Self-assembling nanosystems: Collagen, an escape strategy from plaque formation

Villő Katalin Pálfi

PhD Thesis

Supervisor: Prof. Dr. András Perczel

Program: Synthetic Chemistry, Material Sciences, Biomolecular Chemistry
Head of the Program: Prof. Dr. István Tamás Horváth

Doctorate School of Chemistry
Head of the Doctorate School: Prof. Dr. György Inzelt

Eötvös Loránd University
Institute of Chemistry

2008.

Introduction

Protein aggregates occur in both illnesses as well as in a healthy body. The most prominent case for the former is amyloid and collagen for the latter. The first type of aggregate formation from amyloid and other proteins cannot be controlled (up) to our knowledge, and the resulting structure can not be dismantled by the organism. Collagen formation, however, is thoroughly regulated, and the resulting structure – as can be observed for example in the bone – can also be degraded.

Dozens of ordinary proteins (*e.g.* SH3, β-2 microglobulin, lysozyme, myoglobin) tend to aggregate if misfolded under abnormal cellular conditions, producing an architecture similar to the amyloid peptide, which is responsible for the development of Alzheimer's disease. Similar aggregates play a role in the case of the Creutzfeldt-Jacob disease and Huntington chorea.

These protein aggregates have the same macroscopic structure, so-called amyloid fibrils, and are supposed to have the same or at least similar molecular structure. It is thought that these aggregates are rich in β-sheet structures, where in one peptide chain there are two elongated parts connected by a turn region, and peptide chains are placed parallelly in an endless crystal. In the correct conditions, practically any protein investigated could be trapped in such a toxic β-layer form. Therefore the following question arises: why is the β-sheet structure so stable?

Collagen is an essential extracellular protein. It provides the quarter of the mass of all the proteins in the human body. It is a major structural protein, forming molecular cables that strengthen the tendons and sheets that support the skin and internal organs. Its build-up is hierarchical, the first level is tropocollagen, which consists of three protein chains that are self-assembled into a triple-helical structure. Furthermore, seven tropocollagen triple helices form microfibrils in a hexagonal arrangement.

The triple helix of tropocollagen is built up of almost identical conformational elements (homoconformers) often referred to as the polyproline II or shortly PPII conformation with the following parameters: $\phi \approx -70°$, $\psi \approx +160°$ (ε_L), similar those of β-pleated sheet type aggregates, β_L, ($\phi \approx -150°$, $\psi \approx +150°$). The primary structure of collagen is characterized by the repetitive motif Xxx-Yyy-Gly, (triplet) where the Xxx and Yyy positions are typically occupied by proline (Pro) and hydroxyproline (Hyp or O in short) residues, respectively. This

triple helix is highly hydrated, at least 4 or 5 water molecules per triplet are required to obtain the regular fold. The atomic structure of the water shell around collagen was first described meticulously by Bella et al. and was observed in several other X-ray structures as well. Some water molecules are not just absorbed on the surface of tropocollagen but also maintain a regular pattern. Hydrogen-bond mediated water chains connect different parts of tropocollagens. We refer to these water-chains as "water bridges" when the first and the last water molecules of the chain are connected to the same tropocollagen unit. The first hydration shell of the POG-type collagen consists of minimum 6 and maximum 9 water molecules per triplet unit, of which 5 is attached directly to a collagen atom by a H-bond.

Bella et al. described four different types of water bridges, called α, β, γ and δ according to the type of atoms they are attached to on the surface of the collagen triple-helix. **(Figure 1)**

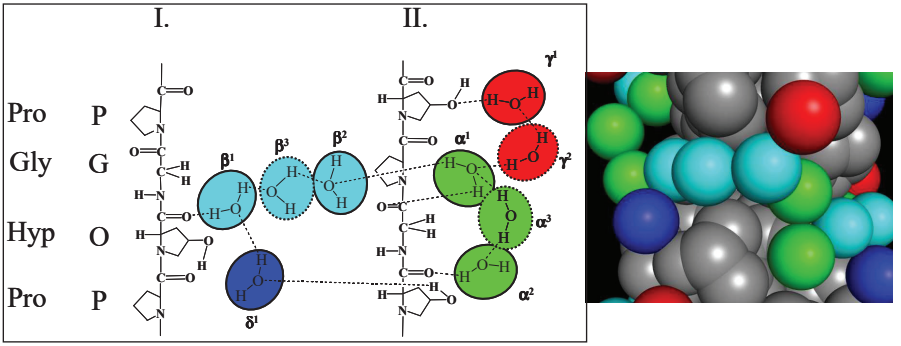

Figure 1. left, Schematic water connectivity (first layer hydration network) of a triple helical collagen. Green: α, light blue: β, red: γ and marine: δ. The names of the water bridges are taken from Bella et al., while labeling of the individual water molecules were completed accordingly. Those H_2Os circled with a solid line can always be found in the X-ray structure, while those encircled by a dashed line can be missing. **right,** The water bridges as seen in the 1V7H PDB structure. The same coloring pattern is used as for the scheme above, while all atoms of the protein are grey. (No hydrogen atoms are shown.)

Also, another type of water molecule can be observed when the Xxx amino acid is not a proline (nor hydroxyproline). This water molecule connects two protein chains in the triple helix, and is called as ζ-type water.

Methods

Amyloid models

To determine the most stable packed structures we performed calculations for two types of peptides: Ac-$(Ala)_3$-NHMe (standing for amino acids that have side chains and are chiral) and Ac-$(Gly)_3$-NHMe (representing an amino acid without a side chain). In these calculations we examined all possible conformations where all the amide groups have two peptide bonds with neighboring peptide molecules.

The crystal structures were optimized at HF/3-21G level, and subsequently single-point calculations were carried out at B3LYP/6-31G(d) level, both using periodic boundary conditions.

Collagen models

To examine the backbone preferences and the triple-helix stabilizing effects the following model systems were constructed and subjected to full *ab initio* geometry optimizations. Six residue long N- and C-protected collagen triple helix models composed of a total of 18 amino acid residues were generated: *i)* glycine only (GGG collagen helix model), *ii)* from L-alanine only (AAA collagen helix model), *iii)* from L-alanine and glycine (AAG collagen helix model), *iv)* from L-alanine and D-alanine, (AAa collagen helix model) *v)* from L-prolines and glycine (PPG collagen helix model), *vi)* from L-prolines and D-alanine (PPa collagen helix model), *vii)* from sarcosine and glycine (SaSaG collagen helix model)and *viii)* form L-proline, L-hydroxyproline and glycine (POG collagen helix model). In all of these cases where more than a single type of amino acid residue is involved, glycines are placed appropriately and the polypeptide chains are suitably shifted: *e.g.* the -Gly-Xxx-Yyy-Gly-Xxx-Yyy-, -Yyy-Gly-Xxx-Yyy-Gly-Xxx- and –Xxx-Yyy-Gly-Xxx-Yyy-Gly- where chains are adjusted head to head. All these models are shown in **Figure 2**.

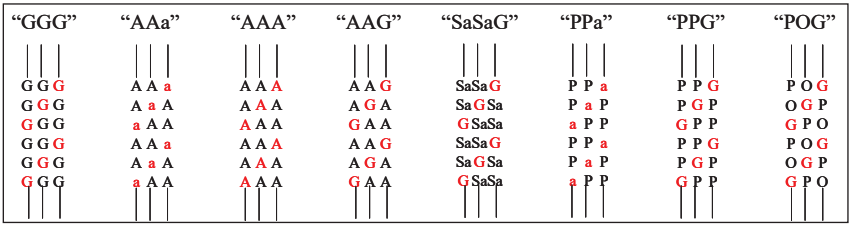

Figure 2. 3x6 residue containing models of different amino acid constitution that are used to examine the different backbone stabilizing effects on the triple helix.

All structures were fully optimized first at the RHF/3-21G and subsequently at the B3LYP/6-31G(d) levels of theory. Finally, energy calculations were completed both at the B3LYP/6-311++G(d,p) and B3LYP/PCM/6-31G(d) levels of theory on *a priory* optimized DFT structures. For the PCM calculations we used the IEF-PCM method and water as a solvent, meaning that ε=78.39 was set.

Also, on all of the energy minimized structures frequency calculations within the harmonic approximation were carried out at the B3LYP/6-31G(d) level of theory. Gibbs free energy and entropy data are direct results of these calculations.

To determine stabilities of water molecules connected to the tropocollagen via H-bonding(s), the following three types of model systems were optimized. The first one was designed to examine the binding of the internal structural water, the so called ζ-type water, for which neither proline nor hydroxyproline can be at position Xxx. Therefore, in the 3x6 amino acid containing model both Xxx and Yyy positions are occupied by alanines. All six essential water molecules were placed in and the overall system was fully optimized at B3LYP/6-31G(d) level of theory.

The second type of supramolecular system was chosen to characterize the horizontal water thread. Thus, here the α– and β–bridges were in focus. Some additional water molecules from neighboring bridges had to be added to the model to ensure the well-defined position of the two bridges of interest and to avoid their shifting.

The third type of supramolecular system was built to characterize the vertical water thread. Thus, here the γ- and δ-bridges were in focus.

All calculations were carried out with the Gaussian03 program

Results

Crystal structures

Peptides that have side chains are represented by an alanine containing tripeptide Ac-$(Ala)_3$-NHMe. For these the three most stable structures are some forms of the β-pleated sheet. The most stable form is the antiparallel, single layer β-pleated sheet, then the parallel single layer β-pleated sheet, and than the parallel multiple layered (3D) β-pleated sheet, that is still only 1.7 kcal·mol^{-1} less stable than the antiparallel form.

The least stable structure is where the tilting angle is −60º. It is interesting, the energy gap between "2-neighbored" and 4- or 6- neighbored structures is quite large, more than 10 kcal·mol^{-1}. That means that the calculations eventually follow the deductions we made by common sense. That is, the presence of a side chain forces such a large distance among the peptide chains that the H-bonding becomes too weak.

Peptides that do not have side chains are built up of glycine (as this is the only amino acid without a side chain). The most stable structure is where one peptide is surrounded by six others, and the tilting angle of the amide groups is 120°. In this superstructure each molecule has a polyproline II-like backbone conformation which was also described by Crick and Rich. The least stable structure is the other hexagonal, where the amide planes are tilted by 60°.
Summarizing, we can state that the most stable conformation for peptides which have side chains is one type of β-sheet, while for polyglycine the most stable aggregation is hexagonal and the amino acids have ε_L conformations, the same as in collagen.

Collagen triple helices

The collagen triple helix stability compared to three individual strands as function of the primary sequence is: AAA<PPa<AAa<AAG<GGG<POG<PPG<SaSaG, where the formation is favored for the GGG, AAG, POG, PPG and SaSaG triplets. The latter tendency is in agreement with general expectations based on experimental melting point data of various triplets embedded in a $(POG)_3$-XYG-$(POG)_4$ sequence. Stability of the experimental models grow in the AAA → AAG → PPG → POG direction.

For the non-imino acid containing models, AAA, GGG, AAa and AAG, the triple-stranded parallel β-pleated sheet structure is stable over the collagen-like triple helical structure both with respect to energy and with respect to Gibbs free energy results. For these four models, by using the PCM solvent model, the relative energies are as follows: $\Delta E^{\text{B3LYP/PCM/6-31G(d)//B3LYP/6-31G(d)}}$ = +6.4, +3.8, +4.3 and +4.7 kcal.mol^{-1} per triplet, respectively. The most stable β-sheet is formed by the "alanine only" model, AAA, at all levels of theory. Conformation selection is reversed for the other four (SaSaG, PPa, PPG and POG) models. In fact, the collagen triple helix of the PPG model becomes more stable by 4.8 kcal·mol^{-1} in vacuum and by 3.8 kcal·mol^{-1} in water. The POG model in its triple helical form is more stable than the POG-sheet model by 3.4 kcal·mol^{-1} in vacuum and 2.0 kcal·mol^{-1} in water. The SaSaG triplet has around the same amount of Gibbs free energy difference between the two secondary structures as the POG, although according to the energy data it is less stable. For the PPa triplet the triple helix formation is only slightly preferred regarding both the energy and the Gibbs free energy data.

The first hydration layer of collagen

The stability order of the different water binding places is as follows: (α^3), (γ^2), α^2, α^1, δ^1, γ^1, β^1,(β^3),(β^2), ζ (or internal). α^3 is the strongest while β^2 is the weakest water binding site.

We have observed that these water bridges self-assemble into left-handed water threads around tropocollagen. Each of the polypeptide chains of collagen forms a left-handed screw, but the triple helix formed from these turns out to be a right-handed supramolecular complex. As Orgel *et al.* have described a filament built up from seven tropocollagen units also evolves as a right-handed screw. Therefore, the left-handed water threads nicely fit into the "gap" between the right-handed triple helical tropocollagen and the collagen filament. This counter-twist is what provides the stability and attachment between strands for twisted ropes, it is interesting to observe the same thing for a "molecular rope".

The function of the extra water molecules in the α, β and γ-bridges can be various. First, their places might serve as "water-hole-conducting" places, meaning that these places can be used by water molecules, when they are flowing between tropocollagen triple helices. That is why Henkelman *et al.* have found that the flow of water molecules in collagen is fluid-like. Second, these "extra" water molecules can serve as molecular "buffers": at high water

concentration molecules occupy these places, whereas in case of low water concentration these places might be left vacant. In this case, of course, the remaining water molecules arrange themselves to have the best H-bonding contacts. Therefore, collagen can act as a sponge: take up and store water, and release it if necessary at a relatively low energy cost, without deteriorating the global fold of the macromolecule.

Summary

Protein oligomers and aggregates can have multiple roles in a living body. Amyloid and similar type of aggregates are formed without strict self-regulation and are associated to several illnesses. On the other hand, associated multichain nanosystems formed by different polypeptide chains such as collagen (β-barrels) are vital in cell and tissue formation . Furthermore, the latter type of nano-associates, formed in a strictly controlled manner such as collagen fibers can be produced or dismantled according to the needs of the living organism.

Plaque formation from different peptides or proteins is the cause of many illnesses. Also, several other proteins were found to form plaques under the proper cellular conditions. These plaques have the same macroscopical forms, and are therefore thought to possess the same microscopical structure: the β-pleated sheet. This leads to the question of why are these β-pleated sheet structures are so much preferred. To answer this question we have proposed a deduction scheme, and for the confirmation of it we have designed and accomplished suitable theoretical calculations on periodical model systems. With this we were among the first researchers who carried out periodical calculations on peptides.

The results show that for a glycine containing peptide the most stable form is a two-dimensional superstructure that was already described by Crick and Rich, where the residues have the ε_L local backbone conformation. The collagen forming residues have the same conformation as these glycine residues. The alanine containing peptide models, however, prefer the form of β-pleated sheet, as these type of structures have ~10 kcal·mol^{-1} less energy than any of the other ones. As the side chains do not let the peptide chains get close enough to each other, the hydrogen bonds get weaker, resulting in the presented energetics of the structures. Therefore the above QM calculations have shown that the deduction holds and that for polypeptides having side chains the only possible closely packed structure is some kind of

sheet. That is why all proteins or peptides that are allowed to adopt a closely packed structure form amyloid-like fibrils so easily.

The conformation that is the most stable for only glycine containing peptides is preserved in a triple-helical collagen structure. We have studied the effect of selected amino acid residues on the stability of the tropocollagen. For this reason we have carried out quantum chemical calculations and compared the stability of the collagen triple helices with those of the β-pleated sheets, the building unit of amyloid. In this way we could analyze the different contributions coming from differently structured amino acids. For the experimentally already studied amino acids these comparisons provided the same results about which amino acids or triplets strengthen collagen formation. Sarcosine, an amino acid that is abundant in our body but is not among the encoded 20 residues, helps to form a triple helix apparently stronger than even those built from PPG or POG triplets. Therefore sarcosine containing artificially synthesized collagens can have future in medical or cosmetical applications.

From our results a simple hypothetical deduction emerges that describes that collagen stabilizing amino acids are needed to reduce the number of H-bonding capacities that are oriented away from the triple helix. This way the importance of bound water molecules is put into different light, as they form hydrogen bonds with the remaining carbonyl groups that are still oriented away from the triple helix.

We have determined the binding strength of the water molecules forming the first hydration layer, to the best of our knowledge for the first time in the literature. According to our calculations a synergistic effect can be observed between the bound waters that are interconnected with strong hydrogen bonds. Furthermore, these water molecules around collagen form threads that wrap it as a net. Tropocollagens are protected by these water nets and are also first contacted to each other through them. Therefore, some biological or medical applications (for example for decreasing the effects of osteoporosis imperfecta) could be based in the future on the maintenance of a strong hydration network around the triple helices.

Related publications

1. Palfi, V.K. and Perczel, A. *How stable is a collagen triple helix? An ab initio study on various collagen and beta-sheet forming sequences.* Journal of Computational Chemistry, 2008. **29**(9): p. 1374-86.

2. Perczel, A., Hudaky P. and Palfi V.K., *Dead-end street of protein folding: Thermodynamic rationale of amyloid fibril formation.* Journal of the American Chemical Society, 2007. **129**(48): p. 14959-14965.