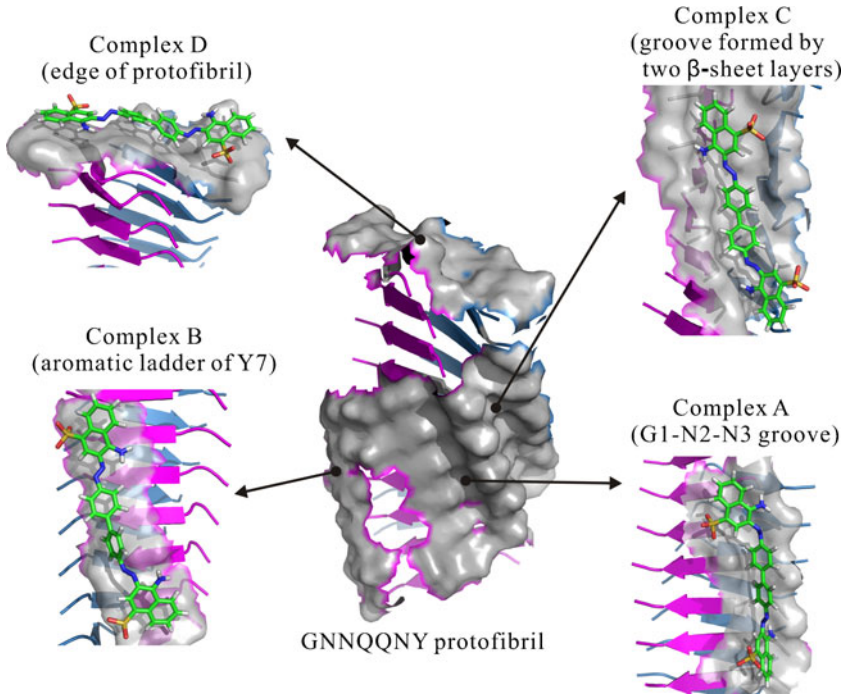

Fig. 2 The binding of CR molecule to four binding sites (**a**–**d**) of GNNQQNY protofibril. The GNNQQNY protofibril structure is shown in center and its four CR binding sites (**a**–**d**) are indicated in gray molecular surfaces. The semitransparent surfaces of four CR binding sites were set for better visualization. The binding of CR toward these four binding sites are also displayed in complex (**a**) (site A: G1-N2-N3 groove; below right), (**b**) (site B: aromatic ladder of Y7; below left), (**c**) (site C: groove formed by two β-sheet layers; above right), and (**d**) (site D: edge of protofibril; above left), respectively. CR is displayed in stick, while H, O, C, and S elements of CR are colored in white, red, cyan, and yellow, respectively

Binding mode of CR toward G1-N2-N3 groove (complex A)

A recent MD study by Wu et al. [31] has identified and characterized two specific binding modes of CR to the GNNQQNY protofibril. They proposed that the primary and secondary binding modes of CR were to the regular G1-N2-N3 groove formed by the β-strands along the wet side of the β-sheet extension direction and the edge of β-sheet layer with its long axis perpendicular to the β-sheet extension direction, respectively. The results of our study also suggested that CR primarily binds to the G1-N2-N3 groove along the β-sheet extension direction with a strong binding free energy of −23.7±1.7 kcal mol^{-1} (Table 1), in qualitative agreement with previous measurements (−22.8± 1.6 kcal mol^{-1}) [31]. This observation is also in good agreement with previous studies, suggesting that the long axis of the CR lies parallel to the long axis of the fibrils [32-36]. Moreover, the apolar part of CR (including six-membered carbon rings plus the diazo groups) is buried in the G1-N2-N3 groove, whereas the polar SO_3^- groups are

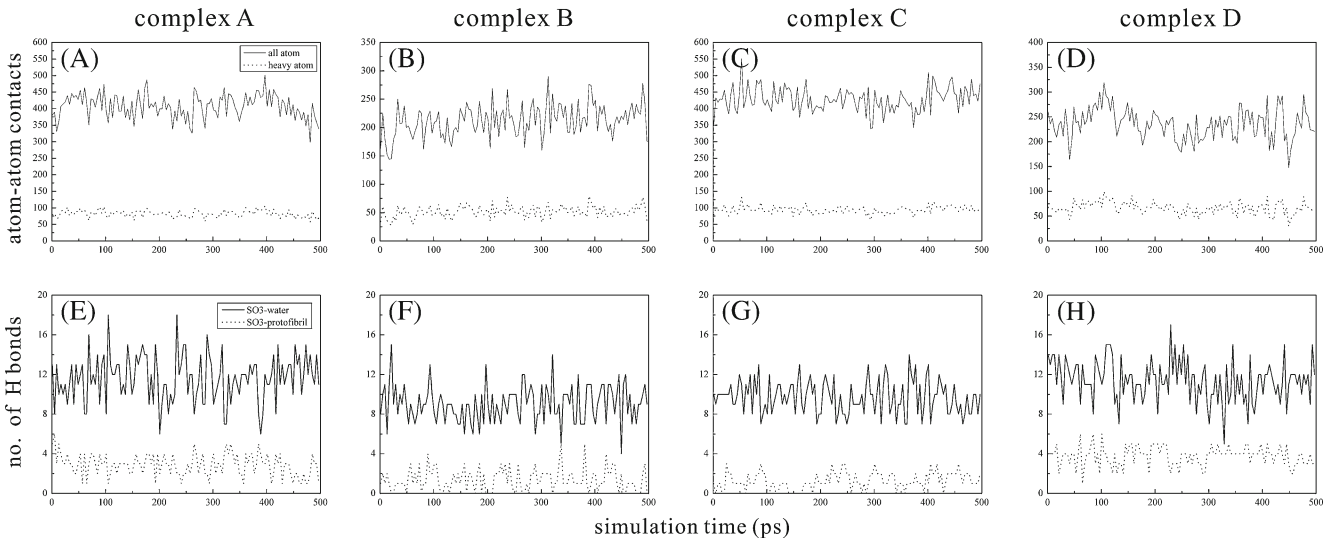

Fig. 3 The evolution of (**a**–**d**) the all- and heavy-atom contacts between CR and GNNQQNY protofibril (within 4 Å cutoff); and (**e**–**h**) of the number of hydrogen bonds between SO_3^- group of CR and protofibril or water for four complexes during 500-ps MD simulations. A hydrogen bond was assigned when the distance between donor and acceptor was ≤2.5 Å and the angle D-H…A was ≥120°

Table 1 Averaged MM-PBSA binding free energy for four CR-GNNQQNY complexes

Complexes	Locations	MM-PBSA binding free energy (kcal/mol)
A	G1-N2-N3 groove	−23.7±1.7
B	aromatic ladder of Y7	−20.5±1.9
C	groove formed by two β-sheet layers	−15.3±2.7
D	the edge of protofibril	−9.4±3.3

exposed to water extensively (Figs. 2 and 3e), suggesting that hydrophobic interactions are the main driving force governing this binding mode.

Binding mode of CR toward aromatic ladder of Y7 (complex B)

In complex B, CR binds with its long axis parallel to the long axis of the fibrils (Fig. 2) with a strong binding affinity (−20.5±1.9 kcal mol^{-1} (Table 1)). It reveals that the hydrophobic or aromatic interactions between the naphthalene ring of CR and the aromatic side chain of Y7 play an important role for binding. Recently, Biancalana et al. [37] suggested that ThT interacts with the β-sheet by docking onto surfaces formed by a single tyrosine ladder, rather than in the space between adjacent ladders, which directly support the existence of binding site B, the aromatic ladder of Y7.

Binding mode of CR toward groove formed by two β-sheet layers (complex C)

In complex C, CR was shown to bind in the repeating shallow grooves between the two β-sheet layers along the main-chain hydrogen-bond direction (Fig. 2) with moderate binding affinity (−15.3±2.7 kcal mol^{-1} (Table 1)). Furthermore, the apolar part of CR was found to fit very well in the groove between the two β-sheet layers while the polar SO_3^- groups were observed to be exposed to solvent (Figs. 2 and 3g). It suggests that hydrophobic interactions play an important role for binding as observed in complex A. A previous study by Childer et al. [38] proposed that CR bind to the laminate grooves formed by two neighboring β-sheet layers with a parallel orientation to the amyloid long axis and colinear with laminate grooves, which is in accordance with complex C observed in this study. In addition, recent experimental studies have reported that Aβ oligomers may have the same types of binding sites as for the $A\beta_{40}$ fibrils in vitro and in vivo [39, 40], suggesting that the binding sites on $A\beta_{40}$ oligomers only represent a subset of ligand binding sites on $A\beta_{40}$ fibrils. Similarly, in our present study, binding site C is formed by two β-sheet layers (protofibril level); in other words, only one β-sheet layer (oligomer level) could not be bound by CR at such a site. Furthermore, GNNQQNY oligomers (with one β-sheet layer) and protofibrils (with two β-sheet layers) share the same CR binding sites A and B as observed in this study, suggesting that the binding sites on GNNQQNY oligomers represent only a subset of ligand binding sites on GNNQQNY protofibrils, in good agreement with the experimental findings [39, 40]. These results imply that lower-order aggregates (such as oligomers with one β-sheet layer) may exhibit a certain degree of β-sheet structures, resulting in the formation of grooves (site A) and Tyr ladders (site B) on the β-sheet surface which could be bound by CR.

Binding mode of CR toward edge of β-sheet layer (complex D)

CR was found to bind to the edge of the protofibril with its long axis perpendicular to the β-sheet extension direction to form complex D, which is also in accordance to the secondary binding mode of CR proposed by Wu et al. [31]. Our results also show that complex D exhibits relatively low binding affinity (−9.4±3.3 kcal mol^{-1} (Table 1)). In addition, our simulations indicate that hydrophobic or aromatic and hydrogen-bonding interactions are important for binding in complex D (see below). Recent studies have pointed out that binding of inhibitors to the edge of amyloid fibrils can block the attachment of the incoming peptide [11-13], suggesting that binding site D can serve as a potential inhibition site to stop β-strand elongation.

In summary, our simulations suggest multiple CR binding sites on the surface of the GNNQQNY protofibril, which is in good agreement with the previous experimental findings [41]. These four CR-protofibril complexes can be classified as a dual binding mode by which CR binding with its long axis parallel (complexes A, B, and C) and perpendicular (complex D) to the long axis of the fibril, respectively, consistent with low-resolution spectroscopic experiments [41-45].

Inhibition mechanism

Binding site D served as potential site

Numerous theoretical attempts have been made to increase atomic insights in the binding interactions between inhibitors and amyloid fibrils. Much experimental data has implicated the extension of amyloid fibrils via monomer addition to their edges [46-48]. Recently, it has been reported that the binding of ibuprofen to Aβ destabilizes the interactions between incoming peptides and fibrils based on an implicit solvent MD simulations [11, 12]. More recently, Lemkul and coworkers [13] employed atomistic, explicit-solvent MD simulations to identify the mechanism of Aβ fibril destabilization by morin, one of the most effective anti-aggregation flavonoids, using a model of the

mature Aβ fibril and found that morin can bind to the edges of fibrils to block the attachment of an incoming peptide. Thus, it is expected that the binding of CR to the edges of the fibril can directly interfere with peptide-fibril interactions. Taken together, this evidence allows us to propose that our binding site D can serve as a potential inhibition site, in which CR binding to two edges of an amyloid protofibril could stop β-strand extension. Furthermore, this relatively low binding affinity of complex D when compared to other complexes is consistent with the fact that CR is a weak inhibitor for amyloid fibril formation (Table 1) [49, 50].

The specific interactions involved in inhibition

In complex D, CR was found to be anchored by two specific interactions: (*i*) hydrophobic or aromatic interactions between the naphthalene ring of CR and the aromatic side chains of Y7 and (*ii*) the hydrogen-bonding interactions between the two SO_3^- groups of CR and the C-terminal carboxyl and amino groups of Y7 (Fig. 4). This supports the view point showing that the aromatic groups of polyphenol inhibitors play an important role in binding to fibrils [49]. Moreover, a recent experimental study by Reinke and Gestwicki [10] has examined the structure-activity relationships of amyloid β-aggregation inhibitors based on curcumin, a natural organic dye sharing a similar chemical scaffold as CR, and reported that the substitutions of two methoxyl groups to the two hydroxyl groups on the terminal group of curcumin, in an equivalent position to the SO_3^- groups of CR, would dramatically reduce its binding affinity. To further ensure the role of SO_3^- groups of CR on binding, they were replaced by two methoxyl groups (denoted as CR-OCH_3) for additional MD simulations and the results are shown in Fig. 5. It is obvious that the atom contacts between CR-OCH_3 and the protofibril were unstable (Fig. 5a) and it tends to dissociate from the protofibril surface during the last 250 ps simulation as shown in Fig. 5b, suggesting that CR-OCH_3 is unable to block the attachment of an incoming peptide, fails to act as an aggregation inhibitor, which is in line with the previous experimental findings [10]. It can be attributed to the fact that the replacement of SO_3^- to OCH_3 group would directly disrupt hydrogen-bonding interactions between CR and the protofibril. The same results can also be observed in Fig. 3h, where two SO_3^- groups of CR form 3.9±0.9 stable hydrogen-bonding interactions with the C-terminal carboxyl and amino groups of Y7, stating the importance of SO_3^- groups of CR toward binding. Overall, complex D allows us to infer that CR binding onto two edges of amyloid protofibrils could stop β-strand extension, which may be one of the possible inhibition mechanisms of amyloid fibril formation and thus represents an attractive binding site for the identification of new potential inhibitors by virtual screening.

The identification of new inhibitors by structure-based pharmacophore virtual screening

Structure-based pharmacophore generation

The above evidences indicate that the naphthalene rings and hydrogen bond acceptors on the terminal part of CR can be chosen as the principal pharmacophore features for inhibition activity. Thus, a structure-based pharmacophore model was defined by the putative bound conformation of CR derived from complex D and, which was further used as a filter for the identification of hit compounds from the NCI database. The resulting model, constituted by two hydrophobic or aromatic features and two hydrogen bond acceptors, is shown in Fig. 6. The overall virtual screening workflow is shown in Fig. 7.

Virtual screening

The NCI database containing 260,071 compounds was selected as a resource for the following virtual screening. In the first step of 2-D ligand-based searching, the criteria in terms of the Lipinski's rule of five (≦5 H-bond donors,

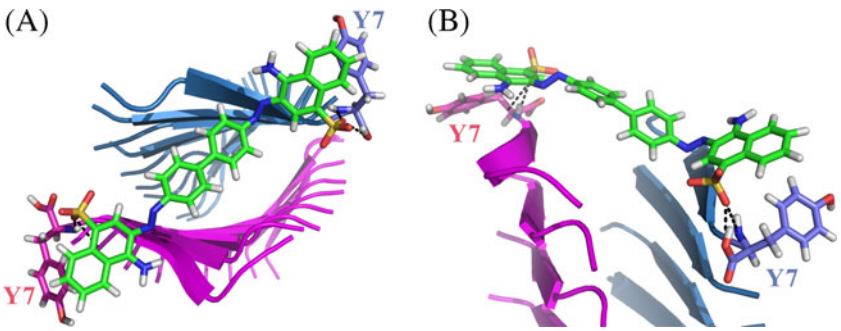

Fig. 4 The (**a**) top view and (**b**) side view of complex **D**. The side chains of Y7 of terminal strands in contact with naphthalene ring of the CR are also shown in stick. CR is displayed in stick, while H, O, C, and S elements of CR are colored in white, red, cyan, and yellow, respectively. Hydrogen bonds between CR and protofibril are indicated in black dash line

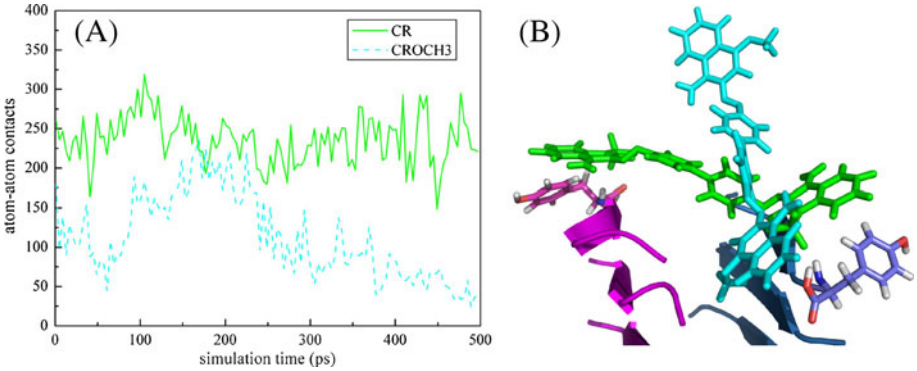

Fig. 5 **a** The evolution of the all-atom contacts between CR or CR analog (CR-OCH3) and GNNQQNY protofibril (within 4 Å cutoff). **b** The snapshot structures of CR-GNNQQNY complex and CR-OCH3-GNNQQNY complex at 500 ps (two snapshots are combined into one for comparison). The side chains of Y7 of terminal strands in contact with CR are also shown in stick. CR and its analog (CR-OCH3) are displayed in green and cyan stick, respectively

≦10 H-bond acceptors, MW≦500 Da, log P≦5) [30] was employed to preselect all the molecules out of the NCI database. After this process, 190,239 compounds were obtained for the following pharmacophore-based virtual screening. The pharmacophoric model was further applied as the filter to screen the 190,239 compounds and led to the selection of a subset of 255 compounds. Subsequently, they were subjected to pharmacophore-based docking calculations to predict the best binding pose of inhibitors in the binding site D of protofibrils and were followed by 500-ps MD simulations and MM-PBSA analysis to evaluate their binding stabilities and affinities. Finally, three hit compounds, namely NSC54429, NSC106129, and NSC279264, were shown to exhibit higher binding affinity than CR toward binding site D. Their chemical structures and corresponding binding affinities are shown in Fig. 8.

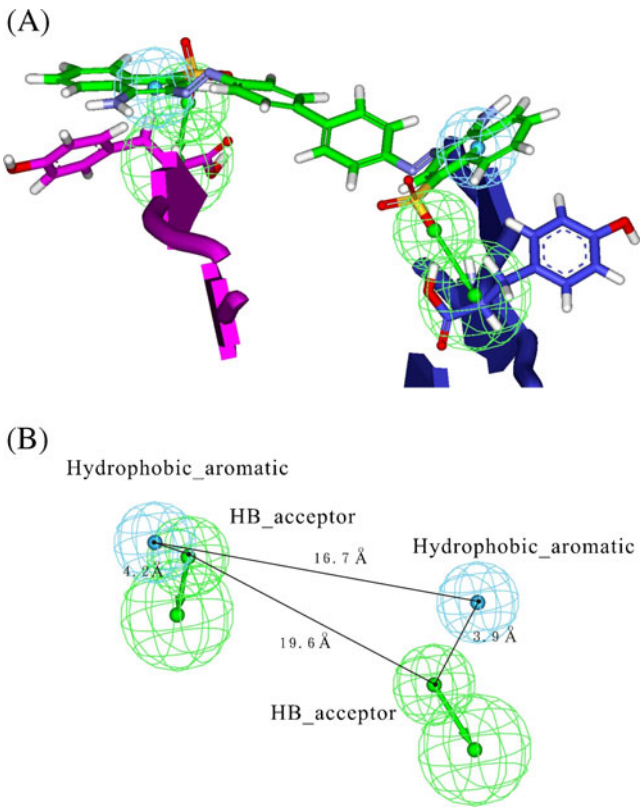

Fig. 6 **a** 3D pharmacophore query generated based on the complex **D**. **b** 3D Pharmacophore model which used as a query for the identification of hit compounds. The pharmacophore features of hydrophobic or aromatic and hydrogen bond acceptor are colored in cyan and green, respectively

Three potential inhibitors

Similar to CR, these three compounds exhibit two hydrophobic or aromatic and two hydrogen bond acceptor features on their terminal groups. In contrast, the central parts of these three compounds exhibit more hydrogen bond

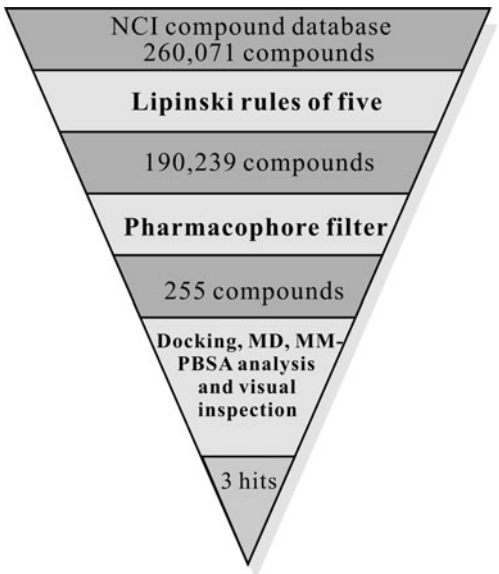

Fig. 7 Schematic for the high-throughput virtual screening protocol in the present study

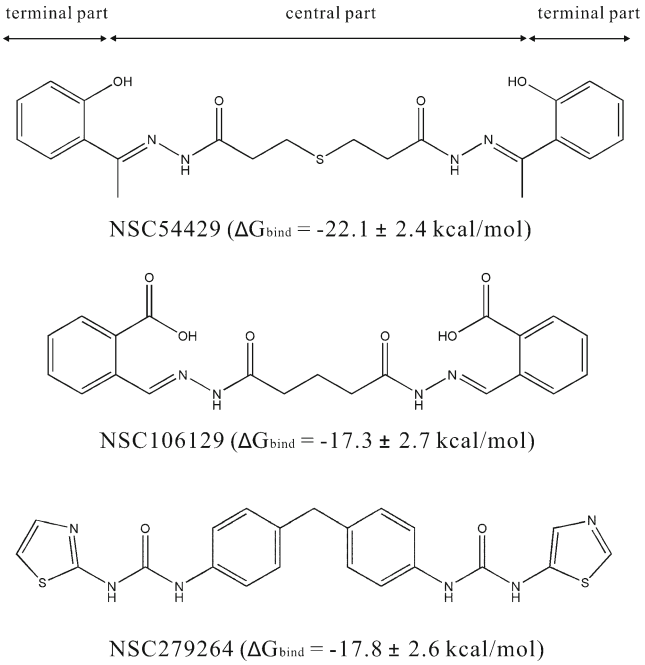

Fig. 8 Chemical structures of three hit compounds and their corresponding binding affinities (ΔG_{bind}) using MM-PBSA approach toward binding site D of GNNQQNY protofibril. The central and terminal parts of these molecules are indicated

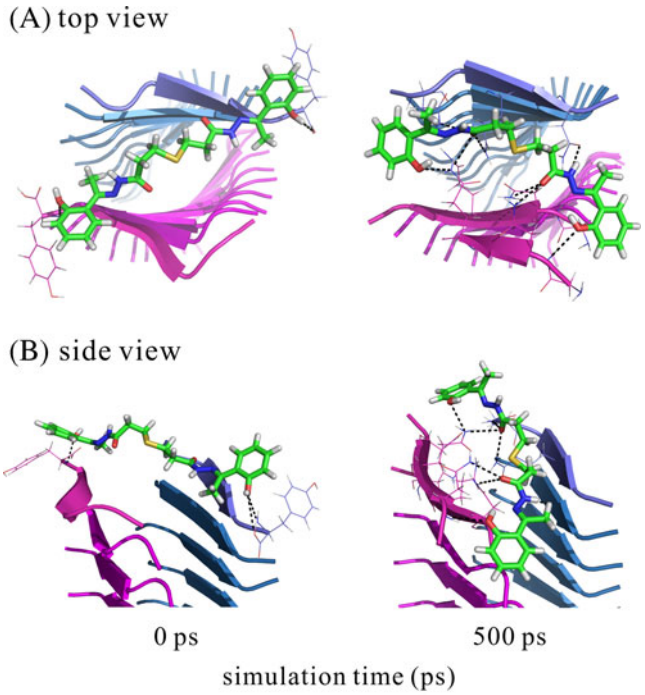

Fig. 9 The snapshots of **A** top view and **B** side view of NSC54429-GNNQQNY complex at 0 and 500 ps. The residues of protofibril which interact with NSC54429 are shown in line and the hydrogen bonds between NSC54429 and protofibril are shown in black dash line

donors and acceptors than CR, giving them the capacity to form more hydrogen bonds with the protofibrils, particularly for NSC54429. It forms ~8 hydrogen bonds with the main chain (backbone) of outer β-strands and the hydrophilic side chains of N2, Q4, and N6, resulting in a stronger binding affinity toward binding site D compared to CR (Fig. 9). It is worth mentioning that the initial binding orientation of NSC54429 was oblique to the β-strand and crossed two β-sheet layers, as predicted by the pharmacophore-based docking calculation. During the 500-ps MD simulation course, the NSC54429 tended to switch its initial orientation and inserted into two β-sheet layers, where the long axis of NSC54429 was parallel to the β-strand to maximize the hydrogen bonds with the protofibril. A recent MD study has indicated that aromatic rings of morin play an important role in driving the initial hydrophobic contacts with the Aβ protofibril, which may promote further insertion of morin into the hydrophobic core of the Aβ protofibril [13]. A similar inhibition mechanism can be observed for NSC54429, implying that NSC54429 may initially deposit on the edge of the GNNQQNY protofibril through both aromatic and hydrogen-bonding interactions, and then switch its orientation to maximize hydrogen-bonding interactions with the protofibril. The entire process allows NSC54429 to anchor to the protofibril, resulting in a more stable complex and thus preventing protofibril elongation. Thus, introducing more hydrogen bond donors or acceptors into the central parts of ligands may enhance binding affinity. These results may provide useful information for the development of new inhibitors to prevent amyloid fibril formation.

Conclusions

In this study, molecular docking, consensus scoring, MD simulations, and MM-PBSA analyses were conducted to investigate the binding sites and modes of CR in GNNQQNY protofibrils and evaluate their binding stabilities and affinities. Our results suggest four binding sites with dual binding modes of CR in protofibrils where CR can bind with its long axis parallel (as shown in complexes A, B, and C) and perpendicular (as shown in complex D) to the long axis of the fibril, which are in good agreement with previous studies [41-45].

Moreover, our results indicate that CR molecule can bind to the edge of the protofibril via hydrophobic or aromatic and hydrogen-bonding interactions to block the attachment of an incoming peptide to the protofibril. Furthermore, with the replacement of the SO_3^- groups by OCH_3 groups on CR, our simulations show that this CR-OCH_3 molecule tends to dissociate from binding site D, suggesting that this analog is unable to bind to site D and thus loses the ability to prevent protofibril from attachment of incoming peptides, which is in accordance with the experimental finding [10]. Taken

together, binding site D was suggested as a possible inhibition site for the identification of new inhibitors using a virtual screening technique. Three hit compounds, namely NSC54429, NSC9966, and NSC2237, were identified using virtual screening technique and shown to exhibit higher affinities than CR. This can be attributed to the fact that the central parts of these compounds exhibit more hydrogen bond donors and acceptors to form hydrogen-bonding interactions with the protofibril, implying that the introduction of these chemical features into central parts of compounds may enhance binding affinity. These results shed light at the atomic level on the interactions between inhibitors and amyloid protofibrils, which may provide useful insights for the design of small-molecule inhibitors of aggregation with therapeutic potential in amyloid fibril-related disease.

Acknowledgments The authors gratefully acknowledge the financial supports from Atomic Energy Council of Taiwan (Project number: ARA010203), National Science Council of Taiwan (Project numbers: 99-2221-E-027-022-MY3, 99-2221-E-027-037-MY2, and 99-2622-E-027-003-CC3), the Institute of Nuclear Energy Research of Taiwan (Project number: 1002001INER081), and National Taipei University of Technology and Taipei Medical University (Project number: NTUT-TMU-100-09).

References

1. Pepys MB (2001) Pathogenesis, diagnosis and treatment of systemic amyloidosis. Philos Trans R Soc Lond Ser B 356:203–211
2. Dobson CM (2005) Prying into prions. Nature 435:747–749
3. Makin OS, Serpell LC (2005) X-ray diffraction studies of amyloid structure. Methods Mol Biol 299:67–80
4. Makin OS, Serpell LC (2005) Structures for amyloid fibrils. FEBS J 272:5950–5961
5. Nelson R, Sawaya MR, Balbirnie M, Madsen A, Riekel C, Grothe R, Eisenberg D (2005) Structure of the cross-β spine of amyloid-like fibrils. Nature 435:773–778
6. Sawaya MR, Sambashivan S, Nelson R, Ivanova MI, Sievers SA, Apostol MI, Thompson MJ, Balbirnie M, Wiltzius JJW, McFarlane HT (2007) Atomic structures of amyloid cross-β spines reveal varied steric zippers. Nature 447:453–457
7. Esposito L, Pedone C, Vitagliano L (2006) Molecular dynamics analyses of cross-β-spine steric zipper models: β-sheet twisting and aggregation. Proc Natl Acad Sci USA 103:11533–11538
8. Howie AJ, Brewer DB (2009) Optical properties of amyloid stained by Congo red: history and mechanisms. Micron 40:285–301
9. Lorenzo A, Yankner BA (1994) β-amyloid neurotoxicity requires fibril formation and is inhibited by congo red. Proc Natl Acad Sci USA 91:12243–12247
10. Reinke AA, Gestwicki JE (2007) Structure-activity relationships of amyloid β-aggregation inhibitors based on curcumin: influence of linker length and flexibility. Chem Biol Drug Des 70:206–215
11. Raman EP, Takeda T, Klimov DK (2009) Molecular dynamics simulations of ibuprofen binding to Aβ Peptides. Biophys J 97:2070–2079
12. Chang WE, Takeda T, Raman EP, Klimov DK (2010) Molecular dynamics simulations of anti-aggregation effect of ibuprofen. Biophys J 98:2662–2670
13. Lemkul JA, Bevan DR (2010) Destabilizing Alzheimer's Aβ42 protofibrils with Morin: mechanistic insights from molecular dynamics simulations. Biochemistry 49:3935–3946
14. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. Proc Natl Acad Sci USA 89:2195–2199
15. Chen R, Weng Z (2003) A novel shape complementarity scoring function for protein-protein docking. Proteins 51:397–408
16. Wu G, Robertson DH, Brooks Iii CL, Vieth M (2003) Detailed analysis of grid-based molecular docking: a case study of CDOCKER─CHARMm-based MD docking algorithm. J Comput Chem 24:1549–1562
17. Krammer A, Kirchhoff PD, Jiang X, Venkatachalam CM, Waldman M (2005) LigScore: a novel scoring function for predicting binding affinities. J Mol Graph Model 23:395–407
18. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DR, Fogel LJ, Freer ST (1995) Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 2:317–324
19. Jain AN (1996) Scoring noncovalent protein-ligand interactions: a continuous differentiable function tuned to compute binding affinities. J Comput Aided Mol Des 10:427–440
20. Muegge I, Martin YC (1999) A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 42:791–804
21. Muegge I (2006) PMF scoring revisited. J Med Chem 49:5895–5902
22. Böhm HJ (1998) Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. J Comput Aided Mol Des 12:309–323
23. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N log (N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
24. Miyamoto S, Kollman PA (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. J Comput Chem 13:952–962
25. Wang J, Morin P, Wang W, Kollman PA (2001) Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. J Am Chem Soc 123:5221–5230
26. Rafi SB, Cui G, Song K, Cheng X, Tonge PJ, Simmerling C (2006) Insight through molecular mechanics Poisson-Boltzmann surface area calculations into the binding affinity of triclosan and three analogues for FabI, the E. coli enoyl reductase. J Med Chem 49:4574–4580
27. Nolde SB, Arseniev AS, Orekhov VY, Billeter M (2002) Essential domain motions in barnase revealed by MD simulations. Proteins 46:250–258
28. Zeng J, Li W, Zhao Y, Liu G, Tang Y, Jiang H (2008) Insights into ligand selectivity in estrogen receptor isoforms: molecular dynamics simulations and binding free energy calculations. J Phys Chem B 112:2719–2726
29. Rastelli G, Rio AD, Degliesposti G, Sgobba M (2010) Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. J Comput Chem 31:797–810
30. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25
31. Wu C, Wang Z, Lei H, Zhang W, Duan Y (2007) Dual binding modes of Congo red to amyloid protofibril surface observed in molecular dynamics simulations. J Am Chem Soc 129:1225–1232
32. Cooper JH (1974) Selective staining of amyloid as a function of amyloid composition and structure. Lab Invest 31:232–238

33. Klunk WE, Pettegrew JW, Abraham DJ (1989) Quantitative evaluation of congo red binding to amyloid-like proteins with a β-pleated sheet conformation. J Histochem Cytochem 37:1273–1281
34. Glenner GG (1980) Amyloid deposits and amyloidosis. N Engl J Med 302:1283–1292
35. Jin LW, Claborn KA, Kurimoto M, Geday MA, Maezawa I, Sohraby F, Estrada M, Kaminksy W, Kahr B (2003) Imaging linear birefringence and dichroism in cerebral amyloid pathologies. Proc Natl Acad Sci USA 100:15294–15298
36. Romhányi G (1971) Selective differentiation between amyloid and connective tissue structures based on the collagen specific topo-optical staining reaction with Congo red. Virchows Arch A Pathol Anat 354:209–222
37. Biancalana M, Makabe K, Koide A, Koide S (2009) Molecular mechanism of thioflavin-T binding to the surface of β-rich peptide self-assemblies. J Mol Biol 385:1052–1063
38. Childers WS, Mehta AK, Lu K, Lynn DG (2009) Templating molecular arrays in amyloid's cross-β grooves. J Am Chem Soc 131:10165–10172
39. Maezawa I, Hong HS, Liu R, Wu CY, Cheng RH, Kung MP, Kung HF, Lam KS, Oddo S, LaFerla FM (2008) Congo red and thioflavin-T analogs detect Aβ oligomers. J Neurochem 104:457–468
40. Chimon S, Ishii Y (2005) Capturing intermediate structures of Alzheimer's β-amyloid, Aβ (1–40), by solid-state NMR spectroscopy. J Am Chem Soc 127:13472–13473
41. Levine H 3rd (2005) Multiple ligand binding sites on Aβ (1–40) fibrils. Amyloid 12:5–14
42. Miura T, Yamamiya C, Sasaki M, Suzuki K, Takeuchi H (2002) Binding mode of Congo red to Alzheimer's amyloid-β peptide studied by UV Raman spectroscopy. J Raman Spectrosc 33:530–535
43. Elhaddaoui A, Delacourte A, Turrell S (1993) Spectroscopic study of Congo red and thioflavin binding to amyloid-like proteins. J Mol Struct 294:115–118
44. Sajid J, Elhaddaoui A, Turrell S (1997) Investigation of the binding of Congo red to amyloid in Alzheimer's diseased tissue. J Mol Struct 408:181–184
45. Pigorsch E, Elhaddaoui A, Turrell S (1994) Spectroscopic study of pH and solvent effects on the structure of Congo red and its binding mechanism to amyloid-like proteins. Spectrochim Acta A Mol Spectrosc 50:2145–2152
46. Esler WP, Stimson ER, Jennings JM, Vinters HV, Ghilardi JR, Lee JP, Mantyh PW, Maggio JE (2000) Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. Biochemistry 39:6288–6295
47. Cannon MJ, Williams AD, Wetzel R, Myszka DG (2004) Kinetic analysis of β-amyloid fibril elongation. Anal Biochem 328:67–75
48. Ban T, Hoshino M, Takahashi S, Hamada D, Hasegawa K, Naiki H, Goto Y (2004) Direct observation of Aβ amyloid fibril growth and inhibition. J Mol Biol 344:757–767
49. Porat Y, Abramowitz A, Gazit E (2006) Inhibition of amyloid fibril formation by polyphenols: structural similarity and aromatic interactions as a common inhibition mechanism. Chem Biol Drug Des 67:27–37
50. Frid P, Anisimov SV, Popovic N (2007) Congo red and protein aggregation in neurodegenerative diseases. Brain Res Rev 53:135–160