STRUCTURAL CHANGES IN PHOSPHOGLYCERATE KINASE DURING FOLDING AND AMYLOID FORMATION

PhD thesis

Gergely Agócs

Doctoral School of Pharmaceutical Sciences

Semmelweis University

Supervisor: Dr. Szabolcs Osváth Ph.D.

Official reviewers:
Dr. Tamás Kardon Ph.D.
Dr. László Grama Ph.D.

Head of the Theoretical Examination Committee:
Dr. Erzsébet Ligeti D.Sc

Members of the Theoretical Examination Committee:
Dr. Miklós Geiszt Ph.D
Dr. László Nagy Ph.D

Budapest, 2012

1. Introduction

Proteins are heteropolymers composed of amino acids of diverse physico-chemical properties that play various roles in living organisms. The mere existence of the chemical structure (*i.e.* amino acid sequence) is not sufficient but the presence of the proper conformation (*i.e.* three dimensional structure) enabling functioning is also necessary. Biologically active structures represent only a *minute fraction* of theoretically possible protein conformations, yet, amazingly, these form in a very short time and essentially without external help *in vivo.* The process is called protein *folding* and the resulting native structure is primarily stabilized by side chain interactions. Nevertheless, there are rare occasions when biologically inactive or pathologic structures form (*misfolding*). Often, these structures are prone to aggregation and form threads of well-defined structure rich in beta-sheets called *amyloid fibrils*; these fibrils are stabilized by interactions between the peptide residues of the protein backbone and are hugely resistant to external impacts. Such structures were reported to be present in several diseases collectively known as *amyloidoses,* including aging-related degenerative illnesses widespread in developed countries. These facts raised amyloid studies to the forefront of biomedical research.

Amyloids were first identified in tissue samples obtained from patients suffering from certain diseases. Later it was proved that in case of a given illness only one (or rarely a few) protein turns into amyloids; these could also be turned into amyloids *in vitro*. Recently several proteins not related to amyloid diseases (such as yeast phosphoglycerate kinase) were shown to form amyloids *in vitro*. It became a more and more widely accepted view that amyloid formation is a general property of proteins. Amyloid studies on non-pathogenic proteins are important to elucidate the difference between pathogenic and well-behaved proteins.

The first step of misfolding is the destabilization of the native fold. Since the destabilized structure still shares some analogy with the original fold the

native structure might influence the kinetics of formation and final structure of amyloids that in turn affects resistance.

Despite the exceptional stability of amyloids there are certain methods to disassemble these aggregates. Since small oligomers are the most dangerous with respect to infectiveness, it is important to characterize the efficiency of disaggregation and the successfulness of the reestablishment of native structure.

Both folding and amyloid formation initiated from the denatured protein follow the same path to the point of the so-called *molten globule* intermediate product. The model used to successfully describe the denatured state–intermediate transformation assumes finite number of hierarchic transition levels. In case of folding further structural changes can be described by kinetics assuming a single barrier crossing. On the contrary, further development toward aggregation and amyloid formation can be described by the Smoluchowski coagulation model, which uses size-independent rate constants for the sub-steps off aggregation. Fitting of these models is possible to the spectroscopically acquired fluorescence intensity change data. Uncovering kinetics can bring us closer to understanding the factors that decide the path of protein conformational changes between folding and amyloid formation.

2. Objectives

There is a long tradition of protein structure research – be it experimental work or computational simulation – in the Department of Biophysics and Radiation Biology. The experience of the research group and especially that of my supervisor as well as the measuring instruments available in the lab and through partnership with other research groups allowed me to answer the questions related to topic described in the introduction. Although my research work is to be considered as basic research, the investigation of amyloid assembly and disassembly and the parameters defining the direction and kinetics of these processes brings us closer to the identification of pharmacologically relevant target points, which increases the chance of elaborating causal therapies.

My research objectives can be summarized as follows:

I. Effect of inter-domain interactions within PGK on its amyloid formation.
   - I.1. Amyloid formation from the elements of the protein system: tryptophan mutants the domains and the whole protein.
   - I.2. Finding a suitable mathematical model for the description of amyloid formation based on fluorescence intensity change data.
   - I.3. Comparing the amyloid formation of the isolated domains and the whole protein.

II. Recovery of native protein from amyloid fibrils formed from PGK.
   - II.1. Amyloid formation from PGK and the characterization the efficiency thereof.
   - II.2. Finding ideal conditions for the disassembly of PGK-amyloids and for the recovery of the native protein.
   - II.3. Characterization of the biological equivalency of the original and the recovered protein with physical and biochemical methods.
   - II.4. Determination of the recovery efficiency of active protein.

III. Studying the kinetics of processes leading to the folding or amyloid formation of the acid denatured PGK.

III.1 Finding solution conditions were the changing of a single parameter is enough to influence if the denatured protein forms amyloid of native structure.

III.2. The determination of the resulting structure at different values of this parameter.

III.3. Finding a mathematical method to describe the formation of native structure as well as amyloid fibrils based on some spectroscopically measured signal.

III.4. Analyzing the kinetics of folding and amyloid formation with the aid of the parameters determined with the above model with special interest in possible common steps, point of separation and factors defined the direction of the process.

3. Methods

The phosphoglycerate kinase protein used for the measurements was expressed in E. coli cells type BL21(DE3)pLysS (Novagen). The gene of the protein was inserted in a pET28a plasmid. The protein was purified on an ÄKTApurifier (GE Healthcare) fast protein liquid chromatography system containing Ni-NTA Superflow (Qiagen) nickel affinity resin. The chromatographic fractions containing protein were run on a polyacrylamide gel (Invitrogen), identified PGK-containing fractions were unified and concentrated in a pressurized ultrafiltration chamber using regenerated cellulose membrane (Millipore) and, finally, were dialyzed into the solution corresponding to the measurement conditions.

For domain interaction experiments, amyloid formation was initiated by addition of 200 mM NaCl to the acid denatured protein. Native fluorescence spectra were registered on a JobinYvon luminometer (excitation at 295 nm). The ratio of the integrals between 310–320 nm and 360–370 nm of the spectra was plotted as a function of time and the Smoluchowski equation was fitted to the curves. Fitting was executed in all cases with Wolfram Research's Mathematica software.

For activity recovery measurements, protein was first turned into amyloids with the method described above, and a set of investigations were performed to identify amyloid fibrils: circular dichroism spectrum was registered with a CD-spectrometer (JASCO 700) between 300–650 nm (path length: 10 cm, increment: 1 nm, detection slit width: 2 nm, detection speed: 50 nm/min); Thioflavin T fluorescence spectrum was recorded at 20 μM dye concentration (excitation: 450 nm, excitation slit width: 4 nm, detection: 458–530 nm, detection slit width: 2 nm, increment: 1 nm, detection speed 2 nm/sec); the size distribution of aggregates was determined with dynamic light scattering on a home built instrument; electron micrographs were also prepared (Hitachi 7100). After decreasing salinity and subsequently elevating pH into the native range,

biological equivalence as well as the efficiency of recovery was determined with calorimetric (MicroCal VP-DSC) measurement of transition temperature and enzyme activity measurements using the *3-phosphoglycerate* + *ATP* $\xrightarrow{PGK}$ *1,3-bisphosphoglycerate* + *ADP* and the *1,3-bisphosphoglycerate* + *NADH* $\xrightarrow{GAPDH}$ *glyceraldehyde phosphate* + *NAD* concatenated reactions – catalytic rate was calculated from the change of NADH concentration while rates at different substrate levels were used to determine maximum catalytic rate and Michaelis constant.

For comparison of the native and amyloid conformational pathways, solution containing the acid denatured protein was mixed into one of a set of citrate-phosphate buffers (10–10 mM, also containing 0.2 M NaCl and ranging from 2 to 7 in pH) with a 1-to-10 volume-to-volume ratio at + 4 °C. Events on the 1 ms – 50 s time scale were initiated by stopped-flow mixing and followed by the detection of native tryptophan fluorescence (Applied Photophysics PiStar-180, excitation: 295 nm, excitation slit width: 5 nm, detection: above 320 nm). On the 30–1800 s timescale tryptophan fluorescence measurements (Horiba Jobin-Yvon Fluorolog-3 fluorimeter, excitation at 295 nm with 0,25 nm bandwidth, detection at 355 nm with 12 nm bandwidth) were done after initiation with hand-mixing. For the measurements spanning to four days I used the same instrument to register series of spectra (excitation at 295 nm with 1 nm bandwidth, detection from 305 to 400 nm with 1 nm increment, 1 nm/s rate and 1.5 nm bandwidth) and the total intensity detected in the spectrum range was used for evaluation. Beside the native tryptophan fluorescence measurements the change in the fluorescence spectrum of 20 μM Thioflavin T solution was also measured (excitation at 450 nm with 4 nm bandwidth, detection between 458–530 nm with 2 nm bandwidth, 1 nm increment), here the ratio of intensities integrated throughout the spectrum of the dye solution (volume: 2625 μl) before and after addition of 100 μl sample was used for evaluation.

4. Results and conclusions

Using primarily spectroscopic measurements, I studied the conformational changes of the model protein phosphoglycerate kinase and the kinetics of these changes during a broad time interval and among a variety of circumstances throughout my work. Based on the evaluation and analysis of the measurements, the following conclusions can be drawn:

I.1. Each member of the protein system created to study the interactions between folding units (*i.e.* both the single tryptophan whole protein mutants and the corresponding domain mutants) was able to aggregate and form amyloids under circumstances similar to that used for the amyloid formation of the wild type PGK (200 mM NaCl in pH = 2 HCl solution). Time range for amyloid formation was on the same order of magnitude (roughly five days) as that in case of the wild type protein. Amyloid formation was proved by electron micrographs in all cases.

I.2. The modified version of the Smoluchowski coagulation equation, which includes the coagulation time and the critical monomer number as variables could be used – besides the wild type protein – in case of three out of the four mutants, *i.e.* in all cases where amyloid formation was accompanied by a significant shift of the fluorescence spectrum. Problems originating from the intensity fluctuations unavoidable in case of long term fluorescence measurements were overcome by a fitting formula that is sensitive to spectrum shape and thus spectrum shift. The model supposing a single-step fluorescence change above a critical monomer number fit very well our measured data.

I.3. There are well defined differences between the fluorescence changes of isolated domains and the corresponding (*i.e.* bearing a tryptophan in the same position) whole proteins which fact lets us suppose different kinetics. While the isolated N-domain indicates practically no change, the fluorescence spectra of the other three mutants shows similar shift towards shorter wavelengths as seen in case of the wild type protein. While the coagulation time

constants can be considered equivalent, the critical monomer number for the isolated C-domain differs significantly from that of the mutant or wild type whole proteins. In general, we concluded that the domains of the whole proteins transform into amyloids with much more concerted kinetics compared to the isolated domains, which can be explained by the absence of interactions between distant residues. The lack of these interactions affect not only kinetics but also the final structure.

II.1. THe efficiency of the classic method for transforming PGK into amyloid fibrils (pH = 2 HCl-solution, 200 mM NaCl) was analyzed by several techniques based on different principles. Electron micrographs, increased fluorescence in Thioflavin T solution, and the characteristic circular dichroism spectrum exhibited in presence of Congo Red dye all confirmed the presence of amyloid structures. The particle size distribution function obtained by dynamic light scattering gives evidence that 99% of the protein is present in form of one type of aggregate. The positivity of amyloid detecting methods and the size distribution function proves that the protein was nearly completely transformed into amyloid.

II.2. Disaggregation of amyloids while preserving the integrity of the primer structure of the constituent protein molecules seemed to be possible by a two-step process: First dialyzing against a salt-free pH = 2 HCl-solution to yield denatured protein monomers, then transferring these among native conditions which resulted in refolding into the native three-dimensional structure. Other methods, notably direct dilution into native conditions, did not prove to be efficient.

II.3. The biological equivalence of the starting and regained protein could be proved by both enzyme kinetic and thermal analytic methods. Kinetic measurements showed that the regained preparation catalyzes the 3-phosphoglycerate + MgATP $\longleftrightarrow$ 1,3-bisphosphoglycerate + MgADP reaction, which is the primary condition for biological equivalence. The transition temperature (melting temperature) of the starting and the regained

protein was identical within error, which supports structural (including conformational) identity of the two samples.

II.4. Growth and subsequent disaggregation of PGK-amyloids occurred among extreme (though not unreal) conditions compared to physiological circumstances. However, the proportion of recovered active enzyme determined by enzyme kinetic measurements falls in the lower part of the range of proportions obtained during the much less destructive refolding experiments. Therefore, not only the recovery efficiency of proteins enclosed in amyloids was determined but it was also concluded that this value is comparable to the PGK-refolding efficiency published by others.

III.1 The acid denatured (pH = 2 HCl-solution) PGK turns into amyloids within days if placed into a pH = 2 HCl-solution also containing 200 mM NaCl. If it is, alternatively, transferred into a solution of the same salinity but with pH = 7.0, the native structure will be recovered. Therefore it is possible to define the direction of conformational changes of the acid denatured protein with the modification of a single parameter. Deviation from nominal pH due to mixing in the studied pH-range was lessened by addition of salts of triprotic acids that provided good buffering conditions.

III.2. Based on spectroscopic evidences we can conclude that within the 7–5 pH-range folding leading to the native conformation is the preferred structural change while in the 4–2 pH-range misfolding and aggregation resulting in amyloid fibrils is the favored direction of structural modification. Time dependence of fluorescence changes reveals that the formation of the native structure is fastest at pH = 7 while the formation of amyloids is the fastest at pH = 2.

III.3. After setting certain boundary conditions, the fluorescence signal was applied to follow protein structural changes numerically. For the acid unfolded–intermediate transformation the finite level hierarchic energy landscape model could be used at each pH-value: I used the same non-exponential functions to fit data as earlier in case of folding studies. The intermediate–native transformation was assumed to be a monomolecular reaction and fitted with a simple exponential function. The intermediate–amyloid transformation was

described with the Smoluchowski coagulation equation assuming size-independent reaction coefficients.

III.4. The hyperfluorescent intermediate product was formed at each pH in the studied range, following the hierarchic energy landscape model in each case, although with with different speed. This process reached its maximum speed at pH = 3.5. Further transformation of the hyperfluorescent intermediate at pH-values closer to 7 led to the native fold in a one-step process. At lower pH-values, however, the size independent coagulation kinetics prevailed that finally resulted in amyloid formation. Folding and amyloid formation, hence, follows similar pathway until the formation of the hyperfluorescent intermediate, although the number and depth of potential wells on the energy landscape depends on pH. Thereafter, the protein follows amyloid formation at lower pH-values and folding into native state at higher pH-values from the two possible pathways. The choice of direction is strongly exclusive: the coexistence of folding and amyloid formation products could not be observed at any of the studied pH-values, *i.e.* the direction of transformations is under thermodynamic control.

5. Summary

Alzheimer's disease, Parkinson's disease and transmissible spongiform encephalopathies are well known incurable illnesses and are member of the so called amyloidosis disease group. Besides the lack of effective causal therapy, they share the property of featuring the deposition of a non-physiologic protein structure: amyloid. Amyloids are protein aggregates known for their exceptional resistance. Amyloid structure is dominated by β-sheets that are stabilized by hydrogen bonds between the atoms of the polypeptide backbone.

Their increasing clinical significance moved the study of amyloid aggregates to the forefront of life science research. An important part of these investigations are the biophysical experiments searching for the cause of the

tendency of proteins to form amyloid, the kinetics of amyloid formation, the stability of amyloids, and the eventual possibility to disaggregate amyloids.

Throughout my investigations, I applied spectroscopic and biochemical methods to gain a deeper insight into the behavior of yeast phosphoglycerate kinase, a popular model protein. I have concluded that the structural domains that originally play role in the formation of the sequence dependent native structure also influence the development of the otherwise sequence-independent secondary bonds during amyloid aggregation. The effects are also demonstrated by the parameters of the function describing the kinetics.

I conducted successful experiments to disaggregate amyloids formed of phosphoglycerate kinase, and furthermore, to restore the enzymatically active protein, as well as to characterize the efficiency of the restoration: I was able to completely disaggregate the amyloids into monomers and then to bring back these monomers in the enzymatically active form.

Finally, I investigated the path of structural changes of the acid-denatured form of phosphoglycerate kinase toward the native and the amyloid state. I have successfully constructed a system where I could determine the direction of structural changes by the modulation of a single parameter: the acidity. Using this system, I concluded that the route of protein structure changes is governed by the energy content of the product (be it amyloid or native protein), that is, the process is under thermodynamic control.

6. Publications published in topic of the dissertation

Agócs G, Szabó BT, Köhler G, Osváth S. (2012) Comparing the folding and misfolding energy landscapes of phosphoglycerate kinase. Biophys J 102 (12):2828-34.

Agócs G, Solymosi K, Varga A, Módos K, Kellermayer M, Závodszky P, Fidy J, Osváth S. (2010) Recovery of functional enzyme from amyloid fibrils. FEBS Lett. 584 (6):1139-42.

Osváth S, Jäckel M, Agócs G, Závodszky P, Köhler G, Fidy J. (2006) Domain interactions direct misfolding and amyloid formation of yeast phosphoglycerate kinase. Proteins 62 (4):909-17.