




Cite this: *Chem. Sci.*, 2017, 8, 5041

# Elucidating the 3D structures of Al(III)–A $\beta$ complexes: a template free strategy based on the pre-organization hypothesis†

Jon I. Mujika, <sup>\*a</sup> Jaime Rodríguez-Guerra Pedregal, <sup>b</sup> Xabier Lopez,<sup>a</sup> Jesus M. Ugalde,<sup>a</sup> Luis Rodríguez-Santiago, <sup>b</sup> Mariona Sodupe<sup>b</sup> and Jean-Didier Maréchal<sup>\*b</sup>

Senile plaques are extracellular deposits found in patients with Alzheimer's Disease (AD) and are mainly formed by insoluble fibrils of  $\beta$ -amyloid (A $\beta$ ) peptides. The mechanistic details about how AD develops are not fully understood yet, but metals such as Cu, Zn, or Fe are proposed to have a non-innocent role. Many studies have also linked the non biological metal aluminum with AD, a species whose concentration in the environment and food has been constantly increasing since the industrial revolution. Gaining a molecular picture of how Al(III) interacts with an A $\beta$  peptide is of fundamental interest to improve understanding of the many variables in the evolution of AD. So far, no consensus has been reached on how this metal interacts with A $\beta$ , partially due to the experimental complexity of detecting and quantifying the resulting Al(III)–A $\beta$  complexes. Computational chemistry arises as a powerful alternative to investigate how Al(III) can interact with A $\beta$  peptides, as suitable strategies could shed light on the metal–peptide description at the molecular level. However, the absence of any reliable template that could be used for the modeling of the metalloprotein structure makes computational insight extremely difficult. Here, we present a novel strategy to generate accurate 3D models of the Al(III)–A $\beta$  complexes, which still circumvents first principles simulations of metal binding to peptides of A $\beta$ . The key to this approach lies in the identification of experimental structures of the isolated peptide that are favourably pre-organized for the binding of a given metal in configurations of the first coordination sphere that were previously identified as the most stable with amino acid models. This approach solves the problem of the absence of clear structural templates for novel metalloprotein constructs. The posterior refinement of the structures *via* QM/MM and MD calculations allows us to provide, for the first time, physically sound models for Al(III)–A $\beta$  complexes with a 1 : 1 stoichiometry, where up to three carboxylic groups are involved in the metal binding, with a clear preference towards Glu3, Asp7, and Glu11.

Received 22nd March 2017  
Accepted 4th May 2017

DOI: 10.1039/c7sc01296a

rsc.li/chemical-science

## Introduction

The quest for cures of neurodegenerative diseases has become a vital field of research in our modern societies and many directions are considered to fight against them, from genetic to pharmacological approaches. Whatever the strategy, decoding key cellular and molecular mechanisms of sickness represents a fundamental objective in this field. Amongst the molecular

aspects shared by most neurodegenerative diseases is the formation of insoluble peptide aggregates. These species are formed through the association of soluble protein fragments known as amyloids. About twenty proteins are already known to form such amyloids as A $\beta$  in Alzheimer's disease (AD) or  $\alpha$ -synuclein in Parkinson's disease (PD).

In AD, two main types of deposits are found: senile plaques and neurofibrillary tangles. The former contains insoluble filaments made of  $\beta$ -amyloid fragments and the latter consists of hyperphosphorylated tau proteins. The  $\gamma$ -secretase enzyme acts on the Amyloid Precursor Protein (APP) membrane protein to produce the A $\beta$ <sub>1–40</sub> and A $\beta$ <sub>1–42</sub> fragments that are found in senile plaques. Despite the fact that the formation of these peptides is a normal process in healthy people, several factors may prompt an imbalance in their concentration and so favoring and accelerating the aggregation in the unhealthy ones. The aggregation is a complex process in which species

<sup>a</sup>Kimika Fakultatea, Euskal Herriko Unibertsitatea (UPV/EHU) and Donostia International Physics Center (DIPC), 20080 Donostia, Euskadi, Spain. E-mail: joni.mujika@ehu.eus

<sup>b</sup>Departament de Química, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. E-mail: JeanDidier.Marechal@uab.cat

† Electronic supplementary information (ESI) available: Detailed computational details, geometry data of all 3D Al(III)–A $\beta$  models, supplementary figure and table. See DOI: 10.1039/c7sc01296a



such as soluble oligomers, paranuclei, protofibrils, and finally insoluble filaments are produced.<sup>1</sup>

As it has been recently discussed in the review by Kepp,<sup>2</sup> three main hypotheses arise to explain the main factors of AD development: the amyloid cascade hypothesis, the oxidative stress hypothesis, and the metal ion hypothesis.<sup>3–5</sup> The metal ion hypothesis claims that metals naturally involved in biological systems, like Fe, Cu, and Zn, have a key role in AD and a large body of evidences indicates that they participate in different stages of the formation of amyloids and their aggregated forms.<sup>6–8</sup> As knowledge about AD increases, other metals coming from human activities are now the usual suspects in the evolution of the disease. Among them, Al(III) has become the main focus of attention. In spite of its absence from our evolutionary process, human intervention (water treatment, acid rain, drugs, *etc.*) has increased its bioavailability and its exposure into organisms in which it may interact at cellular and molecular levels.<sup>9</sup>

How aluminum is involved in AD is still unclear although it has been associated with different functions.<sup>10–13</sup> Aluminum has been detected in senile plaques extracted from the brains of patients with AD<sup>14</sup> and several *in vitro* experiments showed that Al(III) promotes aggregation more efficiently than other metals.<sup>15–19</sup> The analysis of A $\beta$  aggregates using transmission electron microscopy reveals that Al(III) favors the formation of smaller oligomers than the ones made by Cu, Zn, or Fe.<sup>19</sup> Moreover, it was shown that the A $\beta$ -Al complex alters the cell morphology and increases the membrane fluidity.<sup>19</sup> These authors indicated that the small size of the Al(III)-A $\beta$  oligomers combined with their high hydrophobicity may facilitate the crossing of the complex across the cell membrane and, once there, the A $\beta$  fragments may lead to pathological effects. In addition, it has been determined that the amount of A $\beta$  peptides crossing the blood-brain barrier increases by 60% when they are bound to Al(III).<sup>20</sup> Nevertheless, most of the studies on the Al(III)-A $\beta$  complex report whether the A $\beta$  peptide interacts with Al(III),<sup>21–24</sup> or analyzed the physiological morphology adopted by the aggregates.<sup>17,18,25–27</sup> However, to the best of our knowledge there is no investigation on the particular mode of coordination of Al(III) with A $\beta$  peptides at the molecular level. This lack of data is due in part to the inherent difficulties of characterizing the Al(III) coordination shell. In this vein, computational chemistry could become an interesting ally in shedding light on the interaction mode between Al(III) and A $\beta$  peptides.

The prediction of metal binding processes to biomolecules is one of the most challenging exercises that computation could answer. The quest is even harder when dealing with small peptides, because the variability of the first coordination sphere of the metal upon binding is complicated by the intrinsic flexibility of the polymer. Therefore, numerous intermediates could occur along the simulation of metal – binding complexation and first principles simulations are therefore extremely complex. Biased computations appear to be necessary but are also challenging since suitable approximations are needed to drive the modeling. One possible procedure to escape these limitations consists of using molecular mechanics force field-

based computational methods to reach physically sound starting point geometries of the metallopeptide, and then refine the best possible candidates with more accurate calculations like hybrid quantum mechanics/molecular mechanics approaches. The key element of this kind of procedure resides in accounting for the appropriate amount of starting point structural information to narrow the conformational space to explore.

This strategy has been our focus in several recent studies like the prediction of different copper binding A $\beta$  complexes where NMR structures of Zn(II) bound to A $\beta$  peptides were used as templates for an initial homology modeling exercise.<sup>28</sup> However, it is easy to foresee that the same protocol is not valid for building Al(III) metallopeptides. According to the HSAB theory, Zn(II) and Cu(II) are acids of intermediate strength and Al(III) is a hard acid. Consequently, the chemistry of the latter differs from the chemistry of the other two metal ions naturally occurring in biological systems; Al(III) shows a preference for negatively charged oxygen containing groups, such as phosphates and carboxylates,<sup>29</sup> while Zn(II) and Cu(II) tend to interact with N containing groups, as the well established coordination modes of these two cations to the A $\beta$  peptide demonstrate.<sup>7,30</sup> As a consequence, the unique experimental structure of the Zn-A $\beta$  complex available in the Protein Data Bank<sup>31</sup> does not represent a possible starting point for building a 1 : 1 stoichiometric fold of aluminum interacting with  $\beta$ -amyloid. This is a major missing piece of information when aiming to decode the structures of Al(III)-A $\beta$ .

A possible way forward lies in identifying pre-organized structures of the metal free peptide that could host the ion in its best possible coordination environments, and from there generate final metallopeptide models. The basic idea of this approach stands on the possibility of any receptor naturally exploring, at room temperature, microstates that are close to the metal complexed geometries. This concept is recurrent in many molecular fields nowadays from protein-ligand interactions (*i.e.* in the conformational selection hypothesis),<sup>32,33</sup> *de novo* enzyme design,<sup>34–36</sup> metal binding to proteins,<sup>37</sup> cyclic peptides,<sup>38</sup> supramolecular chemistry,<sup>39</sup> and inorganic chemistry especially when dealing with multidentate ligands.<sup>40,41</sup> Although the exact contributions of pre-organization and induced effects in the binding processes are not yet clear – likely depending on the system under investigation – the former offers a possible solution to the absence of a structural template for the Al(III)-A $\beta$  system. Indeed, plenty of structures of the sole A $\beta$  peptide presenting alternative conformations are available in the Protein Data Bank<sup>42</sup> and suitable algorithms could help in identifying those that satisfy the geometric conditions for the best arrangements of the first coordination sphere of aluminum.

Here, we perform a multi-level computational study that ranges from quantum mechanical cluster models to full size system QM/MM calculations and is based on the host-guest pre-organization hypothesis. After decoding the nature of the most stable first coordination sphere of the metal with a prototypical ligand in biological media (*i.e.* amino acids and water) *via* QM calculations, the Protein Data Bank is screened to find geometries of the peptide that could satisfy the requirement of



the best first coordination sphere environments. Then a series of plausible candidates are selected and finally refined with QM/MM approaches, hence providing a certain amount of induced fit effect. The results of our study provide key insights regarding the formation of Al(III)-A $\beta$  complexes, suggest physically sound atomistic scale models of these systems, and add to the still limited panel of molecular tools available for metalloprotein studies with a novel procedure based on an updated molecular hypothesis.

## Computational details

In this section we present our strategy to reach reliable models of Al(III)-A $\beta$ , and this consists of a multi-level protocol where accurate computational chemistry approaches are combined with structural statistic searches. Full technical details are available in the ESI.† The overall methodology (summarized in Fig. 1) can be divided into three successive steps.

### 1. Characterization of the most favorable first coordination spheres of aluminum with amino acid motifs

Only scarce information is available on how Al(III) interacts with biological motifs. Nonetheless, this is a fundamental piece of knowledge when aiming at generating trustworthy models of the interaction of this ion with A $\beta$  peptides. Therefore, a series of first coordination sphere hypotheses were explored *via* accurate quantum mechanics methods using a cluster model

approach. Numerous models that account for the side chain, main chain, and water coordination with Al(III) throughout nitrogen and oxygen coordination were optimized and their energies compared. Only the few lowest energy systems were considered as possible environments for an *in vivo* condition for the full length Al(III)-A $\beta$  system.

### 2. Selection of full length amyloid structures compatible with optimum aluminum first coordination spheres

Hundreds of NMR and X-ray structures of peptides, including A $\beta$  forms, can be found in the Protein Data Bank.<sup>42</sup> However, experimental structures of metal-bound peptides are scarce and to the best of our knowledge only a few Zn(II)-A $\beta$  systems can be found in this database.<sup>43–45</sup> As direct metal mediated folding of a peptide is not yet viable by computational means (*i.e.* MD) we decided to look for amyloid structures that display convenient scaffolds to host an octahedral aluminum. By doing this we aim to identify, within the ensemble of experimental structures available in the PDB, those structures that present the relevant pre-organization of the peptide to satisfy the Al(III) environment of lower energy, as detected in step one. This was performed by: (1) selecting A $\beta$  peptides from the experimental PDB bank database,<sup>42</sup> (2) filtering of the structures *via* a grid protocol to detect suitable arrangements of backbone atoms for coordinating an Al(III), keeping the coordination mode predicted by cluster model calculations, and (3) generating refined 3D candidates for the Al(III)-A $\beta$  structure. For the last point, we set

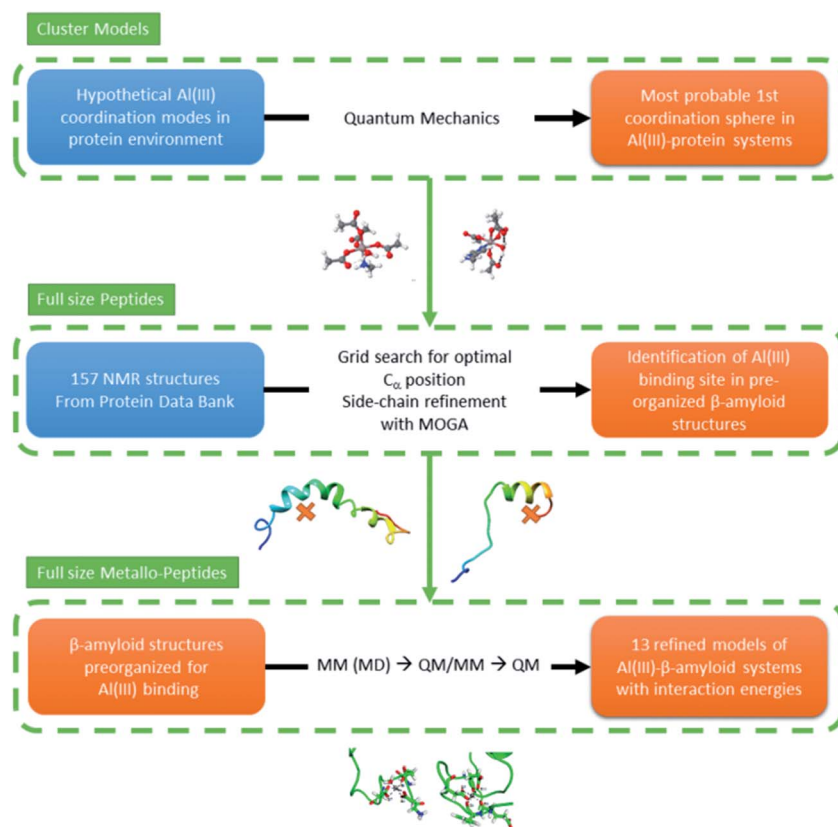


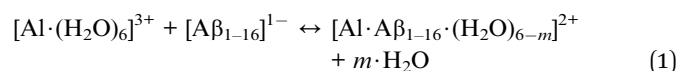
Fig. 1 Schematic illustration of the strategy followed throughout the present work.



up and applied an in-house made genetic algorithm to identify the best position of the metal and the correct orientation of the side chains to generate an adequate Al(III)-A $\beta$  model.

### 3. Refinement of the Al(III)-A $\beta$ structures and calculation of interaction energies

Once a set of Al(III)-A $\beta$  models were built, their structures were equilibrated using molecular dynamics simulations. Three structures from each MD simulation were subsequently minimized using a QM/MM protocol. Afterwards, the interaction energy between Al(III) and the A $\beta$  peptide of each complex structure was determined using pure QM calculations. Since in all of the complexes the 1–11 segment of the peptide accommodates the cation binding site, the 1–16 segment of the A $\beta$  peptide was taken to calculate the interaction energy according to reaction (1), where A $\beta$ <sub>1–16</sub> corresponds to the 1–16 fragment of the apoform of the A $\beta$  peptide, [Al·A $\beta$ <sub>1–16</sub>·(H<sub>2</sub>O)<sub>6–m</sub>] is the complex formed by the fragment with Al(III), and *m* refers to the number of water molecules displaced by the A $\beta$  peptide from [Al·(H<sub>2</sub>O)<sub>6</sub>]. Notice that the structure of the apoform was taken from the complex, and therefore reaction (1) accounts for the interaction energy between the A $\beta$  peptide and Al(III) in the [Al·A $\beta$ <sub>1–16</sub>·(H<sub>2</sub>O)<sub>6–m</sub>] complex structure.



## Results

### 1. Small QM cluster models

The preferential coordination mode of Al(III) was investigated by constructing a series of cluster systems where Al(III) interacts with alternative arrangements of the amino acids of the first coordination sphere. Due to its hard Lewis character, Al(III) markedly shows a preference towards ligands that donate electrons,<sup>46</sup> and in particular towards O-containing amino acids, especially carboxylic acids.<sup>47</sup> Thus, the following building blocks (summarized in Fig. 2) were combined to build the model systems: (a) acetate (C) to mimic aspartic or glutamic

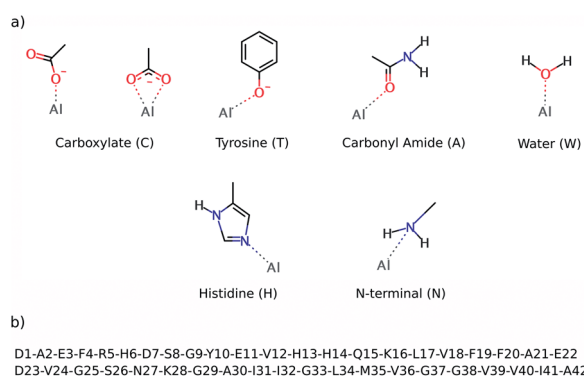
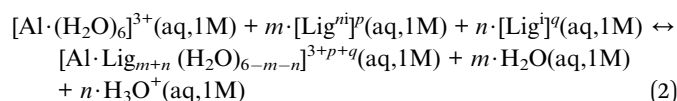


Fig. 2 The top image (a) shows the ligands employed to build alternative QM cluster models for the Al(III) coordination first shell. The below image (b) shows the sequence of the A $\beta$ <sub>1–42</sub> peptide.

acid (both mono- and bidentated to the cation), (b) phenol for tyrosine (T), (c) acetamide for a peptide bond carbonyl amide (A), (d) water (W), (e) 4(5)-methylimidazole for histidine (H), and (f) methylamine for the N-terminal (N). In addition, due to the flexible coordination mode of Al(III),<sup>48</sup> the 4, 5 and 6 coordination numbers were considered. In order to reduce the number of possibilities, at least two non-water ligands were included in the model, with at least two carboxylic groups. The stability of each complex was determined by calculating the reaction energy of reaction (2), where *m* refers to the number of ligands which are not ionized (Lig<sup>ni</sup>) upon their complexation to Al(III), while *n* refers to the number of ligands that are considered neutral in solution and unprotonated in the complex (Lig<sup>i</sup>, see below); *p* and *q* refer to the total charge of Lig<sup>ni</sup> and Lig<sup>i</sup>, respectively. The electronic energy ( $\Delta E$ ) and enthalpy ( $\Delta H^{298}$ ) values of the most stable structures characterized are presented in Table 1 and the four most stable structures are shown in Fig. 3.



The reactants of reaction (2) were considered to be infinitely separated, so the number of ligands enclosed in the complex has a direct consequence for its entropy value, which led to overestimation of the entropic effects. This is especially true in a biological environment, where the reactants are not infinitely separated. Moreover, the calculations on cluster models do not account for the entropic effects ascribed to the conformational change associated with the peptide. So, due to all these factors, the enthalpy values ( $\Delta H^{298}$ ) were used to sort out the stability of all of the cluster models characterized. The results confirm that Al(III) shows a clear preference for carboxylic groups, as four carboxylic groups are present in most of the structures shown in

Table 1 Interaction energies ( $\Delta E$ ) and enthalpies ( $\Delta H^{298}$ ) (in kcal mol<sup>−1</sup>) between Al(III) and its coordination sphere computed on DFT/PCM cluster models and evaluated according to eqn (2). The interaction modes of the carboxylic groups are specified as M (monodentated) or B (bidentated). The number of ligands forming each model and the total charge of the system are also specified

Model	Carb.	#Ligand	Charge	$\Delta E$	$\Delta H^{298}$
4CWA	4M	6	−1	−124.8	−126.1
4CWH	4M	6	−1	−124.1	−125.5
4CW	3M1B	5	−1	−120.7	−125.0
4CW	4M	5	−1	−117.1	−120.4
4CH	3M1B	5	−1	−114.0	−117.8
4C2A	4M	6	−1	−115.0	−117.4
3CWHA	3M	6	0	−115.7	−116.9
4C2H	4M	6	−1	−114.8	−116.5
4CA	4M	5	−1	−112.3	−115.6
4CA	3M1B	5	−1	−112.0	−115.1
3C2WH	3M	6	0	−112.8	−113.7
3C2HA	3M	6	0	−112.1	−113.6
4C	4M	4	−1	−108.6	−113.4
4C	2M2B	4	−1	−108.6	−113.3



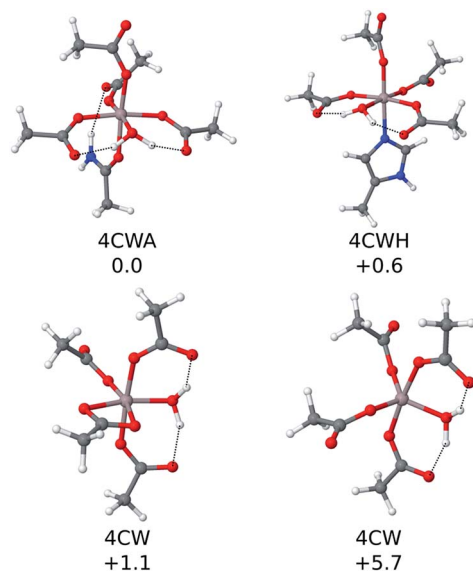


Fig. 3 The four most stable QM cluster model structures that were characterized. Their relative  $\Delta\Delta H^{298}$  values (in kcal mol<sup>-1</sup>) are included.

Table 1. Note that tyrosine is the other negatively charged ligand, but its coordination requires being deprotonated first with a subsequent deprotonation energy penalty ( $\Delta G^{\text{deprot}}$  defined in the ESI<sup>†</sup>), so its coordination to Al(III) is energetically less favorable than the coordination of a carboxylic group.

The first coordination shell of the four most stable structures (illustrated in Fig. 3), apart from the aforementioned four carboxylic groups, includes one water molecule. This molecule provides extra stability, as it can form strong hydrogen bond interactions with the two carboxylic groups monodentately bound to Al(III). In particular, the most stable structure (4CWA) presents a binding site with four monodentated carboxylic groups, a water molecule and a carbonyl oxygen atom completing the coordination shell of the metal. A similar coordination mode is found in 4CWH, the second most stable structure, where a histidine replaces the carbonyl group. The same 4CW motif is found in the next two structures, and they differ in the coordination mode of one of the carboxylic groups; the bidentated one is about 5 kcal mol<sup>-1</sup> more stable than the monodentated one. In both structures, the remaining three carboxylic groups interact monodentately.

Interestingly, the results point to a penta- or hexacoordinated Al(III) ion. The  $\Delta H^{298}$  values of the structures with only 4C groups in the coordination sphere (for instance, 4M and 2M2B structures, the latest two entries of Table 1) are *ca.* -113 kcal mol<sup>-1</sup>, significantly less stable than the structures where the metal coordination shell is completed with additional ligands. Thus, the addition of more ligands to the four carboxylic groups provides an additional stabilization to the system. The carbonyl group appears very often as one of the ligands completing the metal coordination shell. Histidine is also found in some structures.

In summary, the cluster models indicate that the most stable binding site of Al(III) should provide either three or four

carboxylates. The nature of the remaining ligands has a smaller effect, with a water molecule, a carbonyl group, or a histidine being the best candidate. However, it must be pointed out that even though the characterization of the metal binding sites using cluster models provides useful information about the preferred binding site of Al(III) to an A $\beta$  peptide, these types of calculations do not take into account some key effects, such as the influence of the residues located beyond the first coordination shell, the peptide's conformation, or the entropic effects. The full Al(III)-A $\beta$  structure is required to account for these effects (*vide supra*).

## 2. Full size 3D models of the Al(III)-A $\beta$ complex

**Building of reliable structures.** The DFT/PCM calculations on cluster models shown in the preceding subsection provide us with the preferential coordination modes of Al(III) in a protein-like environment. However, the effects of the entire A $\beta$  peptide are not included in these models, even though they could drastically modulate the relative order of energy between the different coordination modes and tune the *in silico* predicted best metal-peptide complexes. Thus, the next step along our multi-level strategy consists of building realistic 3D Al(III)-A $\beta$  amyloid candidates.

Solved structures of A $\beta$  peptides in solution were extracted from the Protein Data Bank (PDB)<sup>42</sup> and were checked to see to what degree they could adapt to Al(III). A total of 157 conformations obtained from 13 different NMR samples was considered (summarized in Table S1<sup>†</sup>). In these structures the A $\beta$  peptide presents a large variety of conformations, thus providing a larger number of templates with which to build a reliable initial structure for the Al(III)-A $\beta$  complex. Next, a grid search was carried out on all of these structures of the A $\beta$  peptide with the aim of finding a conformation(s) that can be taken as pre-organized enough to be coordinated by Al(III). The grid exploration identified 194 sites in which Al(III) could have a viable environment with at least three different amino acid side-chains located consistently with the cluster model results. Note that the C $_{\alpha}$  atoms of the residues were taken as references to give more flexibility to the system (see Computational details in the ESI<sup>†</sup>). Interestingly, only O-containing amino acids (or water molecules) are found in the vicinity of each of the grid points (potential metal sites) and no nitrogen atoms from the backbone or side chains were identified.

After having reduced the group of possible geometries viable for aluminum binding, a Multi-Objective Genetic Algorithm (MOGA) was applied. As usual with MOGA techniques a high number of potential solutions were thrown (over 37 000 results). Those solutions were filtered in such a way that the maximum clashes do not overcome 30 nm<sup>3</sup>; at least three nitrogen or oxygen atoms would remain in the 2.5 Å from the grid point. Using this criterion the final solutions were reduced to 13, which were then analyzed. Surprisingly, while our algorithm gives the possibility of nitrogenated groups interacting with the metal, none of the 13 solutions present Al(III)-nitrogen bonds, and aluminium interacts only with oxygen groups. The oxygen atoms can be either part of the Asp or Glu side chains, although



they sometimes belong to the backbone. From this initial part of the modeling of the entire metalloprotein it became clear that the Protein Data Bank already contains geometries of the peptides that are sufficiently pre-organized to satisfy the geometric and coordination requirements of octahedral Al(III) with either aspartates or glutamates coordinated to it.

**Final models and interaction energies.** The 13 structures chosen from the MOGA search were first solvated with explicit water molecules and then equilibrated using a 10 ns long molecular dynamics simulation. Once equilibrated, three geometries from the MD simulations (two intermediate structures and the last structure of the simulations) were fully optimized *via* a hybrid QM/MM scheme. Finally, the interaction

energies between Al(III) and the A $\beta$  peptide were estimated by performing single-point calculations of the three QM/MM optimized structures, and their average value was computed (more details in ESI†). The computed interaction energies along with the most stable complexes are shown in Fig. 4. For all of the structures characterized, the interaction energies computed on them, along with the distances between Al(III) and the ligands located in the cation first coordination shell, are listed in Tables S2, S3, and S4†. Note that the structures are labelled as ComplX, where X indicates their positions in the stability, that is, Compl1 is the most stable one, Compl2 the second most stable, *etc.* The superscripts refer to the residues located in the Al(III) first coordination shell of the corresponding complex.

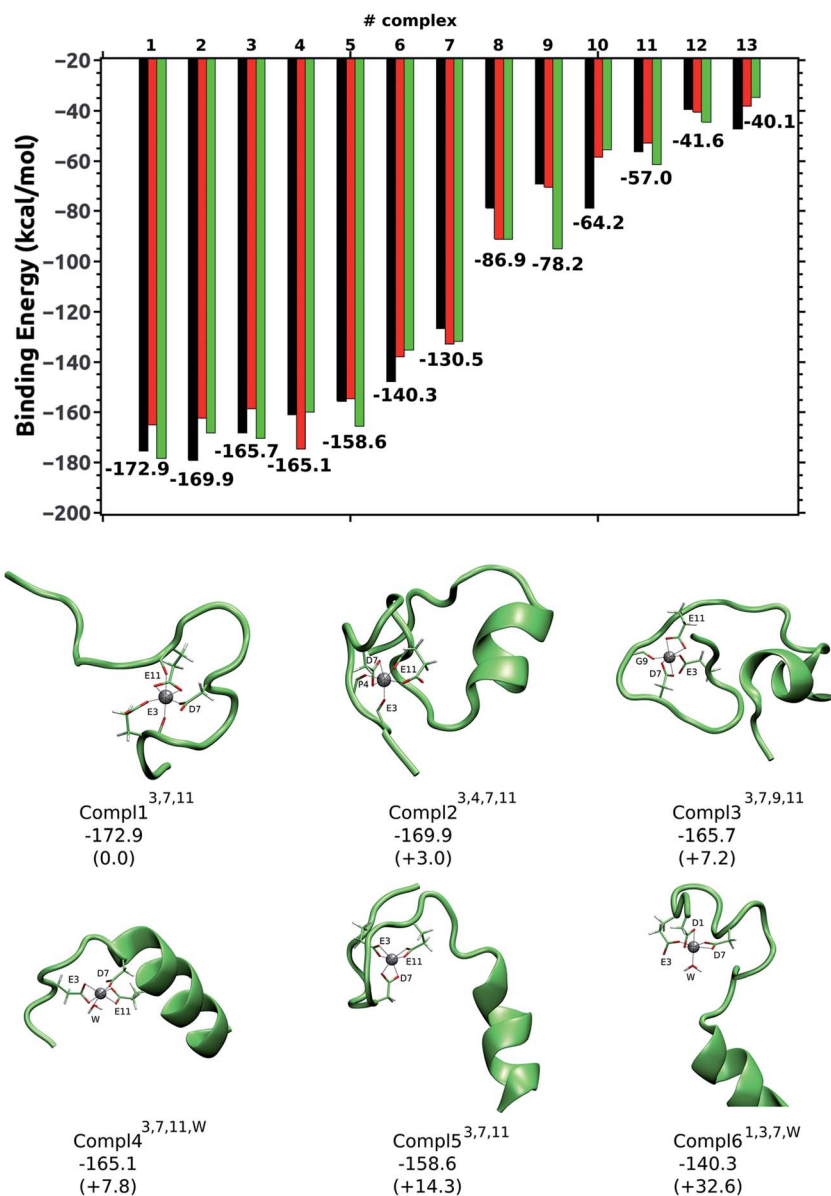


Fig. 4 Top image: the interaction energies computed on the three structures optimized for each of the 13 Al(III)–A $\beta$  complexes. The average interaction energies (in kcal mol<sup>-1</sup>) computed on the three structures are also shown. Lower image: the six most stable Al(III)–A $\beta$  structures characterized. The interaction energies between Al(III) and A $\beta$  peptide and their relative differences (in parentheses) are also shown.



Al(III) presents different coordination modes in the 13 structures, which provides a good scenario to infer the preferential binding mode of the cation. However, all of these structures of the Al(III)-A $\beta$  complex share a similar pattern for the metal; in all of them Al(III) appears hexacoordinated (except in Compl9<sup>7,11,W</sup>) and interacts only with O-ligands. Moreover, in all of the structures only four of the six carboxylic groups found in the A $\beta$ <sub>1-42</sub> sequence (shown in Fig. 2) interact with Al(III), namely Asp1, Glu3, Asp7, and Glu11. Thus, the other two carboxylic groups, Glu22 and Asp23, are predicted not to interact with the cation.

The five most stable structures shown in Fig. 4 present a remarkably more stable  $\Delta E_{\text{aq}}$  value than the remaining structures. The energy difference ( $\Delta\Delta E_{\text{aq}}$ ) between these five structures is less than 15 kcal mol<sup>-1</sup>, while the  $\Delta\Delta E_{\text{aq}}$  values of the next two structures are 32.6 kcal mol<sup>-1</sup> (Compl6<sup>1,3,7,W</sup>) and 42.4 kcal mol<sup>-1</sup> (not shown). Finally, the  $\Delta\Delta E_{\text{aq}}$  values of the remaining complexes are above 85 kcal mol<sup>-1</sup> (see ESI†). Therefore, the five most stable complexes are clearly the most representative ones for the Al(III)-A $\beta$  complex. Interestingly, the same three residues (Glu3, Asp7, and Glu11) appear in the Al(III) first coordination shell of these five structures so they are the most likely residues to be coordinated by Al(III). Among them, Compl1<sup>3,7,11</sup> is the most stable structure, with an interaction energy of -172.9 kcal mol<sup>-1</sup>. There, Glu3 and Asp7 interact monodentately with the cation, while Glu11 does bidentately. The peptide bond carbonyl oxygens of Glu3 and Glu11 also interact with Al(III). The shortest distances with respect to Al(III) correspond to O<sub>δ2</sub>@Asp7 (1.81 Å) and O<sub>ε2</sub>@Glu3 (1.86 Å), while the longest ones correspond to O@Glu11 (2.01 Å) and O<sub>ε2</sub>@Glu11 (2.01 Å). A slightly different arrangement is found in Compl2<sup>3,4,7,11</sup>, with a  $\Delta E_{\text{aq}}$  value of -169.9 kcal mol<sup>-1</sup>. Even though the carboxylic groups of Asp7 and Glu11 interact with Al(III), Glu3 only does so through its peptide carbonyl oxygen, while its carboxylic groups are facing the solvent. Two more carbonyl oxygen atoms (Phe4 and Glu11) complete the metal first coordination shell. In Compl3<sup>3,7,9,11</sup>, with a  $\Delta E_{\text{aq}}$  value of -165.7 kcal mol<sup>-1</sup>, the same three residues (Glu3, Asp7, and Glu11) are coordinated to the cation, but in this case Glu3 binds monodentately, and Asp7 and Glu11 bidentately. The carbonyl oxygen of Gly9 also interacts with Al(III).

The carboxylic groups of Glu3, Asp7, and Glu11 are also involved in the coordination of Al(III) in Compl4<sup>3,7,11,W</sup> and Compl5<sup>3,7,11</sup>, with  $\Delta E_{\text{aq}}$  values of -165.1 and -158.6 kcal mol<sup>-1</sup>, respectively. The only difference between them, apart from the conformation of the A $\beta$  peptide, is that the three carboxylic groups interact bidentately in Compl5<sup>3,7,11</sup>, while in Compl4<sup>3,7,11,W</sup> Asp7 does monodentately and the free coordination place is occupied by a water molecule.

Compl6<sup>1,3,7,W</sup> and Compl7<sup>1,3,11</sup> are the first structures where the Asp1 amino acid interacts with Al(III), along with Glu3 and Asp7 or Glu11. Thus, the coordination modes of Al(III) in Compl6<sup>1,3,7,W</sup> and Compl7<sup>1,3,11</sup> resemble the ones found in Compl4<sup>3,7,11,W</sup> and Compl5<sup>3,7,11</sup>, respectively. However, the substitution of Glu11 or Asp7 by Asp1 has increased their  $\Delta E_{\text{aq}}$  values substantially (more than 25 kcal mol<sup>-1</sup>), suggesting that

the inclusion of Asp1 in the metal first coordination shell destabilizes the complex.

In the remaining structures, which clearly show less favorable interaction energies ( $\Delta\Delta E_{\text{aq}} > 85$  kcal mol<sup>-1</sup>), only two carboxylic groups are coordinated to Al(III), while the coordination shells are fulfilled by two or three water molecules. The large gap in the  $\Delta E_{\text{aq}}$  values of these two sets of structures agrees with the cluster model calculations, indicating that the reduction in the number of carboxylic groups interacting with Al(III) has a drastic effect on the stability of the Al(III)-A $\beta$  complex, and that three carboxylic groups interacting with Al(III) provides good stability to the complex.

It is noteworthy that in the most stable 3D models Al(III) is coordinated into the 1-16 segment of the A $\beta$  peptide, as Cu(II) and Zn(II) do. However, coordination features obtained for Al(III)-A $\beta$  are significantly different to those determined for Cu(II)-A $\beta$ , which involves His6, His13 or His14, carbonyl oxygen, and the terminal NH<sub>2</sub> in component I or His6, His13, His14 and CO(Ala2) for Component II.<sup>7,30</sup> Thus, all of the carboxylate groups are exposed to the solvent in the Cu(II)-A $\beta$  complex, while in Al(III)-A $\beta$  three of them are interacting with the cation. This difference is expected to shape the aggregation pattern of each complex and accounts for the different morphologies shown by the aggregates grown in the presence of each of these cations.<sup>19</sup>

## Conclusions

By translating the pre-organization concept of host-guest interaction for the binding of metal to peptide, we here present relevant insights on the interaction of Al(III) with A $\beta$  and overcome the major limitation so far: the absence of any reliable template experimentally available of Al(III) interacting with middle sized A $\beta$  peptides. To reach final physically sound models, different computational approaches have been combined and this allows us to shed light on the preferential interaction mode between Al(III) and an A $\beta$  peptide and present, for the first time, 3D models of the complex. Remarkably, the results provide a detailed description at the atomistic scale of the Al(III)-A $\beta$  complex.

First, a large set of DFT cluster models were built to model alternative binding sites of Al(III) and to infer its intrinsic coordination mode preferences. Al(III) shows a marked preference for carboxylic groups and up to four of this type of group can interact with the cation in the most stable structures. The results also indicate that the metal coordination shell is completed by water molecules or amide bond carbonyl oxygens. All of these data provide useful information for proceeding to build reliable Al(III)-A $\beta$  complexes.

Next, a large number of structures of A $\beta$  peptides in solution, solved by NMR, were used as tentative models for building the structures of the Al(III)-A $\beta$  complex. This large set of geometries provide enough sampling of different conformations of the A $\beta$  peptide for finding pre-organized conformations of the peptide for adequately coordinating an Al(III) cation. These structures were filtered using a 3D grid algorithm and then coordination rules were applied with the MOGA algorithm. The emerged structures were further refined by MD and QM/MM calculations



in order to obtain reliable geometries of the complex and evaluate the interaction energies between Al(III) and the A $\beta$  peptide on them. This protocol gave rise to 13 model structures as the most suited to represent the Al(III)–A $\beta$  complex.

The analysis of these 13 structures indicates that the number of carboxylic groups determines the interaction energy between Al(III) and A $\beta$ , as those models with three carboxylic groups in the Al(III) coordination shell are clearly the most stable ones. The study predicts that Glu3, Asp7, and Glu11 are the most suited amino acids to be coordinated by Al(III), and that some peptide bond carbonyl oxygens may fill the metal coordination shell.

The study therefore proposes, for the first time, a 3D model for the Al(III)–A $\beta$  complex, and identifies the amino acids involved in the metal binding. The proposed coordination mode of Al(III) to A $\beta$  differs from the ones established for other cations. We believe that this is key information for investigating the biophysics of the Al(III)–A $\beta$  complex, which is fundamental to understand the biochemical role of Al(III) in AD.

Additionally, the present work gives a proof-of-concept of an original multi-level computational pipeline. It overcomes the numerous barriers resulting from the very limited three dimensional information that was available at the beginning of the project. Using the hypothesis that close-to-native metallopeptide geometries are naturally explored by the unbounded form of the biopolymer, this strategy only relies on identifying those geometries inside the conformational ensemble provided by experimental structures of the A $\beta$  peptide that are available in the Protein Data Bank, something reminiscent to strategies used for *de novo* enzyme design. Looking for such pre-organized forms of the coordinating molecule allows the issue of the lack of structural information on the Al(III) metallopeptide to be solved and avoids the necessity of using a metallopeptide resulting from the fixation of chemically different metals as templates. Such a strategy has shown extremely encouraging results, and its applicability to other kinds of ensembles (*i.e.* like those obtained by long range molecular dynamics) is now under investigation. We expect that this could shed light on other molecular systems for which 3D models are difficult to generate *via* standard methodologies.

## Acknowledgements

JIM, XL and JMU recognize the Basque Government (GIC IT588-13). The authors gratefully acknowledge financial support from MINECO (CTQ2012-38496-C05-04, CTQ2014-54071-P, CTQ2014-59544-P and CTQ2015-67608-P projects). The Generalitat de Catalunya is also acknowledged for financial support through the 2014SGR-482 and 2014SGR989 grants and the JRG FI grant. MS acknowledges the Generalitat de Catalunya for the 2011 ICREA Academia award. The SGI/IZO-SGIker UPV/EHU is also acknowledged for computational resources.

## Notes and references

- 1 R. Roychaudhuri, M. Yang, M. M. Hoshi and D. B. Teplow, *J. Biol. Chem.*, 2009, **284**, 4749–4753.
- 2 K. P. Kepp, *Chem. Rev.*, 2012, **112**, 5193–5239.

- 3 M. Necula, R. Kaye, S. Milton and C. G. Glabe, *J. Biol. Chem.*, 2007, **282**, 10311–10324.
- 4 J. Hardy and D. J. Selkoe, *Science*, 2002, **297**, 353–356.
- 5 R. J. Castellani, H. Lee, X. Zhu, G. Perry and M. A. Smith, *J. Neuropathol. Exp. Neurol.*, 2008, **67**, 523–531.
- 6 A. I. Bush, *Trends Neurosci.*, 2003, **26**, 207–214.
- 7 P. Faller and C. Hureau, *Dalton Trans.*, 2009, 1080–1094.
- 8 H. Kozłowski, A. Janicka-Klos, J. Brasun, E. Gaggelli, D. Valensin and G. Valensin, *Coord. Chem. Rev.*, 2009, **253**, 2665–2685.
- 9 A. I. Pogue and W. J. Lukiw, *Front. Neurol.*, 2014, **5**, 7–10.
- 10 S. Bhattacharjee, Y. Zhao, J. M. Hill, M. E. Percy and W. J. Lukiw, *Front. Aging Neurosci.*, 2014, **6**, 2013–2015.
- 11 V. B. Gupta, S. Anitha, M. L. Hegde, L. Zecca, R. M. Garruto, R. Ravid, S. K. Shankar, R. Stein, P. Shanmugavelu and K. S. J. Rao, *Cell. Mol. Life Sci.*, 2005, **62**, 143–158.
- 12 L. Tomljenovic, *J. Alzheimer's Dis.*, 2011, **23**, 567–598.
- 13 P. Zatta, R. Lucchini, S. van Rensburg and A. Taylor, *Brain Res. Bull.*, 2003, **62**, 15–28.
- 14 S. Yumoto, S. Kakimi, A. Ohsaki and A. Ishikawa, *J. Inorg. Biochem.*, 2009, **103**, 1579–1584.
- 15 M. Kawahara, M. Kato and Y. Kuroda, *Brain Res. Bull.*, 2001, **55**, 211–217.
- 16 M. Kawahara, K. Muramoto, K. Kobayashi, H. Mori and Y. Kuroda, *Biochem. Biophys. Res. Commun.*, 1994, **198**, 531–535.
- 17 F. Ricchelli, D. Drago, B. Filippi, G. Tognon and P. Zatta, *Cell. Mol. Life Sci.*, 2005, **62**, 1724–1733.
- 18 P. W. Mantyh, J. R. Ghilardi, S. Rogers, E. Demaster, C. J. Allen, E. R. Stimson and J. E. Maggio, *J. Neurochem.*, 1993, **61**, 1171–1174.
- 19 D. Drago, M. Bettella, S. Bolognin, L. Cendron, J. Scancar, R. Milacic, F. Ricchelli, A. Casini, L. Messori, G. Tognon and P. Zatta, *Int. J. Biochem. Cell Biol.*, 2008, **40**, 731–746.
- 20 W. A. Banks, M. L. Niehoff, D. Drago and P. Zatta, *Brain Res.*, 2006, **1116**, 215–221.
- 21 S. B. Vyas and L. K. Duffy, *J. Protein Chem.*, 1995, **14**, 633–644.
- 22 S. B. Vyas and L. K. Duffy, *Biochem. Biophys. Res. Commun.*, 1995, **206**, 718–723.
- 23 D. Valensin, C. Migliorini, G. Valensin, E. Gaggelli, G. La Penna, H. Kozłowski, C. Gabbiani and L. Messori, *Inorg. Chem.*, 2011, **50**, 6865–6867.
- 24 C. Exley, N. C. Price, S. M. Kelly and J. D. Birchall, *FEBS Lett.*, 1993, **324**, 293–295.
- 25 S. Bolognin, L. Messori, D. Drago, C. Gabbiani, L. Cendron and P. Zatta, *Int. J. Biochem. Cell Biol.*, 2011, **43**, 877–885.
- 26 W.-T. Chen, Y.-H. Liao, H.-M. Yu, I. H. Cheng and Y.-R. Chen, *J. Biol. Chem.*, 2011, **286**, 9646–9656.
- 27 E. House, J. Collingwood, O. Khan, A. Korchazkina, G. Berthon and C. Exley, *J. Alzheimer's Dis.*, 2004, **6**, 291–301.
- 28 J. Ali-Torres, J.-D. Marechal, L. Rodriguez-Santiago and M. Sodupe, *J. Am. Chem. Soc.*, 2011, **133**, 15008–15014.
- 29 R. B. Martin, *Clin. Chem.*, 1986, **32**, 1797–1806.
- 30 S. C. Drew and K. J. Barnham, *Acc. Chem. Res.*, 2011, **44**, 1146–1155.





- 31 S. Zirah, S. A. Kozin, A. K. Mazur, A. Blond, M. Cheminant, I. Segalas-Milazzo, P. Debey and S. Rebuffat, *J. Biol. Chem.*, 2006, **281**, 2151–2161.
- 32 G. G. Hammes, Y.-C. Chang and T. G. Oas, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 13737–13741.
- 33 M. A. Morando, G. Saladino, N. D'Amelio, E. Pucheta-Martinez, S. Lovera, M. Lelli, B. López-Méndez, M. Marenchino, R. Campos-Olivas and F. L. Gervasio, *Sci. Rep.*, 2016, **6**, 24439.
- 34 S. Osuna, G. Jiménez-Osés, E. L. Noey and K. N. Houk, *Acc. Chem. Res.*, 2015, **48**, 1080–1089.
- 35 A. J. T. Smith, R. Müller, M. D. Toscano, P. Kast, H. W. Hellinga, D. Hilvert and K. N. Houk, *J. Am. Chem. Soc.*, 2008, **130**, 15361–15373.
- 36 M. P. Frushicheva, J. Cao and A. Warshel, *Biochemistry*, 2011, **50**, 3849–3858.
- 37 O. Maglio, F. Nistri, V. Pavone, A. Lombardi and W. F. DeGrado, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3772–3777.
- 38 A. E. Wakefield, W. M. Wuest and V. A. Voelz, *J. Chem. Inf. Model.*, 2015, **55**, 806–813.
- 39 J. B. Wittenberg and L. Isaacs, in *Supramolecular Chemistry*, ed. P. A. Gale and J. W. Steed, John Wiley & Sons, Ltd, Chichester, UK, 2012.
- 40 B. W. Parks, R. D. Gilbertson, J. E. Hutchison, E. R. Healey, T. J. R. Weakley, B. M. Rapko, B. P. Hay, S. I. Sinkov, G. A. Broker and R. D. Rogers, *Inorg. Chem.*, 2006, **45**, 1498–1507.
- 41 F. Cisnetti, J.-D. Maréchal, M. Nicaise, R. Guillot, M. Desmadril, F. Lambert and C. Policar, *Eur. J. Inorg. Chem.*, 2012, **2012**, 3308–3319.
- 42 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- 43 A. N. Istrate, S. A. Kozin, S. S. Zhokhov, A. B. Mantsyzov, O. I. Kechko, A. Pastore, A. A. Makarov and V. I. Polshakov, *Sci. Rep.*, 2016, **6**, 21734.
- 44 A. N. Istrate, P. O. Tsvetkov, A. B. Mantsyzov, A. A. Kulikova, S. A. Kozin, A. A. Makarov and V. I. Polshakov, *Biophys. J.*, 2012, **102**, 136–143.
- 45 S. Zirah, S. A. Kozin, A. K. Mazur, A. Blond, M. Cheminant, I. Segalas-Milazzo, P. Debey and S. Rebuffat, *J. Biol. Chem.*, 2006, **281**, 2151–2161.
- 46 A. E. Martell, R. D. Hancock, R. M. Smith and R. J. Motekaitis, *Coord. Chem. Rev.*, 1996, **149**, 311–328.
- 47 E. Rezabal, J. M. Mercero, X. Lopez and J. M. Ugalde, *ChemPhysChem*, 2007, **8**, 2119–2124.
- 48 W. J. Yang, Z. S. Qian, Q. Miao, Y. J. Wang and S. P. Bi, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2396–2401.

