Insights into the molecular mechanism of protein native-like aggregation upon methylglyoxal glycation

Luis M. A. Oliveira[1,4], Ricardo A. Gomes[1,2], Dennis Yang[3], Carlos Família[1], Ana Lages[1], Ana V. Coelho[2], Regina M. Murphy[3] and Alexandre Quintas[1*]

Affiliations

1 Centro de Investigação Interdisciplinar Egas Moniz, Instituto Superior das Ciências da Saúde Egas Moniz, Campus Universitário, Qta. da Granja, Monte da Caparica 2829-511 Caparica, Portugal.

2 Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, 2780-901 Oeiras, Portugal

3 Chemical and Biological Engineering, University of Wisconsin, Madison, WI USA

4 Centro de Química e Bioquímica, Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade de Lisboa 1749-016 Lisboa, Portugal.

[*]**Corresponding Author:** E-mail address: alexandre.quintas@gmail.com Institutional address: Centro de Investigação Interdisciplinar Egas Moniz, Instituto Superior das Ciências da Saúde Egas Moniz, Campus Universitário, Qta. da Granja, Monte da Caparica 2829-511 Caparica, Portugal.

Running Tittle: Methylglyoxal induces protein native-like aggregation

Number of manuscript pages: 37
Number of tables: 5
Number of figures: 6

Abstract

Protein glycation induces structural and stability changes that impair protein function, and is associated with several human neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease and Familial Amyloidotic Polyneuropathy. Recently we have shown that methylglyoxal induces and stabilizes the formation of small native-like aggregates in the amyloidogenic protein insulin and the same was previously shown for α-synuclein. However, the fundamental biophysical mechanism underlying such methylglyoxal-induced protein aggregation is not yet fully understood. In this study, we used the model protein cytochrome c to characterize the specific glycation targets and to investigate the glycation effects on protein structure, stability and aggregation. Methylglyoxal was found modify cytochrome c in a single residue and to induce the formation of cytochrome c native-like aggregates. Additionally, it is shown that methylglyoxal glycation of cytochrome c also results in the formation of a partially unfolded species. Interestingly, the formation of this partially unfolded species is not implicated in the aggregation process, a clear difference from amyloid fibril mechanisms that involve partially or totally unfolded intermediates. Equilibrium-unfolding experiments using guanidinium hydrochloride shows that glycation strongly reduces cytochrome c conformational stability. This reduction is balanced by aggregation that increases conformational stability. The data collected from analytical and spectroscopic techniques along with kinetic analysis based on least-squares parameter fitting and statistical model discrimination permitted the proposal of a comprehensive thermodynamic and kinetic model for native-like aggregation of methylglyoxal glycated cytochrome c.

Keywords: Native-like aggregation, Conformational diseases, Glycation, Methylglyoxal, Cytochrome c

Importance of the work: The manuscript presents important data on the molecular mechanisms of protein aggregation induced by glycation, which has been associated to several human neurodegenerative diseases. Our major findings show that glycation induces protein native-like aggregation and that the aggregation process is thermodynamic and kinetically favored. These results may be extended to the aggregation behaviour of amyloidogenic proteins upon glycation, suggesting a shift from an amyloidogenic pathway to a native-like aggregation process.

Abbreviations Used: α-CHCA: α-cyano-4-hydroxicinamic acid, ACN: Acetonitrile, AGE: Advanced glycation end-products; AIC: Akaike Information Criteria, CD: Circular dichroism, CID: Collision induced dissociation, GdnHCl: Guanidinium hydrochloride, MAGE: Methylglyoxal-derived advanced glycation end-products, MGH: Hydroimidazolone, MALDI: Matrix-assisted laser-desorption ionization, SEC: Size exclusion chromatography, TOF: Time of flight

Introduction

More than 40 human diseases are associated with the deposition of normally soluble and functional proteins into insoluble aggregates having well-defined morphological, structural and tinctorial properties (1). These include a long and unbranched shape when fibrils are analyzed with transmission electron microscopy (2) or atomic force microscopy (3), the ability to bind specific dyes (4; 5) and a typical cross-β pattern, when analyzed by X-ray diffraction (2). Many of these diseases involve protein aggregation due to genetic mutations, abnormal proteolysis or *post*-translational modifications in disease-related proteins. Among the different known *post*-translational modifications, protein glycation has gained increased attention in the context of several human pathologies, such as age-related disorders and neurodegenerative diseases of amyloid type. In this irreversible *post*-translational modification, protein amino groups (N-terminal and arginine and lysine side chains) are modified by carbonyl compounds

leading to the formation of advanced glycation end-products (AGE). Methylglyoxal, a non-enzymatic by-product of dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate production during glycolysis in eukaryotic cells, is the main intracellular glycation agent, reacting with amino groups in lipids, nucleic acids and proteins with the formation of methylglyoxal-derived advanced glycation end-products (MAGE) (6). Protein glycation has been reported in numerous neurodegenerative diseases to be involved in amyloidosis (7-11) and several studies suggest that glycation may be an early event in protein aggregation (10; 12; 11). Moreover, recent studies showed that glycation by methylglyoxal resulted in inhibitory effects on the formation and extension of fibrils and in the stabilization of native-like oligomeric species (13; 14). The formation of these aggregates *in vivo* may be a key step for cellular damage, cytotoxicity and neurotoxicity, as increasing evidence shows that the severity of cognitive impairment in protein misfolding diseases correlates with the levels of small oligomeric species and not with the large amyloid fibrils (15-18).

The propensity of normally folded proteins to form amyloid-like fibrils increases in conditions that favor the partial unfolding of the native state across the major unfolding energy barrier. These include low pH, high temperature, or the presence of organic solvents (19-21). However increasing evidence is now accumulating that folded proteins also retain a significant tendency to aggregate with no need of unfolding as first obligatory step (22-25). Despite the identification of unfolding-independent aggregation pathways, the molecular mechanisms and the structural and stability determinants that promote aggregation from native-like states are not yet fully understood. As a consequence, an in-depth understanding of the biochemistry and biophysics of the native-like aggregation process upon glycation will be crucial to disclose fundamental mechanisms underlying these diseases as well as to further develop therapeutic strategies.

In the present work we have focused our attention on cytochrome c, a protein recently used as a model to study protein aggregation (26). Model proteins offer several advantages as they are abundant and inexpensive, and their biophysics have been well characterized. The mechanistic principles gathered from model proteins can lead to hypotheses about how glycation triggers protein aggregation that can later be tested with disease-related proteins, which most of the time are difficult to work with. Among several model proteins available for biophysical studies, cytochrome c is excellent because it is highly soluble, and its ready production facilitates experiments requiring the large quantities of protein necessary to detect and characterize early aggregation intermediates. These features are paramount advantages for performing a detailed characterization of methylglyoxal effect on the protein aggregation process.

In this study, we show that cytochrome c is converted into native-like aggregates upon methylglyoxal glycation, and that partial unfolding represents an off-pathway to aggregation. Circular dichroism (CD) experiments and kinetic model discrimination demonstrate that formation of glycated cytochrome c aggregates is thermodynamically and kinetically favored.

Results

Characterization of cytochrome c glycation sites by mass spectrometry

The detection of cytochrome c glycation and assignment of modified amino acid residues were performed by mass spectrometry after trypsin digestion. A comparative analysis of peptide mass spectra from the glycated and unmodified cytochrome c reveals major differences with several new peptides appearing exclusively in the glycated cytochrome c (Figure 1A and Table I). These new peaks can be caused by the occurrence of miscleavages

associated with defined mass increases of specific MAGE. Since only lysine and arginine residues are irreversibly modified, tryptic digestion of glycated proteins would then produce peptides with at least one miscleavage associated with a defined mass increase that is characteristic of a specific MAGE. For example, the new m/z peak at 1404.727 may correspond to the cytochrome c peptide 90-100 (TEREDLIAYLK) with m/z 1350.726 plus 54 Da, a mass increase characteristic of a hydroimidazolone modification. Likewise, the new peak at m/z 1532.821 may correspond to the cytochrome c peptide 89-100 (KTEREDLIAYLK) or the peptide 90-101 (TEREDLIAYLKK), with the exact same mass of 1478.821, plus a MGH modification (54 Da mass increment). To unequivocally confirm these data, MS/MS experiments, using collision-induced dissociation (CID), were performed to provide sequence information. When using the CID fragmentation technique, bond breakage mainly occurs through the peptide bond leading to b-ions (when the charge is retained by the amino-terminal fragment) or y-ion (when it is retained by the carboxy-terminal fragment). Thus, if an amino acid residue is modified, the particular y and complementary b ions, which encompasses the modification, will have the particular amino acid mass value plus 54.018 Da for hydroimidazolone. With the peptide with m/z of 1404.727 $TER_{MGH}EDLIAYLK$, we clearly observed the b3 ion that corresponds to a TER with 54 Da increase of MGH-modified arginine residue (Figure 1B). All other identified b ions show the same mass increment. The detection of y9 ion with a mass increase of 54 Da further confirms the glycation of R92 with the formation of a hydroimidazolone (Figure 1B). Likewise, the MS/MS data in figure 1C further confirms the glycation of R-92. In addition, the sequence information gathered for this peptide allowed the identification of peptide $KTER_{MGH}EDLIAYLK$ (89-100) instead of peptide $TER_{MGH}EDLIAYLKK$ (90-101), both composed by the same amino acid, hence having the exact same molecular mass.

From this analysis, only R-92 was found to be glycated with the formation of the MAGE hydroimidazolone. No lysine modification was detected in our experimental conditions. It is

worth mention that upon cytochrome c digestion and subsequent analysis of the tryptic digests, high sequence coverage (between 70 and 80 %) was obtained.

Structural characterization of the glycated cytochrome c species

To investigate the effects of methylglyoxal-induced glycation on cytochrome c aggregation, samples of native and glycated cytochrome c at two different incubation times were analyzed by size exclusion chromatography (figure 2). Cytochrome c appears at time 0 as a single molecular species, with an apparent molecular mass of 12.9 kDa (elution volume – 12.96 ml) (figures 2A and 2B). The native cytochrome c monomer is stable in solution during the time of analysis. In contrast, SEC elution profiles of glycated protein reveal an increasing complexity of *post*-glycation cytochrome c forms with the emergence of four new main species. These species show apparent molecular masses consistent with the dimeric, trimeric and tetrameric forms of cytochrome c (figure 2B). Moreover, another molecular species appears to be present at a higher elution volume (figure 2B, fraction U). The SDS-PAGE analysis reveals that the U fraction of cytochrome c has the same electrophoretic mobility of cytochrome c monomer (figure 2C). This result discards the idea of protein fragments and suggests that the U fraction is in fact a monomeric form of glycated cytochrome c.

To evaluate the secondary structure of the different species of *post*-glycation cytochrome c, far-UV circular dichroism (CD) was used. Figure 3A shows the CD spectra of the native cytochrome c together with the monomeric, dimeric and trimeric forms of glycated protein, at 25 ºC and physiological ionic strength and pH. A direct analysis of the CD spectra shows that the secondary structure content of both native and glycated forms of cytochrome c is very similar. The spectral minima at 208 and 222 nm and a positive band bellow 200 nm, which dominates all spectra, are characteristic of α-helical structures (27). Nevertheless, CD spectra of aggregates show minimal differences when compared to monomers. Aggregates

show no loss in secondary structure, but a slight readjustment of secondary structural elements distribution in relation to monomeric forms. CD spectra deconvolution of native and glycated species using the CDSSTR algorithm (28), allowed determination of the secondary structure contents presented in Table II. This analysis confirmed the mainly α-helical content of native and glycated cytochrome c. Still, the analysis of CD spectra of the dimeric and trimeric forms of glycated cytochrome c revealed an increase in β-sheet content with a corresponding loss of α–helix and β–turns. CD spectra deconvolution of native cytochrome c also shows the presence of a small fraction of β-sheet content. Although not detected either by X-ray diffraction or by NMR, Calvert and colleagues have observed a similar feature by FTIR (29). The CD spectra of the monomeric form of cytochrome c with a higher retention time (U fraction) shows no secondary regular structural elements (figure 3B), consistent with the appearance of an unfolded monomeric form of glycated cytochrome c. This new populated state of monomeric protein species shows that cytochrome c glycation causes cytochrome c unfolding and aggregation.

To evaluate alterations on the three-dimensional structure resultant of structural changes at secondary level, tryptophan intrinsic fluorescence was used. Tryptophan has an emission maxima ranging from 320 to 355 nm depending on the polarity of the local environment. Hence, protein fluorescence may be used as a diagnostic tool to probe the conformational state of a protein. Figure 3C shows an overlay of the fluorescence spectra obtained with native cytochrome c and the glycated monomeric, dimeric and unfolded species. Fluorescence spectra show that no shift in native cytochrome c fluorescence emission λ_{max} occurs, when compared with the monomeric and dimeric forms of glycated cytochrome c. These results indicate that the tryptophan environment is not affected by glycation or by aggregation, which suggests that the three-dimensional structure of cytochrome c is preserved in these processes. Nevertheless, a red-shift in fluorescence emission λ_{max} for the unfolded species is observed, showing that tryptophans are in a very polar micro-environment. This is

expected in an open conformation of the polypeptide chain, confirming the unfolded nature of this species.

Conformational stability of cytochrome c forms

To understand thermodynamically how glycation induces aggregation, we characterized the conformational stability of the glycated monomeric and dimeric forms of cytochrome c and compared to the native protein. Protein conformational stability can be accessed by analysing the dependence of the ellipticity values at 222 nm on guanidinium hydrochloride concentration. Native cytochrome c unfolds by a two-state model (M ↔ U) (30), without evidence regarding glycation-induced species. For glycated cytochrome c forms there are several possible unfolding pathways. We hypothesized that unfolding could occur in the presence or in the absence of intermediates and that for the dimer, the final unfolded form can be a dimer or a monomer. Thus, unfolding of the monomeric glycated cytochrome c can be described by a two-state M ↔ U or by a three-state model M ↔ I ↔ U, while the dimeric glycated cytochrome c can be described by one of the following five models: M ↔ U; D_2 ↔ I_2 ↔ 2U; D_2 ↔ 2I ↔ 2U and D_2 ↔ I_2 ↔ U_2. To discriminate among these unfolding pathways, SEC experiments in the presence of increasing amounts of GdnHCl were performed (figure 4A and 4B). In the 1.5-2.5 M GdnHCl concentration range, the glycated monomer and dimer are in the transition state between native and unfolded forms (figure 4C). Figure 4 A shows a progressive and direct conversion of the dimer to the unfolded form, suggesting that unfolding intermediates are not populated to any measurable extent. Retention times decrease progressively with the increase of denaturant concentration because the denaturant promotes a partial expansion of the hydrodynamic radius of the protein in both dimeric and monomeric forms. When compared with the cytochrome c monomer (figure 4 B), apparently both monomeric and dimeric species unfold to the same unfolded species, suggesting that the

unfolding pathway of the glycated dimer can be described by a $D_2 \leftrightarrow 2U$ model. The total absence of an unfolded dimeric form shows that aggregates are non-covalent and that cross-links are not observable in this denaturant concentration range. Glycated monomer also shows a direct conversion from the native to the unfolded form and no unfolding intermediates are observed, suggesting that glycated monomer unfolds through a two-state model (figure 4 B).

In order to establish the thermodynamic parameters for native and glycated cytochrome c (monomer and dimer), monomeric species were analysed according to a two-state unfolding model $M \leftrightarrow U$, while dimeric species was analysed considering equation 1 that describes an approximation to determine the effective ΔG (31). Both fits were made using the linear extrapolation method (32) in a non-linear least squares fitting procedure and yielded values for $\Delta G^{o}(H_2O)$, the conformational stability, and m, the dependence of ΔG^{o} on denaturant concentration. Table III shows the values obtained from the curves in figure 4C for $\Delta G^{o}(H_2O)$, m and C_m, the denaturant concentration at the midpoint of the unfolding transition. The magnitude of $\Delta G^{o}(H_2O)$ for the native cytochrome c fits well within typical values obtained for denaturation of similar size proteins (33) and the value of m explains the rapid transition between the two well defined states represented in native cytochrome *c* curve. When glycation occurs, a significant decrease on cytochrome *c* conformational stability from 8.1 ± 0.6 to 3.1 ± 0.6 kCal.mol^{-1} is observed. Additionally, C_m also suffers a decrease from 2.5 ± 0.4 to 2.2 ± 0.8 M, confirming the reduction of conformational stability upon glycation.

When compared to the glycated monomer, glycated dimer shows an increase of conformational stability from 3.1 ± 0.6 to 8.4 ± 1.3 kCal.mol^{-1} (table III). These results on conformational stability of native and glycated forms of cytochrome c may thermodynamically elucidate the aggregation process observed in size exclusion chromatography (figure 2). Methylglyoxal irreversibly decreases cytochrome c stability, possibly mainly due to loss of positive charges at the surface, forcing the modified monomer to aggregate, which leads to the

formation of a more stable species. The decrease in conformational stability of cytochrome c upon glycation is balanced with an increase of conformational stability gain upon aggregation.

Kinetic analysis of cytochrome c aggregation

Besides aggregation, cytochrome c undergoes a conversion to a partially unfolded form. Thus, despite the native-like structure of *post*-glycated cytochrome c aggregates, it is possible that the partially unfolded form participates in the aggregation pathway. To clarify the aggregation mechanism, a kinetic analysis of time course obtained by SEC was performed. We investigated two possible cytochrome c aggregation models. The first model assumes that folded monomers (M) reversibly convert to partially unfolded monomers (U), and that the aggregation proceeds through folded monomers; with dimers (D), trimers (T3) and tetramers (T4) formed by addition of M. Thus, the reaction pathway for model 1 is:

$$\mathrm{M} \rightleftharpoons \mathrm{U}$$

$$\mathrm{M} + \mathrm{M} \rightleftharpoons \mathrm{D}$$

$$\mathrm{M} + \mathrm{D} \rightleftharpoons \mathrm{T3}$$

$$\mathrm{M} + \mathrm{T3} \rightleftharpoons \mathrm{T4}$$

The differential equations describing the kinetic pathway for model 1 are:

$$\frac{d[U]}{dt} = k_1[M] - k_{-1}[U]$$

$$\frac{d[M]}{dt} = -k_1[M] + k_{-1}[U] - 2k_2[M]^2 + 2k_{-2}[D] - k_3[M][D] + k_{-3}[T3] - k_4[M][T3] + k_{-4}[T4]$$

$$\frac{d[D]}{dt} = k_2[M]^2 - k_{-2}[D] - k_3[M][D] + k_{-3}[T3]$$

$$\frac{d[T3]}{dt} = k_3[M][D] - k_{-3}[T3] - k_4[M][T3] + k_{-4}[T4]$$

$$\frac{d[T4]}{dt} = k_4[M][T3] - k_{-4}[T4]$$

where k_i is the forward and k_{-i} is the reverse rate constant for each reaction. The second model assumes that aggregates are formed by continuous addition of unfolded monomers:

$$M \rightleftharpoons U$$

$$U + U \rightleftharpoons D$$

$$U + D \rightleftharpoons T3$$

$$U + T3 \rightleftharpoons T4$$

The equations for model 2 are:

$$\frac{d[U]}{dt} = k_1[M] \quad k_{-1}[U] \quad 2k_2[U]^2 \mid 2k_{-2}[D] \quad k_3[U][D] \mid k_{-3}[T3] \quad k_4[U][T3] \mid k_{-4}[T4]$$

$$\frac{d[M]}{dt} = -k_1[M] + k_{-1}[U]$$

$$\frac{d[D]}{dt} = k_2[U]^2 - k_{-2}[D] - k_3[U][D] + k_{-3}[T3]$$

$$\frac{d[T3]}{dt} = k_3[U][D] - k_{-3}[T3] - k_4[U][T3] + k_{-4}[T4]$$

$$\frac{d[T4]}{dt} = k_4[U][T3] - k_{-4}[T4]$$

The concentration of each species was measured over time by SEC, and parameter estimates were obtained by fitting the two kinetic models to the experimental data using the software package Athena Visual Studio. Initially, we assumed that all four reactions were reversible, and each model had 8 parameters. However, in some cases the fitted reverse rate constant was negative or had large confidence intervals, indicating that the parameters were not required. Therefore the model equations were modified to eliminate these rate constants and the number of parameters was reduced to five and six for models 1 and 2, respectively. A summary of the fitted parameters is given in Table IV. There are several points worth noting about the parameter estimates for model 2.

1. We were unable to calculate a 95% confidence interval for k_2, thus leaving the parameter as indeterminate. This suggests that k_2 is not well known. However, k_2 was not removed from the model since it is required for the production of dimers in model 2.
2. k_2 and k_{-2} are two to three orders of magnitude higher than some of the other parameter estimates. The concentration of U at early time points is small and the large k_2 is needed to account for the increase in dimer concentration. However, at later time points as the concentration of U increases, a large k_{-2} is needed to counter the effect of the large k_2.

The AIC scores for model 1 and model 2 were calculated and are summarized in table V. Model 1 has a lower residual sum of squares and a lower AIC score compared to model 2 and therefore is the preferred model. Both models are able to simulate the concentration profiles relatively well (Figure 5), although the deviation from the experimental data for U is noticeably worse for model 2. However, the uncertainty in parameter k_2 as well as the large value for parameters k_2 and k_{-2} in model 2 indicate that model 2 is not an adequate model for describing

the aggregation process. Considering these results, model 1 is a better candidate model to explain the aggregation pathway of cytochrome c, meaning that the aggregation mechanism occurs through a direct non-covalent interaction of *post*-glycated monomeric cytochrome c subunits and the partial unfolding of cytochrome c is off-pathway.

Discussion

One of the key issues to elucidate the pathogenesis of protein deposition diseases is the mechanism by which such proteins are converted from their native-state into high molecular aggregates. At least two major pathways have been identified to date (1). In some cases aggregation starts from fully or partially unfolded conformational states (1). These can be either native states of intrinsically unstructured proteins or non-native states formed by a folded protein through an unfolding event. This unfolding event leads to the exposure of aggregation prone peptide segments, which are able to form intermolecular interactions, thus triggering aggregation (1). More recently, a different aggregation pathway has been identified where aggregation of normally globular proteins can occur, starting from conformational states directly accessible from native states via thermal fluctuations, mutations, *cis*-*trans* proline isomerization or *post*-translational modifications, but in all cases with no need of transition across the major energy barrier for unfolding (34). They can be either transiently formed conformations accessible through fluctuations of the native state (35) or conformational states permanently populated due to *post*-translational modifications or other events, i. e. states with a thermodynamic higher energy than that of the fully folded state (36).

In the current work we have investigated the effects of methylglyoxal-modification of cytochrome c on structural, stability and aggregation properties. When methylglyoxal reacts with proteins, a process involving multi-step non-enzymatic reactions between the carbonyl groups of methylglyoxal and free amino groups in protein molecules leads to the formation of

irreversible AGEs. The side chains of arginine and lysine residues and N-terminal amino group are the available targets for glycation. To characterize *post*-glycation cytochrome c modified residues, MALDI-TOF/TOF was used. Mass spectrometry data showed that only arginine 92 is modified by methylglyoxal with the formation of a hydroimidazolone adduct. This result is consistent with known preferred arginine modification by methylglyoxal (37).

Recent reports have shown that methylglyoxal-modification of different proteins leads to the formation and stabilization of small and soluble aggregated species that retain native-like structure (13; 14). Here we show that, once glycated with methylglyoxal, cytochrome c starts an aggregation process wherein the formed aggregates retain native-like structure. In spite of maintaining the α–helical fold typical of cytochrome c, glycated cytochrome c aggregates have a slight increase of β-sheet content as judged by CD. This increase in β-sheet content may be justified by the aggregation process, since it is known that β-sheet produces a better environment for intermolecular interactions through hydrogen bonding (38). Additionally, tryptophan intrinsic fluorescence did not show any differences between native and glycated (monomer and dimer) cytochrome c species, suggesting that the three-dimensional structure of cytochrome c remains unchanged during aggregation.

Our data suggest that the most profound effect of glycation is the reduction of conformational stability. It is likely that the pronounced decrease of $\Delta G(H_2O)$ from 8.1 ± 0.6 to 3.1 ± 0.6 kCal.mol^{-1} is due to the loss of positive charge(s) at the protein's surface, which can decrease the protein's pI, increase hydrophobicity and slightly change protein core interaction, judging by the decrease of *m* value. A reduction in protein stability is a step to populate unfolded species due to a lower Ku (unfolding constant) and to aggregation, principally when new hydrophobic segments are exposed to solvent. Aggregation leads to increasing conformational stability, possibly by minimizing the number of hydrophobic side-chains exposed to the solvent through the formation of intermolecular hydrogen bonds in the β-sheet segments.

We wondered whether the native monomeric form of glycated cytochrome c is the aggregation unit, or if partial unfolding is required for aggregation. Kinetic modeling of cytochrome c aggregation time course data coupled with model discrimination allowed us to distinguish between two different models. Although both models were able to capture the kinetic data reasonably well, our analysis indicated that the primary means of cytochrome c aggregation is by continuous addition of monomeric glycated cytochrome c to form dimers, trimers and tetramers. The results also suggest that the partial unfolding of monomeric cytochrome c is reversible, but that subsequent aggregation steps are essentially irreversible.

Based on our results, we propose a model for native-like aggregation of methylglyoxal-glycated proteins where the aggregation occurs due to loss of conformational stability of the modified protein. Consequently, an aggregation process is started directly through monomer addition in a process that is thermodynamically and kinetically favoured. Interestingly, the glycated cytochrome c unfolded species is an off pathway byproduct, a clear difference from the amyloidogenic aggregation pathways (figure 6). These results can apparently be extended to the aggregation behaviour of amyloidogenic proteins upon glycation. In fact both insulin and α-synuclein, which are involved in amyloid diseases, show less amyloid fibril formation after glycation and both retain significantly the native three-dimensional structure during the aggregation process. In conclusion, glycation of amyloidogenic proteins can lead to a shift from an amyloidogenic pathway to a native-like aggregation through a process that is thermodynamically and kinetically favoured.

Material and Methods

Protein preparation and glycation

Horse heart cytochrome c (Sigma) was dissolved in 50 mM potassium phosphate buffer, pH 7.0 and 150 mM of NaF to a final concentration of 10 $mg.ml^{-1}$. For the methylglyoxal-

derived glycation of cytochrome c, the protein solution was incubated with 10 mM methylglyoxal (a kind gift from Dr. Carlos Cordeiro, Centro de Química e Bioquímica, FCUL, Lisbon, Portugal) at 37 ºC in sterile conditions. Control samples were treated in the same way but without methylglyoxal addition. Aliquots were collected in sterile conditions at defined times from 0 to 7 days and immediately analysed.

Characterization of cytochrome c glycation by methylglyoxal using mass spectrometry and dot-blot analysis

Methylglyoxal-derived glycation of cytochrome c was characterized by mass spectrometry. Control and glycated cytochrome c was separated by 1D-polyacrylamide gel electrophoresis (1D-PAGE) and in gel trypsin digestion was performed as described (39). Briefly, gel protein bands were washed in milliQ water and destained in 50% acetonitrile (ACN) and subsequently with 100% ACN. Cys residues were reduced with 10 mM DTT and alkylated with 50 mM iodoacetamide. Gel pieces were vacuum dried and rehydrated in digestion buffer containing 50 mM NH_4HCO_3 and 6.7 ng/µL of trypsin (Sigma, proteomics grade, Promega) at 4 ºC. Upon rehydration, the supernatant was discarded and 20 µL of 50 mM NH_4HCO_3 were added. Digestions were allowed to proceed at 37 ºC overnight (16-18 hours). After digestion, the remaining supernatant was removed and stored at -20 ºC until analysis. The obtained peptide mixture was purified and concentrated by solid-phase extraction using home-made R2 Pore microcolumns (Applied Biosystems) as previously described (39). Peptide mixture was eluted directly onto the MALDI target plate with 0.5 µl of α-CHCA matrix (5 $mg.ml^{-1}$) prepared in 50% (v/v) acetonitrile with 0.1% (v/v) formic acid. Sample peptides were analysed in a MALDI-TOF-TOF mass spectrometer 4800 plus (Applied Biosystems) in positive reflectron mode for peptide mass determination. The mass spectrometer was externally calibrated using des-Arg-Bradykinin (904.468 Da), angiotensin 1 (1296.685 Da), Glu-Fibrinopeptide B (1570.677

Da), ACTH (1-17) (2093.087 Da), and ACTH (18-39) (2465.199) (4700 Calibration Mix, Applied Biosystems). Mass spectra were collected in a result-independent acquisition mode, typically using 1000 laser shots per spectrum and a fixed laser intensity of 3500V.

The identification of glycated amino acid residues was performed as described (40; 41). Briefly, a glycated cytochrome c peptide should be exclusively present in the MS spectrum of the modified protein and have a miscleavage associated with a defined mass increment of a specific MAGE (72 Da for the lysine specific MAGE carboxyethyl-lysine (CEL) and 54, 80 and 144 Da for the arginine-specific MAGE hydroimidazolones (MGH), argpyrimidine and tetrahydropirimidine respectively). Tandem MS sequence information allowed the unequivocal identification of MAGE-modified peptides and also assignment of specific modified amino acid. The peptides of interest (*i.e.*, having a mass consistent with the mass increment of the modifications by methylglyoxal) were selected for MS/MS experiments using Collision Induced Dissociation (CID), with 1 kV collision energy and an air pressure of 106 torr. Two thousand laser shots were collected for each MS/MS spectrum using a fixed laser intensity of 4500 V. Raw data were generated by the 4000 Series Explorer Software v3.0 RC1 (Applied Biosystems). The MS/MS data were analysed by DataExplorer™ software V4.5 (Applied Biosystems) combined with manual inspection of the assigned sequence.

Cytochrome c aggregation

Aggregation of cytochrome c upon glycation by methylglyoxal was monitored by size exclusion chromatography (SEC). Samples were analysed by SEC at incubation times of 0 and 7 days, after filtration with a 0.2 µm Whatman filter. SEC was performed with a HPLC Jasco PU-2080 Plus isocratic pump with an UV detector JASCO 2075. The mobile phase was 5 mM sodium phosphate buffer pH 7 with 150 mM NaF. Separation was achieved on a molecular exclusion analytical column (Amersham-Pharmacia Superdex™ 75 10/300 GL) at a flow rate of 0.4 ml/min and the eluting peaks were monitored at 280 nm. The monomeric and unfolded

forms of glycated cytochrome c were also separated by SDS-PAGE on a Bio-Rad Mini-Protean 3 system, using 15 % separation gel and a 4 % stacking gel. Proteins were stained with Comassie Brilliant Blue (42).

Structural characterization of cytochrome c species

After 7 days incubation with methylglyoxal, glycated and control samples of cytochrome c were separated and individually collected by SEC, as described, and analysed by circular dichroism (CD) and fluorescence spectroscopy. Secondary structure analysis was performed by far-UV CD in a Jasco J810 spectropolarimeter equipped with a temperature control unit Julabo F25. Far UV (185-260 nm) CD spectra were recorded with 0.1 cm (linear) path length quartz cuvette. For each spectrum, three scans were averaged. Protein concentration was previously determined by absorbance at 410 nm (ε_{410}=1.06x10^{5} M^{-1} cm^{-1}) in a UV-Visible spectrophotometer Jasco V-530. For protein secondary structure estimation, CD spectra were deconvoluted using CDSSTR (28) deconvolution algorithm on Dichroweb (43; 44). CD spectra of the samples' buffer were recorded and subtracted from the protein spectra.

Tertiary structure changes were probed by tryptophan red shift fluorescence analysis. Fluorescence spectra were measured in a Perkin Elmer LS50B instrument equipped with a Julabo F12 temperature control unit, in quartz cuvettes with 1 cm excitation light path. Intrinsic protein fluorescence emission was recorded between 300 and 400 nm with excitation at 280 nm, an increment of 0.5 nm, an integration time of 1 s and 7 nm slits for both excitation and emission.

Determination of the unfolding pathway for glycated cytochrome c dimer

The presence of intermediates in the unfolding pathway of the glycated cytochrome c dimer was evaluated by SEC in denaturing conditions. To capture eventual unfolding intermediates, glycated monomer and dimer were first separated and collected by SEC in the above-mentioned non-denaturing conditions. Samples (100 µl) were then incubated in 5 mM phosphate buffer pH 7 with 150 mM NaF supplemented with guanidinium hydrochloride (GdnHCl) at the concentrations of 1.5, 2.0 and 2.5 M for 2 days. This range of GdnHCl concentrations is in the unfolding transition phase of the glycated cytochrome c dimer. After protein denaturation, samples were again analysed by SEC. During separation, GdnHCl was present in the mobile phase at sample's concentration.

Conformational stability of cytochrome c species

Conformational stability of glycated and non-glycated cytochrome c species was measured by CD in GdnHCl-induced protein unfolding experiments at 25 ºC after 48 hours incubation. GdnHCl-induced denaturation was found to be reversible for all cytochrome c forms studied, as judged by SEC experiments. CD denaturation curves were constructed using ellipticity values at 222 nm. Monomeric and dimeric species were analysed according to a two-state unfolding model $M \leftrightarrow U$ and $D_2 \leftrightarrow 2U$ respectively using the linear extrapolation method (32) in a non-linear least squares fitting procedure. This yielded values for $\Delta G^{\circ}(H_2O)$, the conformational stability, and m, the dependence of ΔG° on denaturant concentration. C_m, the denaturant concentration at the midpoint of the unfolding transition was calculated as $Cm = \Delta G^{\circ}(\mathrm{H_2O})/m$. Denaturation curves for monomeric species were analysed considering the equation developed by Santoro & Bolen (45; 46). For dimeric species, $\Delta G^{\circ}(H_2O)$ was determined from the estimation of the effective ΔG, as described (31). Briefly, the equilibrium denaturation curve of the glycated cytochrome c dimer was analysed using the same equation as for the monomeric species and the effective ΔG was calculated considering equation 1.

$$\Delta G_{eff}(H_2O) = \frac{m}{n} C_m + R(n), \text{ where } R(n) = \frac{n-1}{n} RT \ln 2 \quad (1)$$

where $\Delta G_{eff}(H_2O)$ represents the effective ΔG; *m/n* is the *m*-value of a monomeric unit of the *n*-mer and *R(n)* is a residual function of *n*, the oligomeric state of the protein. *R(n)* is responsible for the differences in $\Delta G_{eff}(H_2O)$ values between different oligomeric states of a same protein. Assuming that in non-denaturing conditions the fraction of unfolded protein will be minimal (a reasonable assumption for a stable protein under aqueous condition), $\Delta G^{o}(H_2O)$ of the cytochrome c dimer can be directly related to $\Delta G_{eff}(H_2O)$ using equation 2 (31).

$$\Delta G_{eff}(H_2O) = \frac{\Delta G^{o}(H_2O)}{n} + \frac{RT}{n} \ln\left(n C_t^{n-1}\right) \quad (2)$$

where $\Delta G^{o}(H_2O)$ is the conformational stability of the protein and C_t is the total protein molar concentration in monomer units. In equation 2, $\Delta G^{o}(H_2O)/n$ is the $\Delta G^{o}(H_2O)$ per monomer unit (47). This term is concentration-independent and represents an intrinsic property of the protein. In contrast, the second term depends only on the concentration and oligomeric state of the protein, and represents the loss of entropy involved in the oligomerization process (31). Parameters were obtained by fitting the derived equations to the experimental data by non-linear regression using the Solver add-on for Microsoft Excel.

Model discrimination and estimation of the kinetic constants of aggregation

The kinetic parameters of glycated cytochrome c aggregation were determined by time-course analysis. Time-course data were obtained by SEC as described above. Samples were analysed after different incubation periods ranging from 0 to 400 hours. In this analysis, two different models were tested. In model 1, the glycated monomeric cytochrome c is the aggregation unit and the partially unfolded form of cytochrome c is off-pathway. In model 2, it is assumed that glycated cytochrome c primary unfolds to a partially unfolded form that subsequently initiates the aggregation process. The model parameters were regressed using the software program Athena Visual Studio 14.2 (Athena Visual Software, Naperville, IL). Athena Visual Studio provides the user with a Windows based interface for inputs into equation solving and parameter estimation programs. Athena utilizes an implicit integrator package, DDAPLUS, for solving initial-boundary value problems described by sets of mixed differential and algebraic equations (48; 49). Parameter estimation from multi-response data is provided by non-linear least squares regression using the GREGPLUS algorithm (50; 49) while confidence intervals on parameters are calculated from the covariance matrix. Model equations and concentration data for each species (unfolded monomers, monomers, dimers, trimers and tetramers) were programmed into Athena along with initial guesses for the parameter values.

The Akaike Information Criteria (AIC) was used to determine the best model. The AIC is a measure of the relative goodness of fit of a model and is derived via information theory (51). AIC is calculated from the weighted residual sum of squared errors Φ_{resid}, the number of parameters n_{param} and the number of data points n_{data} (equation 3). The model with the lowest AIC value is preferred. The value of AIC for a given model increases as Φ_{resid} and n_{param} increases. Therefore, the AIC discourages overparameterization as increasing the number of parameters will result in a higher AIC score. The weighted residual sum of squares was determined by first weighting the data points. Weighting the data ensures that each data point contributes equally to the total sum of squares. The weighting was performed by assuming

that the error (σ) associated with each data point was 10% of the value of the data point itself. The weighted residual sum of squares was calculated using equation 4 where y_i and $f(x_i)$ refer to the measured concentration and predicted concentration respectively.

$$AIC = \frac{2n_{param}}{n_{data}-n_{param}} + \ln\left[\frac{\Phi_{resid}}{n_{data}-n_{param}}\right] \quad (3)$$

$$\Phi_{resid} = \sum_{i=1}^{n} \frac{(y_i - f(x_i))^2}{\sigma_i^2} \quad (4)$$

Acknowledgements

We thank Dr. Carlos Cordeiro for the gift of the methylglyoxal. We wish to acknowledge Professor Ana Ponces Freire, Dr. António E. N. Ferreira and Dr. Carlos Cordeiro for helpful discussions and all the assistance provided. This work was supported by grants PTDC/QUI/73430/2006, SFRH/BD/23604/2005 (L.M.A.O), SFRH/BPD/41037/2007 (R.A.G.) from the Fundação para a Ciência e a Tecnologia, Ministério da Ciência e Tecnologia, Portugal, and CBET-0930102 (DTY) from the National Science Foundation, USA.

References

1. Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem **75**:333-366.
2. Sunde M, Blake C (1997) The structure of amyloid fibrils by electron microscopy and X-ray diffraction. Adv Protein Chem **50**:123-159.
3. Gosal WS, Myers SL, Radford SE, Thomson NH (2006) Amyloid under the atomic force microscope. Protein Pept Lett **13**:261-270.

4. Nilsson MR (2004) Techniques to study amyloid fibril formation in vitro. Methods **34**:151-160.

5. Krebs MR, Bromley EH, Donald AM (2005) The binding of thioflavin-T to amyloid fibrils: localisation and implications. J Struct Biol **149**:30-37.

6. Brownlee M (1995) Advanced protein glycosylation in diabetes and aging. Annu Rev Med **46**:223-234.

7. Smith MA, Taneda S, Richey PL, Miyata S, Yan SD, Stern D, Sayre LM, Monnier VM, Perry G (1994) Advanced Maillard reaction end products are associated with Alzheimer disease pathology. Proc Natl Acad Sci U S A **91**:5710-5714.

8. Yan SD, Chen X, Schmidt AM, Brett J, Godman G, Zou YS, Scott CW, Caputo C, Frappier T, Smith MA, et al. (1994) Glycated tau protein in Alzheimer disease: a mechanism for induction of oxidant stress. Proc Natl Acad Sci U S A **91**:7787-7791.

9. Castellani R, Smith MA, Richey PL, Perry G (1996) Glycoxidation and oxidative stress in Parkinson disease and diffuse Lewy body disease. Brain Res **737**:195-200.

10. Munch G, Luth HJ, Wong A, Arendt T, Hirsch E, Ravid R, Riederer P (2000) Crosslinking of alpha-synuclein by advanced glycation endproducts--an early pathophysiological step in Lewy body formation? J Chem Neuroanat **20**:253-257.

11. Gomes R, Sousa Silva M, Quintas A, Cordeiro C, Freire A, Pereira P, Martins A, Monteiro E, Barroso E, Ponces Freire A (2005) Argpyrimidine, a methylglyoxal-derived advanced glycation end-product in familial amyloidotic polyneuropathy. Biochem J **385**:339-345.

12. Choi YG, Kim JI, Jeon YC, Park SJ, Choi EK, Rubenstein R, Kascsak RJ, Carp RI, Kim YS (2004) Nonenzymatic glycation at the N terminus of pathogenic prion protein in transmissible spongiform encephalopathies. J Biol Chem **279**:30402-30409.

13. Lee D, Park CW, Paik SR, Choi KY (2009) The modification of alpha-synuclein by dicarbonyl compounds inhibits its fibril-forming process. Biochim Biophys Acta **1794**:421-430.

14. Oliveira LM, Lages A, Gomes RA, Neves H, Familia C, Coelho AV, Quintas A (2011) Insulin glycation by methylglyoxal results in native-like aggregation and inhibition of fibril formation. BMC Biochem **12**:41.

15. Caughey B, Lansbury PT (2003) Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. Annu Rev Neurosci **26**:267-298.

16. Crowther DC, Kinghorn KJ, Miranda E, Page R, Curry JA, Duthie FA, Gubb DC, Lomas DA (2005) Intraneuronal Abeta, non-amyloid aggregates and neurodegeneration in a Drosophila model of Alzheimer's disease. Neuroscience **132**:123-135.

17. Danzer KM, Haasen D, Karow AR, Moussaud S, Habeck M, Giese A, Kretzschmar H, Hengerer B, Kostka M (2007) Different species of alpha-synuclein oligomers induce calcium influx and seeding. J Neurosci **27**:9220-9232.

18. Lauren J, Gimbel DA, Nygaard HB, Gilbert JW, Strittmatter SM (2009) Cellular prion protein mediates impairment of synaptic plasticity by amyloid-beta oligomers. Nature **457**:1128-1132.

19. Guijarro JI, Sunde M, Jones JA, Campbell ID, Dobson CM (1998) Amyloid fibril formation by an SH3 domain. Proc Natl Acad Sci U S A **95**:4224-4228.

20. Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, Dobson CM (1999) Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci U S A **96**:3590-3594.

21. Villegas V, Zurdo J, Filimonov VV, Aviles FX, Dobson CM, Serrano L (2000) Protein engineering as a strategy to avoid formation of amyloid fibrils. Protein Sci **9**:1700-1708.

22. Chow MK, Ellisdon AM, Cabrita LD, Bottomley SP (2004) Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. J Biol Chem **279**:47643-47651.

23. Pedersen JS, Christensen G, Otzen DE (2004) Modulation of S6 fibrillation by unfolding rates and gatekeeper residues. J Mol Biol **341**:575-588.

24. Plakoutsi G, Taddei N, Stefani M, Chiti F (2004) Aggregation of the Acylphosphatase from Sulfolobus solfataricus: the folded and partially unfolded states can both be precursors for amyloid formation. J Biol Chem **279**:14111-14119.

25. Bemporad F, Chiti F (2009) "Native-like aggregation" of the acylphosphatase from Sulfolobus solfataricus and its biological implications. FEBS Lett **583**:2630-2638.

26. Singh SM, Hutchings RL, Mallela KM (2011) Mechanisms of m-cresol-induced protein aggregation studied using a model protein cytochrome c. J Pharm Sci.

27. Sreerama N, Woody RW (2004) Computation and analysis of protein circular dichroism spectra. Methods in enzymology **383**:318-351.

28. Johnson WC (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. Proteins **35**:307-312.

29. Calvert JF, Hill JL, Dong A (1997) Redox-dependent conformational changes are common structural features of cytochrome c from various species. Archives of biochemistry and biophysics **346**:287-293.

30. Hammack B, Attfield K, Clayton D, Dec E, Dong A, Sarisky C, Bowler BE (1998) The magnitude of changes in guanidine-HCl unfolding m-values in the protein, iso-1-cytochrome c, depends upon the substructure containing the mutation. Protein Sci **7**:1789-1795.

31. Park C, Marqusee S (2004) Analysis of the stability of multimeric proteins by effective DeltaG and effective m-values. Protein Sci **13**:2553-2558.

32. Pace CN (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods in enzymology **131**:266-280.

33. Bartalesi I, Bertini I, Ghosh K, Rosato A, Turano P (2002) The unfolding of oxidized c-type cytochromes: the instructive case of Bacillus pasteurii. J Mol Biol **321**:693-701.

34. Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. Nat Chem Biol **5**:15-22.

35. Canet D, Last AM, Tito P, Sunde M, Spencer A, Archer DB, Redfield C, Robinson CV, Dobson CM (2002) Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. Nat Struct Biol **9**:308-315.

36. Banci L, Bertini I, D'Amelio N, Gaggelli E, Libralesso E, Matecko I, Turano P, Valentine JS (2005) Fully metallated S134N Cu,Zn-superoxide dismutase displays abnormal mobility and intermolecular contacts in solution. J Biol Chem **280**:35815-35821.

37. Lo TW, Westwood ME, McLellan AC, Selwood T, Thornalley PJ (1994) Binding and modification of proteins by methylglyoxal under physiological conditions. A kinetic and mechanistic study with N alpha-acetylarginine, N alpha-acetylcysteine, and N alpha-acetyllysine, and bovine serum albumin. J Biol Chem **269**:32299-32305.

38. Creighton TE. 1993. Proteins: Structures and Molecular Properties, W. H. Freeman and Company, New York.

39. Gobom J, Nordhoff E, Mirgorodskaya E, Ekman R, Roepstorff P (1999) Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. J Mass Spectrom **34**:105-116.

40. Gomes RA, Vicente Miranda H, Silva MS, Graca G, Coelho AV, Ferreira AE, Cordeiro C, Freire AP (2006) Yeast protein glycation in vivo by methylglyoxal. Molecular modification of glycolytic enzymes and heat shock proteins. FEBS J **273**:5273-5287.

41. Gomes RA, Oliveira LM, Silva M, Ascenso C, Quintas A, Costa G, Coelho AV, Sousa Silva M, Ferreira AE, Ponces Freire A, Cordeiro C (2008) Protein glycation in vivo: functional and structural effects on yeast enolase. Biochem J **416**:317-326.

42. Wilson CM (1979) Studies and critique of Amido Black 10B, Coomassie Blue R, and Fast Green FCF as stains for proteins after polyacrylamide gel electrophoresis. Analytical biochemistry **96**:263-278.

43. Lobley A, Whitmore L, Wallace BA (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. Bioinformatics (Oxford, England) **18**:211-212.

44. Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic acids research **32**:W668-673.

45. Bolen DW, Santoro MM (1988) Unfolding free energy changes determined by the linear extrapolation method. 2. Incorporation of delta G degrees N-U values in a thermodynamic cycle. Biochemistry **27**:8069-8074.

46. Santoro MM, Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry **27**:8063-8068.

47. Boudker O, Todd MJ, Freire E (1997) The structural stability of the co-chaperonin GroES. J Mol Biol **272**:770-779.

48. Caracotsios M, Stewart WE (1985) Sensitivity analysis of initial value problems with mixed ODEs and algebraic equations. Comput Chem Eng **9**:359–365.

49. Caracotsios M, Stewart WE. 2008. Computer-Aided Modeling of Reactive Systems, Wiley, Hoboken, NJ.

50. Stewart WE, Caracotsios M, Sorensen JP (1992) Parameter estimation from multiresponse data. AIChE J **38**:641–650.

51. Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**:716–723.

Tables

Table I - Assignment of glycated amino acid residues. In all cases, the observed peptide mass has a mass increment specific of a methylglyoxal-derived AGE modification. The specific MAGE are indicated in the table and the modified amino acid residue is highlighted in the peptide sequence. For these three glycated cytochrome c peptides, tandem MS data confirms the glycated arginine 92. (MGH – hydroimidazolone).

Observed mass (Da)	Theoretical mass (Da)	Peptide sequence	Mass Increase (Da)	MAGE	Glycated residue
1404.727	1350.726	TE**R**EDLIAYLK (90-100)	54.001	MGH	R92
1532.821	1478.821	KTE**R**EDLIAYLK (89-100)	54.000	MGH	R92
1660.911	1606.916	KKTE**R**EDLIAYLK (88-100)	53.995	MGH	R92

Table II – Distribution of the structural element fractions obtained by deconvolution of CD spectra using CDSSTR algorithm available on Dichroweb. The NRMSD parameter represents the normalized root mean square deviance.

Structural element	α-Helix	β-Sheet	β-Turns	Unordered structure	NRMSD
Native Cytochrome c	0.32	0.08	0.18	0.41	0.024
Glycated Monomer	0.32	0.08	0.19	0.40	0.021
Glycated Dimer	0.25	0.19	0.15	0.40	0.056
Glycated Trimer	0.24	0.19	0.15	0.41	0.057

Table III – Thermodynamic parameters from GdnHCl unfolding studies of native and glycated cytochrome c. Parameters were obtained by a direct fit of the model equations to experimental data in figure 4C. $\Delta G^{o}(H_2O)$, protein conformational stability; *m*, the dependence of ΔG^{o} on denaturant concentration; *Cm*, the denaturant concentration at the midpoint of the unfolding transition.

	$\Delta G^{o}(H_2O)$ (kcal·mol^{-1})	*m* (kcal·mol^{-1}.M^{-1})	*Cm* (M)
Native Cytchrome c	8.1 ± 0.6	3.2 ± 0.3	2.5 ± 0.4
Glycated Monomer	3.1 ± 0.6	1.4 ± 0.2	2.2 ± 0.8
Glycated Dimer	8.4 ± 1.3	-	2.2 ± 0.5

Table IV: Fitted parameter values and their 95% confidence intervals obtained by regression to experimental data. A tilde (~) between brackets indicate an indeterminate parameter. NR = not required

	k_1 (h^{-1})	k_{-1} (h^{-1})	k_2 ($mM^{-1}\ h^{-1}$)	k_{-2} (h^{-1})	k_3 ($mM^{-1}\ h^{-1}$)	k_4 ($mM^{-1}\ h^{-1}$)
Model 1	1.6 ± 0.4 x 10^{-3}	5.7 ± 2.3 x 10^{-3}	6.9 ± 0.4 x 10^{-4}	NR	3.2 ± 0.2 x 10^{-3}	4.5 ± 0.5 x 10^{-3}
Model 2	3.4 ± 0.4 x 10^{-3}	1.1 ± 0.2 x 10^{-2}	2.6 (~)	6.4 ± 1.3 x 10^{-1}	1.6 ± 0.2 x 10^{-2}	2.2 ± 0.4 x 10^{-3}

Table V: The Akaike Information Criteria (AIC) score for models 1 and 2.

	n_{param}	n_{data}	Weighted Φresid	AIC score
Model 1	5	45	2.17×10^2	1.94
Model 2	6	45	7.73×10^2	3.29

Figure Legends

Figure 1 - Detection of MAGE-modified cytochrome c peptides and location of glycated amino acids by mass spectrometry. A - Representative section of the MALDI-TOF/TOF spectra of peptides from non-modified and glycated cytochrome c. New *m/z* peaks, absent from the control, are clearly detected in the mass spectra of the glycated cytochrome c (highlighted in red). These new *m/z* values correspond to a cytochrome c peptide plus the mass increment characteristic of a hydroimidazolone modification (54 Da). These peptides were analyzed by MS/MS, confirming the glycation of the arginine residue 92. B - Tandem MS spectrum of a glycated cytochrome c peptide with *m/z* 1404.7269, showing the y and b fragment ions. The detected fragment ions arise from the amino acid sequence TE**R**EDLIAYLK, with a hydroimidazolone modification on the arginine residue. C - Tandem MS spectrum of a glycated cytochrome c peptide with *m/z* 1532.8164, showing the y and b fragment ions. The detected fragment ions arise from the amino acid sequence KTE**R**EDLIAYLK, with a hydroimidazolone modification on the arginine residue.

Figure 2 – Size exclusion chromatograms of native and glycated cytochrome c. A – Native cytochrome c; B – Cytochrome c after incubation with methylglyoxal. All samples have the same protein concentration (0.85 mM) and were eluted with a flow rate of 0.4 ml/min. The labels in chromatograms have the following meaning: T4 – Glycated tetramer; T3 – Glycated trimer; D – Glycated dimer; M – Glycated monomer; U – High elution volume species.

To understand the nature of the U species, a SDS-PAGE gel was performed comparing the migration of the U and M species (C). The gel shows an identical migration for both species discarding the possibility of cytochrome c fragments and suggesting that the U species are in fact an unfolded form of cytochrome c.

Figure 3 - Structural characterization of the different species of native and glycated cytochrome c. A – Secondary structure of native cytochrome c (black), glycated monomer (blue), dimer (red) and trimer (green) evaluated by far-UV CD. B – Far-UV CD spectra of the high elution volume species (U) showing absence of secondary structure elements. C – Tertiary structure of native cytochrome c (black solid) and glycated monomer (blue), dimer (red) and high elution volume species (U) (black dash) evaluated by intrinsic fluorescence emission. All experiments were made in 5 mM phosphate buffer pH 7 supplemented with 150 mM NaF at 25 ºC after 7 days incubation with methylglyoxal. Native cytochrome c was submitted to the same conditions without the presence of methylglyoxal.

Figure 4 – Determination of the conformational stability of native and glycated cytochrome c species. A - To determine the unfolding mechanism of cytochrome c glycated species, glycated dimer was collected from SEC in non-denaturing conditions and incubated with 1.5 (black), 2 (dark grey) and 2.5 M (light grey) of GdnHCl at 25 ºC for 2 days. Then aliquots were analysed by SEC and eluted with the same incubation buffer. The chromatograms show no unfolding intermediates. B – Comparison of the unfolding pathways of the glycated monomer (blue) and glycated dimer (red). Both species were previously separated and collected from SEC in non-denaturing conditions, incubated with 2 M GdnHCl for 2 days and then analysed by SEC containing GdnHCl in the mobile phase. The chromatograms show that both species unfold to the same denatured monomeric form and no intermediates are detected to any measurable extent. C - Guanidinium hydrochloride equilibrium denaturation curves of native cytochrome c (black – primary axis), glycated monomer (blue – secondary axis) and glycated dimer (red – secondary axis) at pH 7 and 25 ºC monitored by ellipticity at 222 nm. For the monomeric forms of cytochrome c, the equation used represents a two-state unfolding model (M↔U), while for the dimeric form, $\Delta G º(H_2O)$ was determined considering the effective ΔG (31) and applying a two-state unfolding model (D_2↔2U). The curves are non-linear least squares fits to the

equations representing the entire denaturation curve and using a linear extrapolation method to the experimental circular dichroism data. The insets are the residues plots.

Figure 5 - Model simulations using parameter estimates from Athena Visual Studio for model 1 (solid line) and model 2 (dashed line). Different models were fitted to experimental time course data obtained by SEC, describing the variation of each glycated cytochrome c species along time. Model 1 assumes that the aggregation proceeds through folded monomers with dimers, timers and tetramers formed by monomer addition. Model 2 assumes that larger aggregate species are formed by continuous addition of unfolded monomers. A - Concentration of monomers. B - Concentration of Unfolded monomers. C - Concentration of dimers. D - Concentration of trimers. E - Concentration of tetramers.

Figure 6 - Molecular model for cytochrome c aggregation upon methylglyoxal glycation. The native protein is primarily irreversibly glycated by methylglyoxal forming a glycated monomer. This monomer, which has a reduced stability, can follow two different pathways. It can undergo a rapid equilibrium with a partial unfolded monomeric form, or it can irreversibly go through an aggregation pathway by the sequential addition of glycated monomers.

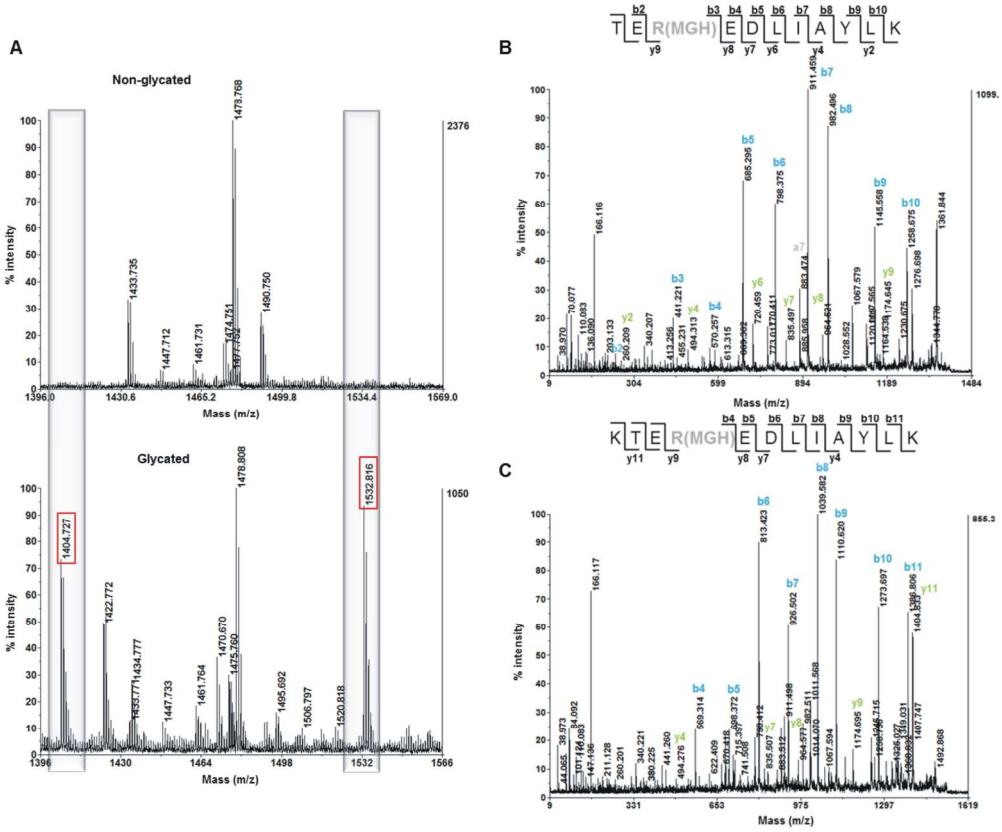

Figure 1 - Detection of MAGE-modified cytochrome c peptides and location of glycated amino acids by mass spectrometry. A - Representative section of the MALDI-TOF/TOF spectra of peptides from non modified and glycated cytochrome c. New m/z peaks, absent from the control, are clearly detected in the mass spectra of the glycated cytochrome c (highlighted in red). These new m/z values correspond to a cytochrome c peptide plus the mass increment characteristic of a hydroimidazolone modification (54 Da). These peptides were analyzed by MS/MS, confirming the glycation of the arginine residue 92. B - Tandem MS spectrum of a glycated cytochrome c peptide with m/z 1404.7269, showing the y and b fragment ions. The detected fragment ions arise from the amino acid sequence TEREDLIAYLK, with a hydroimidazolone modification on the arginine residue. C - Tandem MS spectrum of a glycated cytochrome c peptide with m/z 1532.8164, showing the y and b fragment ions. The detected fragment ions arise from the amino acid sequence KTEREDLIAYLK, with a hydroimidazolone modification on the arginine residue.

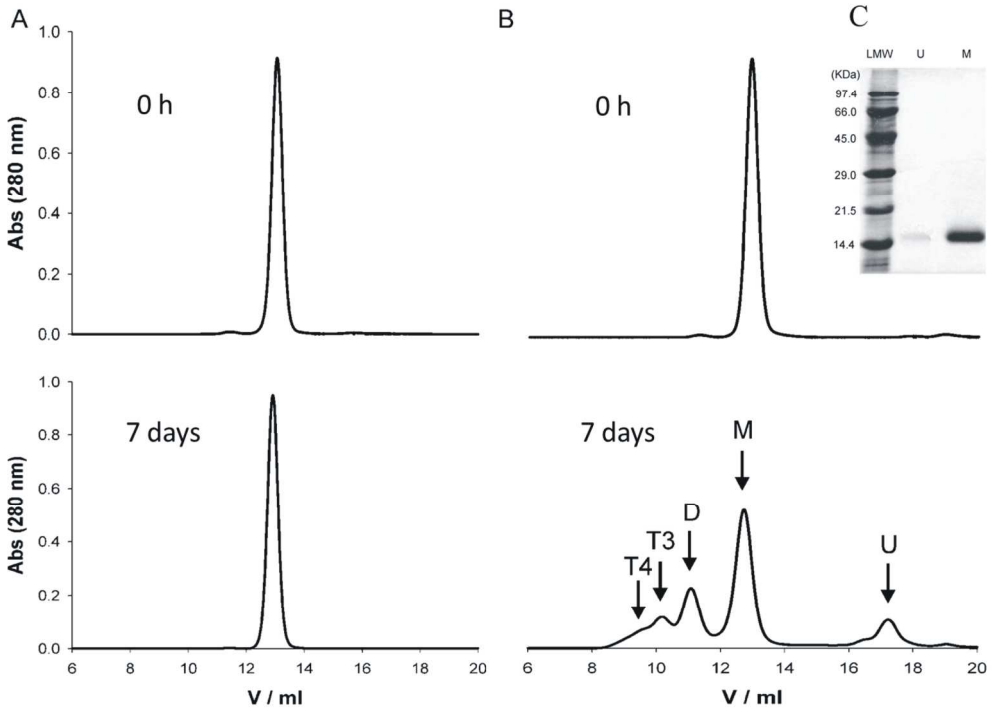

Figure 2 – Size exclusion chromatograms of native and glycated cytochrome c. A – Native cytochrome c; B – Cytochrome c after incubation with methylglyoxal. All samples have the same protein concentration (0.85 mM) and were eluted with a flow rate of 0.4 ml/min. The labels in chromatograms have the following meaning: T4 – Glycated tetramer; T3 – Glycated trimer; D – Glycated dimer; M – Glycated monomer; U – High elution volume species.

To understand the nature of the U species, a SDS-PAGE gel was performed comparing the migration of the U and M species (C). The gel shows an identical migration for both species discarding the possibility of cytochrome c fragments and suggesting that the U species are in fact an unfolded form of cytochrome c.

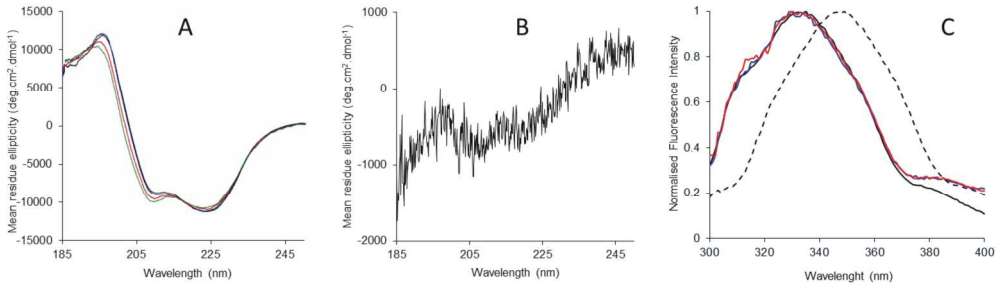

Figure 3 - Structural characterization of the different species of native and glycated cytochrome c. A – Secondary structure of native cytochrome c (black), glycated monomer (blue), dimer (red) and trimer (green) evaluated by far-UV CD. B – Far-UV CD spectra of the high elution volume species (U) showing absence of secondary structure elements. C – Tertiary structure of native cytochrome c (black solid) and glycated monomer (blue), dimer (red) and high elution volume species (U) (black dash) evaluated by intrinsic fluorescence emission. All experiments were made in 5 mM phosphate buffer pH 7 supplemented with 150 mM NaF at 25 ºC after 7 days incubation with methylglyoxal. Native cytochrome c was submitted to the same conditions without the presence of methylglyoxal.

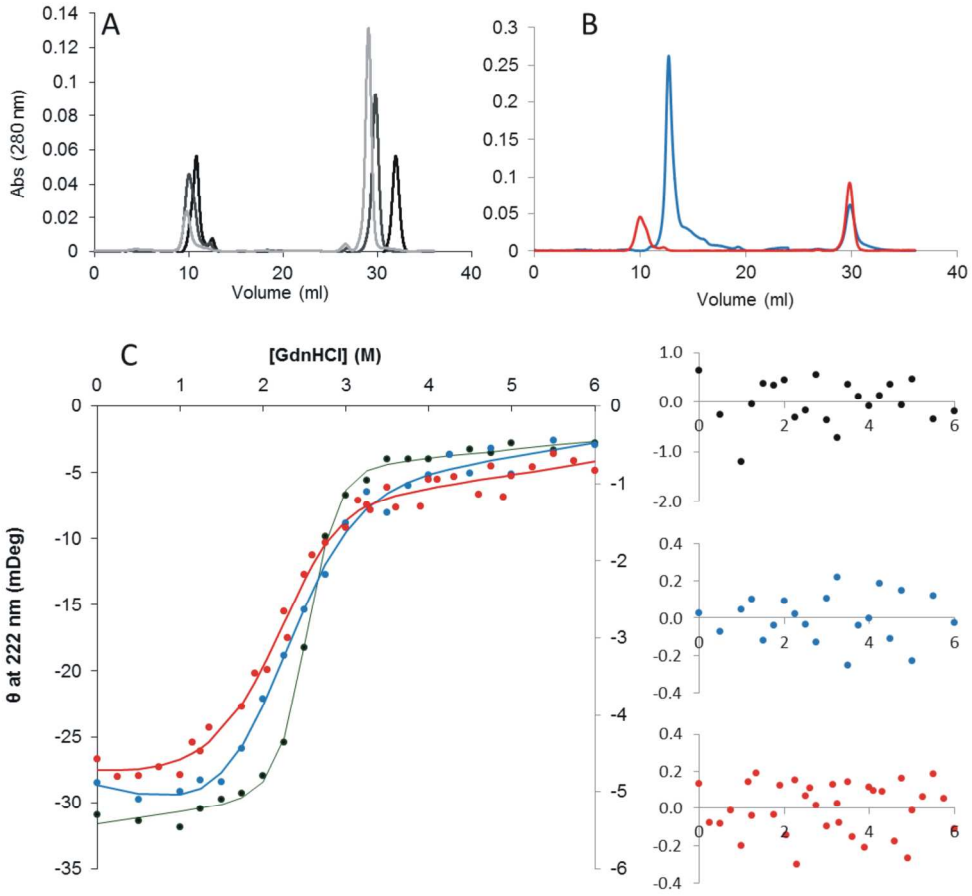

Figure 4 – Determination of the conformational stability of native and glycated cytochrome c species. A - To determine the unfolding mechanism of cytochrome c glycated species, glycated dimer was collected from SEC in non-denaturing conditions and incubated with 1.5 (black), 2 (dark grey) and 2.5 M (light grey) of GdnHCl at 25 ºC for 2 days. Then aliquots were analysed by SEC and eluted with the same incubation buffer. The chromatograms show no unfolding intermediates. B – Comparison of the unfolding pathways of the glycated monomer (blue) and glycated dimer (red). Both species were previously separated and collected from SEC in non-denaturing conditions, incubated with 2 M GdnHCl for 2 days and then analysed by SEC containing GdnHCl in the mobile phase. The chromatograms show that both species unfold to the same denatured monomeric form and no intermediates are detected to any measurable extent. C - Guanidinium hydrochloride equilibrium denaturation curves of native cytochrome c (black – primary axis), glycated monomer (blue – secondary axis) and glycated dimer (red – secondary axis) at pH 7 and 25 ºC monitored by ellipticity at 222 nm. For the monomeric forms of cytochrome c, the equation used represents a two-state unfolding model (M↔U), while for the dimeric form, $\Delta G^{o}(H2O)$ was determined considering the effective ΔG (31) and applying a two-state unfolding model (D2↔2U). The curves are non-linear least squares fits to the equations representing the entire denaturation curve and using a linear extrapolation method to the experimental circular dichroism data. The insets are the residues plots.

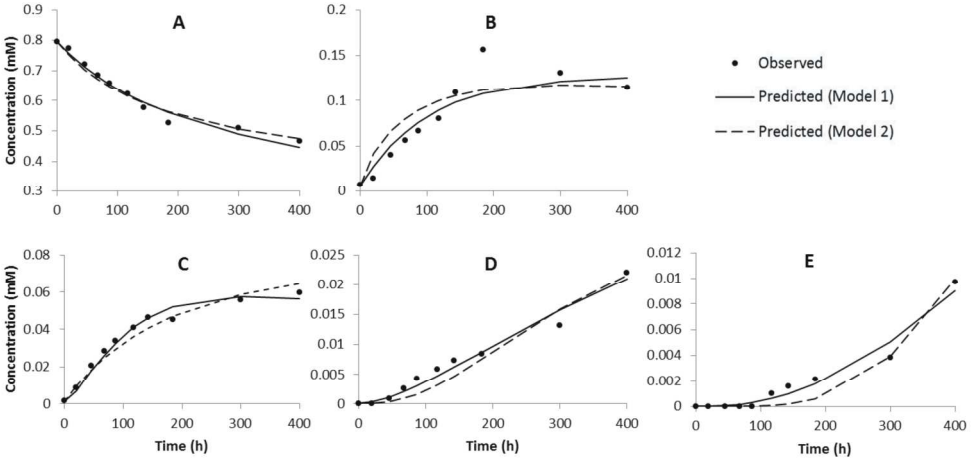

Figure 5 - Model simulations using parameter estimates from Athena Visual Studio for model 1 (solid line) and model 2 (dashed line). Different models were fitted to experimental time course data obtained by SEC, describing the variation of each glycated cytochrome c species along time. Model 1 assumes that the aggregation proceeds through folded monomers with dimers, timers and tetramers formed by monomer addition. Model 2 assumes that larger aggregate species are formed by continuous addition of unfolded monomers. A - Concentration of monomers. B - Concentration of Unfolded monomers. C - Concentration of dimers. D - Concentration of trimers. E - Concentration of tetramers.

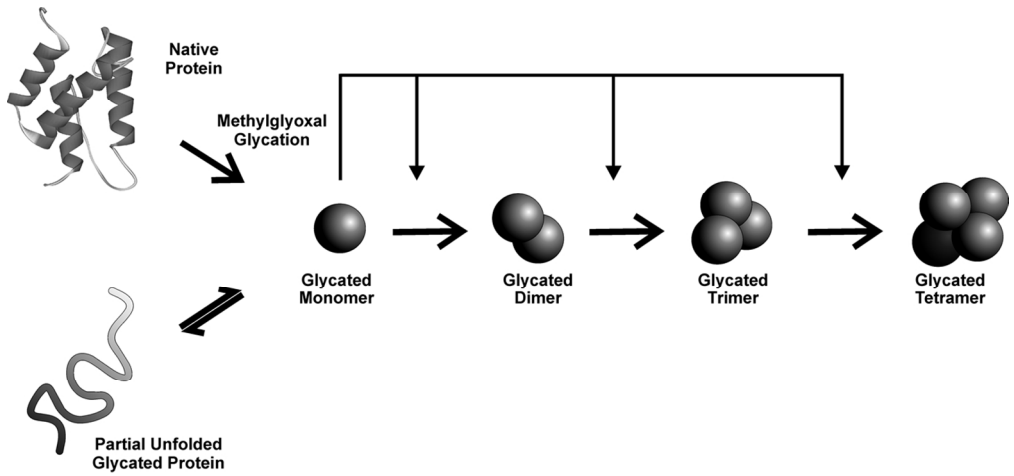

Figure 6 - Molecular model for cytochrome c aggregation upon methylglyoxal glycation. The native protein is primarily irreversibly glycated by methylglyoxal forming a glycated monomer. This monomer, which has a reduced stability, can follow two different pathways. It can undergo a rapid equilibrium with a partial unfolded monomeric form, or it can irreversibly go through an aggregation pathway by the sequential addition of glycated monomers.