# Molecular Mechanisms Underlying Primary Hyperoxaluria Type 1 and New Therapeutic Approaches

Noel Mesa Torres

Tesis Doctoral


**Universidad de Granada**
**Programa de Doctorado en Química**
**Granada, 2014**

---

Departamento de Química Física
Facultad de Ciencias

Editor: Editorial de la Universidad de Granada
Autor: Noel Mesa Torres
D.L.: GR 264-2015
ISBN: 978-84-9083-276-9

Memoria presentada para optar al Grado de Doctor por la Universidad de Granada en el Programa de Doctorado en Química.

Granada, 24 de septiembre de 2014

Fdo.: Noel Mesa Torres
Ingeniero Químico por la Universidad de Granada

VºBº Director de la tesis:

Fdo.: Dr. Ángel L. Pey Rodríguez
Departamento de Química Física
Facultad de Ciencias
Universidad de Granada

El doctorando Noel Mesa Torres y el director de la tesis Dr. Ángel Luis Pey Rodríguez garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 24 de Septiembre de 2014

| | |
|---|---|
| Director de la Tesis | Doctorando |
| Fdo.: Dr. Ángel L. Pey Rodríguez | Fdo.: Noel Mesa Torres |

*A mis padres,*
*a Lola,*
*a mis amigos y compañeros.*
*Gracias por estar siempre ahí.*

«*Ingrato es el que sólo en secreto es agradecido. Ingrato es quien niega el beneficio recibido; ingrato, quien no lo restituye; pero de todos, el más ingrato es quien lo olvida. Nadie apunta en su agenda los favores recibidos.*» Lucio Anneo Séneca.

Enfrentarse a nuevos retos es estimulante y conlleva cierto vértigo. Por suerte, he sido muy afortunado y durante los últimos cuatro años he estado rodeado de grandes personas sin las cuales esta tesis doctoral no hubiera sido posible en modo alguno.

En primer lugar, he de dar las gracias a José Manuel por darme la oportunidad de aprender y formar parte de su laboratorio. Gracias también a Beatriz y a mis compañeros Inma, Álvaro, Sergio, Valeria, Fadia, Asunción, Encarnación, Esperanza y Mª Carmen, por ayudarme siempre que lo necesité. También he de agradecer a todos las personas que forman parte del departamento (en especial a Antonio Parody).

Mención especial se merecen aquellos que me soportaron día a día. Personas de las que he aprendido mucho (dentro y fuera del laboratorio) y que han marcado un momento muy importante de mi vida. Carles, Bertrand, Fran, Rocío, Pedro y Adela, ¡Gracias!. No obstante, si hay alguien que me ha tenido que soportar especialmente es la persona que me ha dirigido y enseñado durante estos años. Una persona que ha tenido una gran paciencia y ha creído en mí. ¡Muchas gracias Ángel!.

Quiero dar las gracias a Eduardo, Armando e Israel. También quiero dar las gracias a Antonio, María, Fran y Ana por enseñarme tan pacientemente a trabajar con los *gusanillos*.

Por último, aunque no menos importante, me gustaría dar las gracias a mi familia. A Lola y a mis padres, porque su cariño y apoyo incondicional ha sido, y es, la piedra angular de mi vida.

**ABSTRACT.** Primary hyperoxaluria type 1 (PH1) is an autosomal recessive inborn error of metabolism caused by mutations in the *AGXT* gene, which encodes for the enzyme alanine:glyoxylate aminotransferase (AGT).[1] Two main non-pathogenic polymorphisms are found in the *AGXT* gene: the most common *major* allele and the so-called *minor* allele, which is the most frequent in patients. The biological role of the human AGT enzyme is to create a glyoxylate sink in peroxisomes of hepatocytes through the transamination of L-alanine to pyruvate and glyoxylate to glycine in the presence of the coenzyme pyridoxal-5'-phosphate (PLP).[2,3] AGT deficiency causes glyoxylate accumulation that is subsequently oxidized to oxalate. Calcium oxalate crystals result in a progressive renal failure and a life-threatening systemic build-up of oxalate.[4] A limited number of genotypes have been reported to respond to pharmacological doses of pyridoxine (PLP precursor), but the only curative option is the liver and kidney transplantation, which causes significant morbidity and mortality.[5] About 150 mutations have been found in the *AGXT* gene and several molecular mechanisms seem to contribute to AGT loss-of-function such as mitochondrial mistargeting, protein aggregation, accelerated degradation or catalytic effects.[6]

In this thesis:

1) We have performed a comprehensive *in vitro* characterization of the most common non-pathogenic polymorphisms and disease-causing mutations to obtain information about disease mechanisms. We found that the apo state of PH1 mutations, on the *minor* allele, display a much lower kinetic stability that is overcome by the binding of coenzyme through stabilization of the native structure.[7] Moreover, we have observed a partial correlation between destabilization of apo AGT dimer, kinetic trapping by molecular chaperones of partially folded states, intracellular reduced foldability and enhanced mistargeting in cells, suggesting a key role of the protein homeostasis defects in the pathogenesis of PH1.[8–11]

2) In order to increase our knowledge about the *in vivo* interactions of AGT with elements of the protein homeostasis system along its folding process, we have considered whether the nematode *C. elegans* is a suitable model for PH1.[12,13] We have characterized an orthologous of human AGT protein in the nematode and also discussed the role of glyoxylate in *C. elegans*. Our results suggest that this nematode might not be suitable for a hyperoxaluric phenotype but *C. elegans* could be a proper model to study proteostasis interactions of AGT.

3) We have applied the consensus approach to engineer an enhanced human AGT protein that could be potentially used in gene therapy (GT) or enzyme replacement therapy (ERT).[14–16] This simple approach has been suitable for enhancing the *in vitro* AGT activity and stability by the optimization of interactions in the native structure. The efficacy of the consensus approach to improve GT and ERT is being currently tested. Nonetheless, the improvement in the crystallization process of engineered variants suggests that this approach could also be a simple strategy to obtain structural models of proteins that display stability issues.

4) Finally, we have also studied the energetic basis underlying mutational effects on AGT kinetic stability by considering non-pathogenic, disease-causing and consensus-based variants. Chemical and thermal denaturations support that kinetic stability of human AGT depends on changes in thermodynamic stability and aggregation propensity of partially/globally unfolded states. We propose that the *minor* allele is a lower limit for AGT stability and foldability that may explain the high frequency of misfolding mutants in PH1 and the pivotal role of molecular chaperones in the PH1 pathogenesis.[17–19]

**RESUMEN.** La hiperoxaluria primaria tipo 1 (HP1) es una enfermedad genética autosómica recesiva causada por mutaciones sobre el gen *AGXT* que codifica para la enzima alanina:glioxilato aminotransferasa (AGT).[1] El gen *AGXT* presenta dos polimorfismos no patogénicos: el más común alelo mayor, y el denominado alelo menor, el cual es más frecuente en pacientes. La función biológica de la enzima AGT humana es crear un sumidero de glioxilato en los peroxisomas de hepatocitos mediante la transaminación de L-alanina a piruvato y glioxilato a glicina en presencia del coenzima piridoxal-5'-fosfato (PLP).[2,3] La deficiencia en la función de la enzima AGT provoca la acumulación de glioxilato, el cual puede oxidarse hasta oxalato. La formación de cristales de oxalato cálcico conduce a un progresivo fallo renal y a la deposición de oxalato en todo el organismo provocando un fallo multi-sistémico.[4] Solo un número limitado de genotipos han mostrado respuesta al tratamiento farmacológico con piridoxina (precursor de PLP), y el único tratamiento genérico es el transplante simultáneo de hígado y riñón, el cual está asociado con unos altos niveles de morbilidad y mortalidad.[5] Más de 150 mutaciones han sido encontradas sobre el gen *AGXT* y varios mecanismos moleculares parecen contribuir a la pérdida de función de la enzima AGT, tales como localización errónea en mitocondria, agregación, degradación acelerada o defectos catalíticos.[6]

En esta tesis:

1) Hemos realizado un estudio *in vitro* para caracterizar los polimorfismos no patogénicos y causantes de enfermedad más comunes para obtener información sobre los mecanismos de la enfermedad. Hemos encontrado una baja estabilidad cinética del estado apo de las mutaciones de HP1 sobre el alelo menor, la cual es corregida mediante la unión del coenzima por estabilización del estado nativo.[7] Además, hemos observado una correlación parcial entre la desestabilización del estado apo del dímero AGT, y el aumento de estados parcialmente plegados asociados a chaperonas moleculares, la reducción de la plegabilidad y el aumento de la localización subcelular errónea en células[8–11]. Esto sugiere que la homeostasis proteica juega un papel importante en la patogenicidad de HP1.

2) Con el objetivo de aumentar nuestro conocimiento sobre las interacciones *in vivo* de la enzima AGT con diferentes elementos del sistema de homeostasis proteica durante su plegamiento, hemos considerado el empleo del nemátodo *C. elegans* como posible modelo para la HP1.[12,13] Para ello hemos realizado la caracterización *in vitro* de una proteína ortóloga de la enzima AGT humana en *C. elegans* y discutido el papel del glioxilato en este nemátodo con el fin de crear un fenotipo de hiperoxaluria.

3) Hemos aplicado la aproximación de consenso para diseñar una enzima AGT humana mejorada que pueda ser potencialmente usada en terapia génica o reemplazamiento enzimático.[14–16] Esta simple aproximación ha resultado adecuada para incrementar la actividad y estabilidad *in vitro* de la enzima AGT mediante la mejora de las interacciones de la estructura nativa. El empleo de esta aproximación para mejorar la terapia génica y de reemplazamiento enzimático está aún bajo estudio. Sin embargo, la mejora obtenida en el proceso de cristalización de variantes de consenso sugiere que esta aproximación podría también ser empleada como una estrategia simple para la cristalización de proteínas con problemas de estabilidad.

4) Finalmente, también hemos estudiado las bases energéticas que subyacen los efectos mutacionales sobre la estabilidad de la enzima AGT, considerando desde variantes no patogénicas, patogénicas y de consenso. Los estudios por desnaturalización química y térmica indican que la estabilidad cinética de la enzima AGT humana depende de cambios en la estabilidad termodinámica y en la propensidad a fenómenos de agregación de los estados desplegados o parcialmente desplegados,

donde la existencia de un límite inferior de estabilidad y plegabilidad para el alelo menor, puede explicar la alta frecuencia de mutantes HP1 que afectan al plegamiento, destacando el papel principal de las chaperonas moleculares en su patogenicidad.[17–19]

# Abbreviations

| | |
|---|---|
| AGT: | alanine:glyoxylate aminotransferase |
| ASA: | accesible surface area |
| CaOx: | calcium oxalate |
| CD: | circular dichroism |
| CHO: | chinese hamster ovary |
| DSC: | differential scanning calorimetry |
| $\Delta H$: | enthalpy of unfolding |
| $E_a$: | activation energy |
| ERT: | enzyme replacement therapy |
| GT: | gene therapy |
| ORF: | open reading frame |
| PH: | primary hyperoxaluria |
| PH1: | primary hyperoxaluria type 1 |
| PLP: | pyridoxal-5'-phosphate |
| PMP: | pyridoxamine-5'-phosphate |
| SEC: | size exclusion chromatography |
| TCA: | tricarboxylic acid |
| $T_m$: | temperature where the maximun of the transition occurs |

# Contents

# Part 1

# Introduction

CHAPTER 1

# Hyperoxalurias

The term **hyperoxaluria** is defined as the clinical sign characterized by an elevated level of urinary oxalate.[20] Despite oxalate is a useful metabolite for plants and some bacteria, in mammals is a final product of metabolism where the production:elimination ratio is finely balanced. In humans, the main route to remove oxalate is by excretion through the urinary tract. Therefore, the kidney plays an important role in oxalate elimination, and a direct relation between plasma concentration and urinary levels of oxalate has been found.[20] At physiological pH, oxalate is able to form soluble salts with sodium and potassium, but in the presence of calcium it results in the formation of insoluble calcium oxalate (CaOx) that is the main component of kidney stones. This physical property of its calcium salt is the reason of the life-threatening properties of high oxalate levels.

High levels of oxalate synthesis results in a systemic deposition of kidney stones (urolithiasis), that leads to a progressive renal failure until a final stage characterized by a widespread CaOx deposition (oxalosis), which causes a multisystemic organ failure.[21]

Hyperoxaluria can be classified in two categories according to the origin of the oxalate. In **primary hyperoxalurias** (PH), the cause of the excessive oxalate is genetic while in **secondary hyperoxalurias** is environmental. However, in both cases the clinical signs are similar and indistinguishable.[20]

The term primary hyperoxaluria covers a number of genetic disorders involved in glyoxylate metabolism. To date, mutations in three different autosomic genes have been described to be involved in PH with recessive inheritance pattern. Due to the low rate of incidence, these disorders are classified as rare diseases*.[22] The **primary hyperoxaluria type 1** (PH1) (*Mendelian Inheritance in Man* number MIM# 259900) is the result of mutations on the *AGXT*

*According to the European Commission on Public Health, rare disease is defined as «life-threatening disease that is of such low prevalence (fewer than 1 in 2000 people) that special combined efforts are needed to address them»

gene that codifies for the peroxisomal enzyme *L-alanine:glyoxylate aminotransferase* (AGT); the **primary hyperoxaluria type 2** (MIM# 260000) is caused by mutations in the gene that encodes for the cytosolic enzyme *glyoxylate reductase/hydroxypyruvate reductase* and finally, the **primary hyperoxaluria type 3** (MIM# 613597) is associated with mutations on the gene that encodes for the mitochondrial enzyme *4-hydroxy-2-oxoglutarate aldolase*. Among the three primary hyperoxalurias, PH1 is the most common and severe.[4]

### 1.1. Primary hyperoxaluria type 1

PH1 represents 80% of the primary hyperoxalurias and can be considered a peroxisomal disorder. Peroxisomes are involved in several vital metabolic functions for the cell. It has been estimated that the human peroxisome contains around 50 different enzymatic activities acting in different metabolic routes, such as the metabolism of reactive oxygen species, fatty acids oxidation or glyoxylate metabolism.[23]

Peroxisomal disorders can be classified into two groups, those linked to peroxisome biogenesis (peroxisome biogenesis disorders) by mutations in the *PEX* genes, and those linked to deficiencies in a peroxisomal enzyme that alters a specific metabolic route.[24] PH1 can be included in the latter group.

**1.1.1. Clinical manifestations of PH1.** Clinical manifestations of PH1 are very heterogeneous and there is not a clear genotype-phenotype correlation. This suggests the participation of modifier genes and/or environmental factors in the onset of the disease.[25–27]

From a biochemical point of view, the characteristic phenotype of PH1 is the elevated synthesis of oxalate. Dysfunctions at the organ level are due to the physical consequences of CaOx deposition. These deposits are thought to cause a damage in the epithelial cells of the kidney and to act synergistically to extend the CaOx deposition.[20]

First symptoms of PH1 can appear during the first few years of life due to the formation of stones in kidney and urinary tract in 82% of the cases.[28] However, the onset of PH1 can range from childhood to adult age. When patients have a proper renal function, the oxalate level is only slightly incremented. Nonetheless, this level will be increased with the progressive reduction of the renal capacity,[20] so the adequate diagnosis of PH1 can be delayed.

The damage caused by the CaOx deposition is translated into different effects according to the affected tissue such as urolithiasis in kidney or myocarditis in heart.

**1.1.2. Diagnosis and epidemiology.** The medical prognosis of PH1 patients is difficult and early diagnosis and treatment are crucial.[4] The lack of definitive criteria for diagnosis has led to late and erroneous diagnosis of PH1.[25] However, in the last years, experts from Europe (OxalEurope) and from the United States (Oxalosis and Hyperoxaluria Foundation) have established recommendations to provide diagnosis and therapeutic treatments for PH1 patients.[5] Although the high level of oxalate is the main indicator of the disease, this is not a definitive indicator so the use of biochemical, enzymologic and genetic analyses are required to complete the diagnosis.[4,5]

The average prevalence rate of PH1 is estimated in 1 to 3 cases per 1,000,000 people-year and its average incidence rate is 1:100,000 born alive-year.[28] Moreover, misdiagnosis may contribute to the underestimation of PH1 prevalence.

**1.1.3. Current treatments.** Traditional treatments of PH1 are focussed on disease symptoms rather than its causes, and the main goal is to keep as much as possible the renal function.[29] When the diagnosis is made in an early stage of the disease, the first strategy is to reduce the amount of oxalate in the body by reducing the intake of dietary oxalate and increasing the fluid intake. The next strategy is to prevent the deposition of CaOx by the use of crystallization inhibitors.[5] The administration of pyridoxine hydrochloride (vitamin $B_6$ precursor) is associated with a reduction of the urinary oxalate for some PH1 patients.[5,30–32]

Dialysis is only recommended in advanced stages of the disease, when the renal function is compromised. Hemodialysis can be used only temporarily during and after organ transplantation,[5,33] in which the recommended strategy is the combined liver-kidney transplantation before the systemic oxalosis stage.[34]

Alternatively to traditional therapies, there are other approaches under investigation such as the degradation of oxalate,[35] the enzyme replacement therapy (ERT), the somatic gene therapy (GT), or the use of chemical and pharmacological chaperones[4] to regulate the AGT protein homeostasis i.e., biological pathways within cells that control biogenesis, folding, trafficking and degradation of AGT.[36]

## 1.2. Glyoxylate metabolism

The **glyoxylate cycle** is characterized by two enzymatic activities: isocitrate lyase, which catalyzes the cleavage of isocitrate to succinate and glyoxylate; and malate synthase, which catalyzes the synthesis of malate from glyoxylate and acetyl-CoA. This cycle allows the synthesis of carbohydrates by bypassing the tricarboxilic acid cycle (TCA) (see Figure 1.2.1). It is generally accepted that the glyoxylate cycle is only available in bacteria, fungi, nematodes and germinating seeds of some higher plants. Mammals lack this cycle, although these two enzymatic activities have been found in mammalian tissues with an unknown role.[20,37]

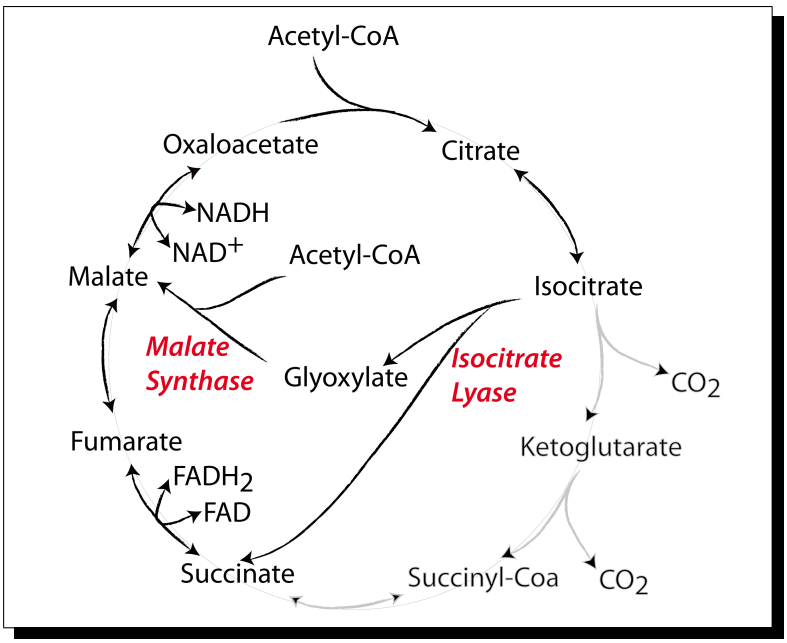


FIGURE 1.2.1. **Overview of the glyoxylate cycle.** The glyoxylate cycle bypasses the decarboxylation steps of the tricarboxylic acid cycle.

Glyoxylate is managed in humans by different metabolic routes termed **the glyoxylate metabolism**, which has a significant role in intermediary liver metabolism.[38] These metabolic routes are found mainly in hepatocytes, where peroxisomes play the role of a glyoxylate sink along with mitochondria[4] in a minor scale. The main molecules involved in glyoxylate metabolism are represented in Figure 1.2.2.

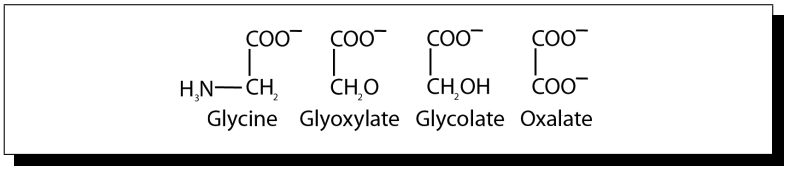


FIGURE 1.2.2. **Main molecules involved in glyoxylate metabolism.**

**1.2.1. Synthesis and transformation of glyoxylate.** In healthy humans, hydroxypyruvate[39] and L-4-hydroxyproline[40,41] are the two main sources of cellular glyoxylate, and peroxisomes and mitochondria respectively, are the main organelles responsible for their metabolism.

In **peroxisomes**, the main glyoxylate precursors are glycine and glycolate (see Figure 1.2.3). The oxidative deamination of glycine to glyoxylate is carried out by the enzyme *D-amino acid oxidase*, while the oxidation of glycolate to glyoxylate is catalyzed by *glycolate oxidase*. Both reactions generate hydrogen peroxide in peroxisomes as a by-product that can be eliminated by *catalase*. Nonetheless, the main contribution to glyoxylate formation is the oxidation of glycolate.[42,43]

Under normal conditions, glyoxylate is metabolized to glycine by the peroxisomal enzyme AGT and/or diffuses to the cytosol where it can be metabolized back to glycolate by the enzyme *glyoxylate reductase/hydroxypyruvate reductase*.[44] The excess of glyoxylate, which is not metabolized to glycine or glycolate, can be oxidized to oxalate in the peroxisome by the enzyme *glycolate oxidase*, or in the cytosol by the enzyme *lactate dehydrogenase*. Glyoxylate and glycolate, can diffuse from peroxisome to cytosol because its membrane is permeable to small hydrophilic solutes through channels made by the peroxisomal membrane protein Pxmp2.[45]

**Mitochondria** produce around 20% of the glyoxylate through the metabolism of L-4-hydroxyproline[41] (see Figure 1.2.3). The glyoxylate produced in mitochondria is metabolized by the enzyme *glyoxylate reductase* and/or by a mitochondrial *alanine:glyoxylate aminotransferase* enzyme that is termed AGT2.[44] AGT2 is encoded in the *AGXT2* gene (chromosome 5 –5q13–), resulting in a class III pyridoxal phosphate enzyme of 514 residues that is suggested to be involved in blood pressure regulation by degradation of methylarginines.[46]

**1.2.2. Oxalate production.** Oxalate is a secondary product that results from glyoxylate metabolism,[38] thus the liver is the main responsible of oxalate

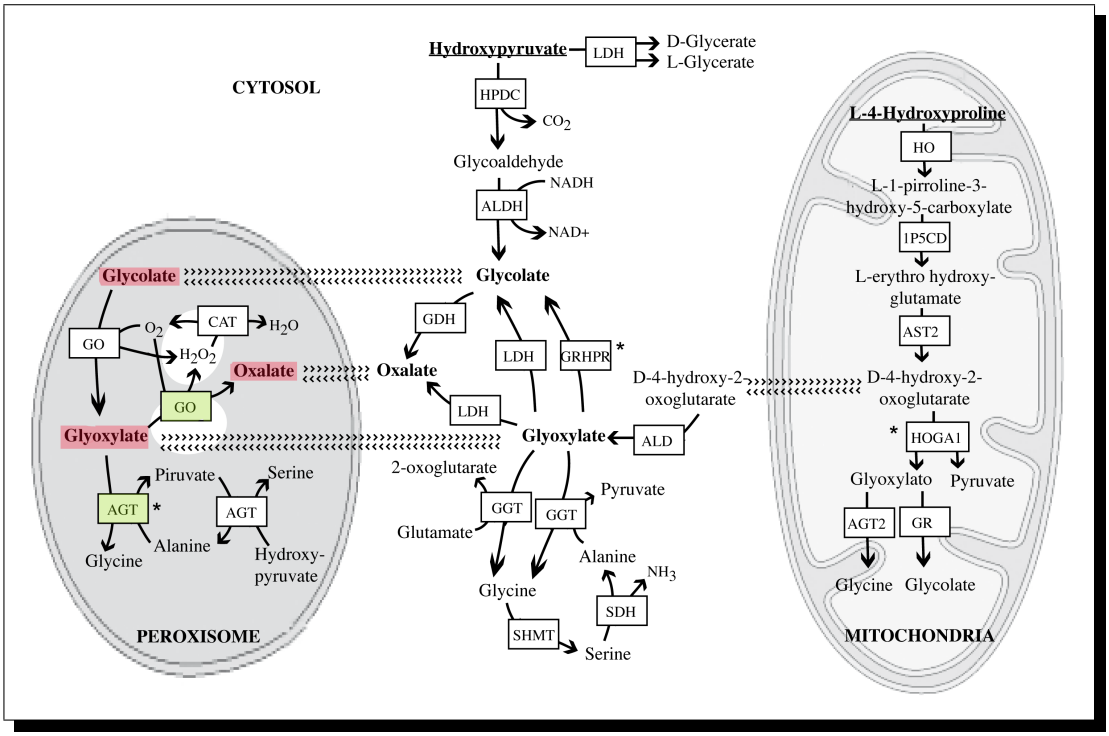


FIGURE 1.2.3. **Main routes of the glyoxylate metabolism in a human hepatocyte.** *AGT, alanine glyoxylate aminotransferase; CAT, catalase; GO, glycolate oxidase; DAO, D-amino acid oxidase/glycine oxidase; GRHPR, D-glicerate dehydrogenase/glyoxilate reductase; LDH, lactate dehydrogenase; SDH, serine dehydratase; SHMT, serine hydroxymethyltransferase; ALDH, aldehyde dehydrogenase; HPDC, hydroxypyruvate dehydrogenase; GDH, glycolate dehydrogenase; HO, hydroxyproline oxidase; 1P5GD, 1 pirroline-5-carboxylate dehydrogenase; AST2, aspartate aminotransferase 2; HOGA1, 4-hydroxy-2-oxoglutarate aldolase; AGT2, alanine:glioxilate aminotransferase 2; GR, glyoxylate reductase; ALD, cytosolic aldolases. *Enzymes involved in primary hyperoxalurias.*

synthesis. Some enzymes involved in the oxalate formation are specific from hepatocytes (AGT or *glycolate oxidase*), but others are expressed in different tissues (*lactate dehydrogenase*) and therefore minor contributions from other tissues cannot be discarded. Only around 10% of the excreted oxalate has an exogenous origin due to the low capacity of the human digestive tract to absorb oxalate.[20] When oxalate synthesis is elevated, oxalate can be transported from the liver to other tissues. Oxalate is not harmful in the liver, and its toxic effects are mainly located in the kidneys where the oxalate is excreted to the urine.

The knowledge of metabolic routes that leads to oxalate formation has been improved in the last decades. However, the aetiology of calcium oxalate is far away to be completely understood due to difficulties in reconstructing these metabolic pathways in hepatocyte models.[20,47]

CHAPTER 2

# Alanine:Glyoxylate Aminotransferase

The human AGT enzyme was first purified and characterized by Thompson and Richardson in 1967, in which the possible role of AGT in glyoxylate and oxalate metabolism was then suggested.[48] The catalytic activity of AGT was found to resemble that of *serine:pyruvate aminotransferase*, but it was proved, years later, that both activities belonged to the same enzyme. In 1986, Danpure and co-workers linked the primary hyperoxaluria type 1 to a deficiency in the AGT enzyme.[1] The human AGT enzyme is a liver-specific peroxisomal aminotransferase (E.C. 2.6.1.44) that has the biological role of glyoxylate detoxification.

## 2.1. Structure and function

From the point of view of PLP enzymes, the **structure** of AGT is classified as a Fold Type I (aspartate aminotransferase family). Its quaternary structure has two independent active sites that lie on the dimer interface.[49] The catalytically active form of AGT is a homodimer of 86 kDa with two molecules of coenzyme bound, constituting the holoenzyme*. As all of known aminotransferases, the AGT enzyme needs the coenzyme pyridoxal-5'-phosphate (PLP) (see Figure 2.1.1) to perform its function.

Each monomer of AGT is constituted by 392 residues that according to its crystal structure can be divided into three domains[50] (see Figure 2.1.2). The **N-terminal domain** (first 20 residues) is an extension that barely has secondary structure and is placed over the surface of the other monomer; the **central domain** (next 260 residues) has most of the residues that form the active site and the dimerization interface; and the **C-terminal domain** (last 110 residues) contains, among other functions, the peroxisome targeting sequence.[51]

*In this document, the term «holoenzyme» will designate the form of the enzyme where the coenzyme is as pyridoxal-5'-phosphate. Otherwise, it will be explicitly indicated.

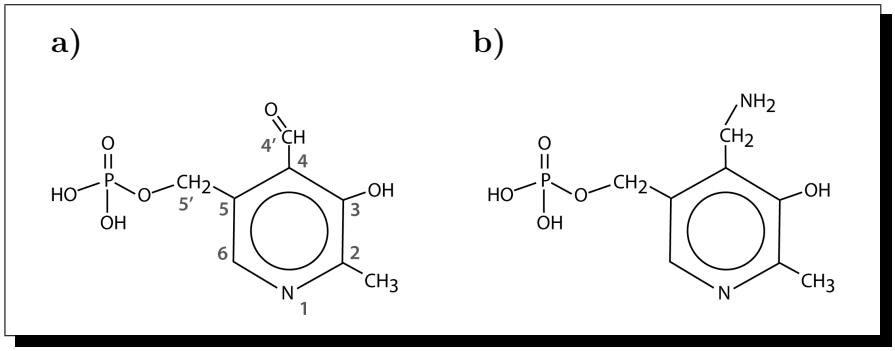


FIGURE 2.1.1. **Structure of active forms of the $B_6$ vitamin. a)** pyridoxal-5'-phosphate (PLP) and **b)** pyridoxamine-5'-phosphate (PMP).

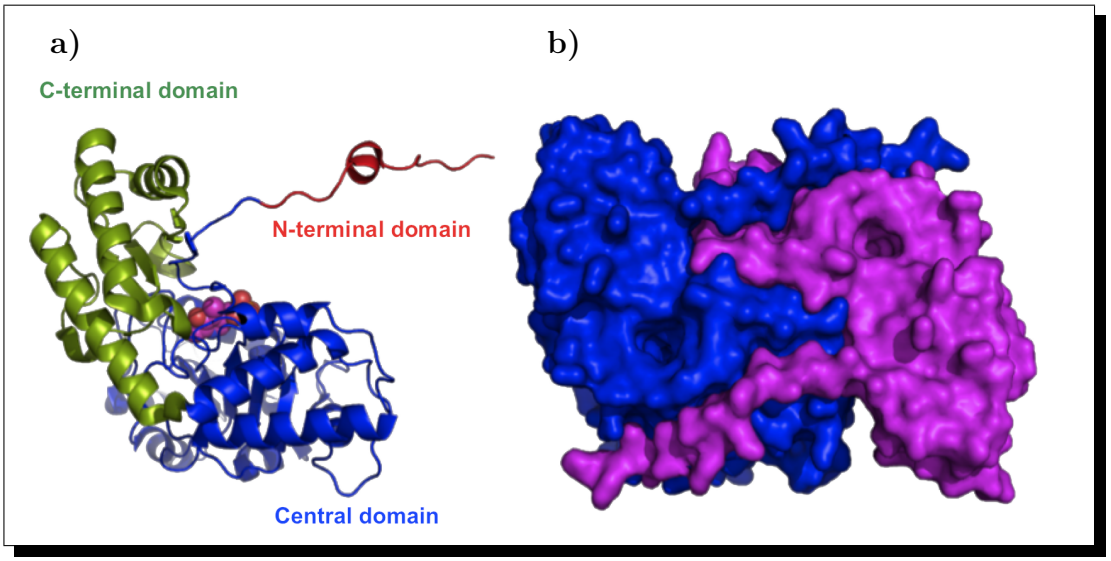


FIGURE 2.1.2. **Structure of the AGT enzyme obtained by X-ray diffraction (PDB ID 1h0c).[50] a)** secondary structure of the AGT monomer where PLP is represented by spheres. **b)** surface representation of the AGT dimer. *Images were created with Pymol Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.*

From the point of view of the **function**, AGT catalyzes the transamination of glyoxylate to glycine using L-alanine as amino donor, resulting into pyruvate. The transamination mechanism was established through works performed by Snell and Braunstein and it has several steps that are schematized in the Figure 2.1.3. The specificity of the reaction is controlled by steroelectronic effects (hypothesis of Dunathan) where the bond to be broken is situated parallel to the conjugated $p$ orbitals of the $\pi$ system constituted by the Schiff base and pyridine ring.[52]

In human AGT, the PLP molecule is bound to the active site through an internal Schiff base between its C-4' carbon and the $\varepsilon$-amino group of the Lysine 209 (see Figure 2.1.3-a). The addition of the amino acid substrate results in a transaldimination, converting the internal Schiff base into an external Schiff base with the $\alpha$-amino group of the substrate (see Figure 2.1.3-b). In the next steps, the $\alpha$-hydrogen of the substrate is transferred to the C-4' of the PLP through a quinonoid (see Figure 2.1.3-c), to form a ketimine intermediate (see Figure 2.1.3-d). This intermediate is then hydrolyzed to PMP and the keto acid product (see Figure 2.1.3-e). This process is completed by the other half reaction in which an amino acid is formed from a keto acid substrate, including the regeneration of PLP.[53]

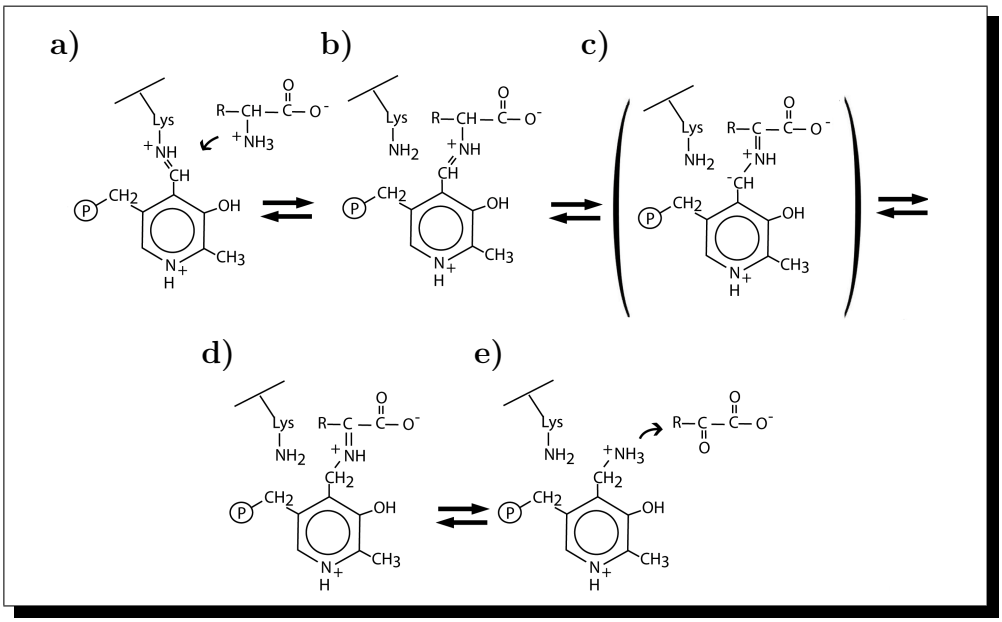


FIGURE 2.1.3. **Mechanism of the aminotransferase reaction.** The scheme ilustrates the half reaction of the transamination. The other half is done by a reverse process.

The **post-translational translocation to the peroxisome** of AGT is mediated by a peroxisome targeting sequence type 1 (PTS1), which is located in the C-terminal domain. This translocation is carried out by the interaction with the peroxisomal receptor Pex5p.[54,55]

The C-terminal tripeptide of human AGT is KKL, where two of them match the SKL consensus PTS1 tripeptide.[56] The tripeptide KKL has not been detected in other peroxisomal proteins following the PTS1 route and previous works have proved that this sequence is insufficient to direct the peroxisomal

targeting.[54] Moreover, the crystal structure of the AGT-Pex5p complex (see Figure 2.1.4) reveals that other residues (extended PTS1: residues 381-388 and 327-330) are involved in the AGT-Pex5p interactions, which are crucial for the recognition and maintenance of the complex.[57]

The AGT protein is expressed at high levels in hepatocytes and it has low affinity for the PTS1-binding domain of human Pex5p protein.[58] It has been suggested that the relation between the affinity of the PTS1 for the PTS1-binding domain of human Pex5p and the expression levels of peroxisomal proteins imported by this route, allows the existence of a relatively uniform population of the Pex5p-protein complexes to import these proteins at an adequate level to develop their function.[58]

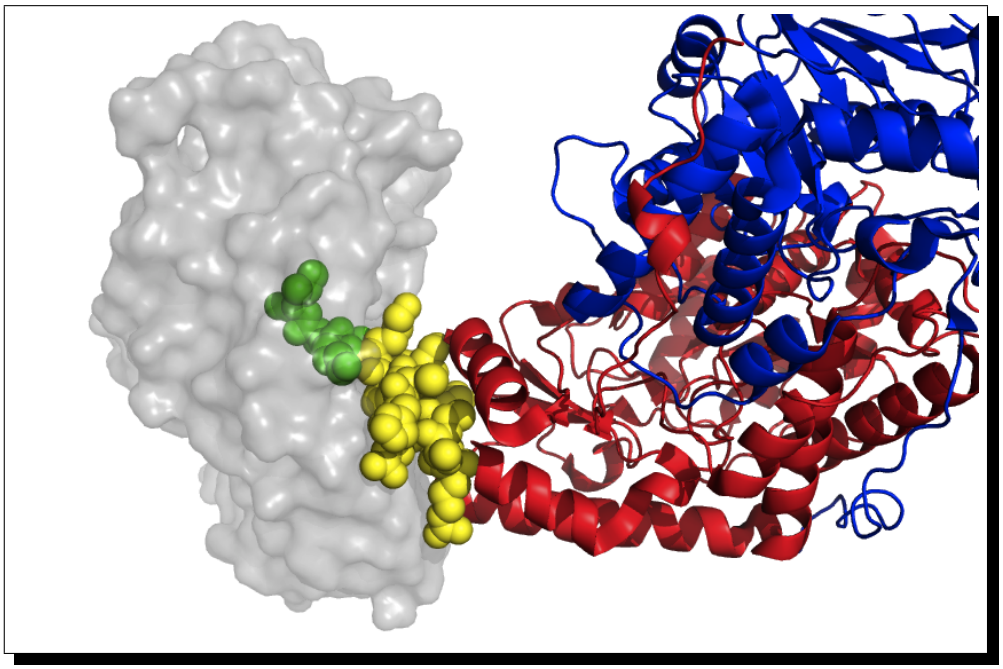

FIGURE 2.1.4. **Structure of the complex AGT–PTS1-binding domain of the human peroxisomal receptor Pex5p obtained by X-ray diffraction (PDB ID 3R9A)**. The PTS1-binding domain (grey) interacts with C-terminal PTS1 tripeptide (green) and residues from the extended PTS1 (yellow). *Image was created with Pymol Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.*

## 2.2. Alleles and pathogenic variants

The AGT enzyme is encoded by the ***AGXT* gene**.[1] This gene has 11 exons comprising around 10 kb in the chromosome 2 (2q37.3). There are two

main alleles for the *AGXT* gene in non-PH1 subjects. The most common allele or haplotype is termed the *major* allele (*AGXT-Ma*) and the other one is termed the *minor* allele (*AGXT-Mi*). The main differences at protein level are the substitution of a leucine to a proline at the $11^{th}$ position (P11L) and the substitution of a methionine to a isoleucine at the $340^{th}$ position (I340M) of the polypeptide chain†.[6] A comparison of the protein properties for both alleles is shown in the Table 2.2.1.

TABLE 2.2.1. **Main properties between the protein product of the *major* and *minor* allele**.

| *Property* | *AGT − Ma* | *AGT − Mi* |
|---|---|---|
| Frequence [1] | 80% | 20% |
| Substitutions | *N/A* | P11L; I340M |
| Localization | Peroxisomal | Peroxisomal [2] |
| Stability | Reference value | Slighly lower |
| Activity | Reference value | Similar or Slighly lower |

*N/*A, not applicable.
[1] In healthy individuals.
[2] In the present work, although it has been reported some degree of mistargeting to mitochondria in previous works.[3]

More than **150 pathogenic mutations** over the *AGXT* gene are associated with PH1.[6] 50% of these mutations are missense mutations in which a single nucleotide change results in a codon for another amino acid. Around 13% are nonsense mutations that result in a premature stop codon while the 12% affect the mRNA splicing. Finally, the last 25% of these mutations correspond to deletions and insertions of different size.[6]

There are different molecular mechanisms underlying the protein defect in PH1 (see Table 2.2.2). These molecular mechanisms can involve: i) mitochondrial mistargeting, at which the AGT enzyme is imported to the mitochondria;[51] ii) protein aggregation;[59–61] iii) accelerated degradation;[60] and/or iv) catalytic defects.[62] Nonetheless, the molecular details of the main mechanisms for most common mutants remain unclear. Most PH1 mutations co-segregate with the

†The polypeptide chain encoded by the *major* allele is the wild type form of the human AGT and is termed as AGT-Ma. On the other hand, the protein product of the *minor* allele is termed as AGT-Mi.

*minor* allele and the existence of synergic effects with the P11L substitution has been proposed.[63]

TABLE 2.2.2. **Most frequent PH1 mutations on human AGT enzyme**

| *Mutation* | *Allele background* | *Frequency* [1] | *Molecular phenotype* |
|---|---|---|---|
| G170R | Minor | 30% | Mistargeting to mitochondria |
| I244T | Minor | 9% [2] | Aggregation |
| F152I | Minor | 1-7% | Mistargeting to mitochondria |
| G41R | Minor/Major | 1% | Aggregation |
| G82E | Major | <1% | Catalytic defect |

[1] Frequency over characterized pathogenic alleles.
[2] The frequency of this mutation reaches the 90% in the Canary Islands.

CHAPTER 3

# Protein misfolding and human diseases

Proteins are the molecular machinery responsible for the most essential functions of life. Although information to reach the native state is encoded in their primary structure (amino acid sequence),[64] it has been highlighted in the last decades that many newly synthesized proteins require a complex cellular machinery to reach their native state *in vivo*.[65] This complex machinery constitutes an essential quality control and its adequate balance is vital for the survival of the cell.

## 3.1. Protein folding *in vitro*

In the energy landscape approach, protein folding is described as a process taking place on a rugged free-energy surface, which is a function of the conformational properties of the unfolded and folded states of the protein. This surface is energetically channelled, and protein folding is a consequence of the free energy reduction as native interactions are established (see Figure 3.1.1).

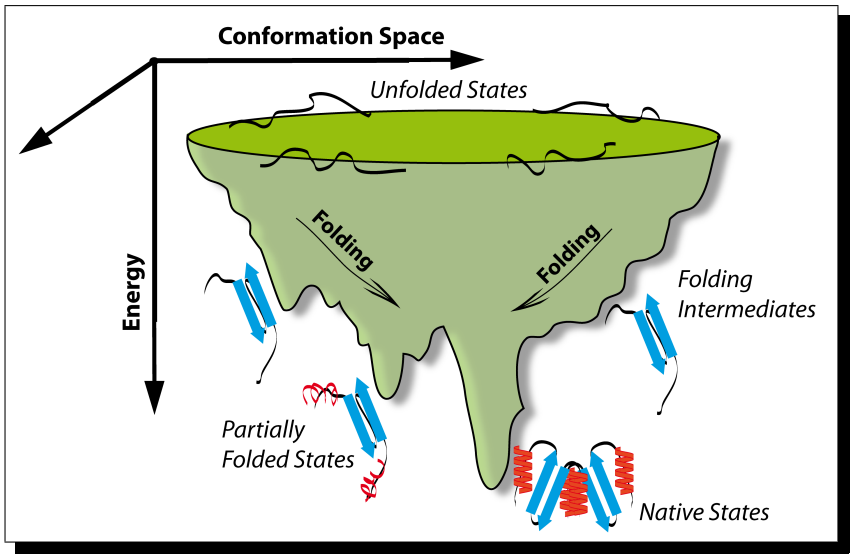


FIGURE 3.1.1. **Energy landscape view of the protein folding.** Multiple conformations are funneled to the native state. Aggregation can occur from partially folded states that can be protected by molecular chaperones.

In this energy landscape, it is not necessary to postulate the existence of a unique folding pathway (that emerged from the Levinthal's paradox)[66] because the protein can fold following different routes, in which partially folded states can be populated. Kinetic traps can be viewed as local energy minima due to roughness on this energy landscape, where the protein remains in a partially folded state. The roughness of the energy landscape may be modified by mutations that stabilizes kinetic traps. Moreover, the energy landscape of a protein can also be modulated *in vivo* by environmental stresses and the action of the cellular protein homeostasis network that maintains the health of the proteome. Consequently, the knowledge obtained for protein folding *in vitro* is not enough to be properly translated to the process of the protein folding in a cell.

### 3.2. Protein folding *in vivo*

The protein folding *in vivo* of nascent chains is linked to ribosomes and largely influenced by the environmental conditions of the cell. The rate of protein folding *in vivo* is affected by the rate of mRNA translation.[67] Protein synthesis is relatively slow (5-9 residues/sec in eukaryotes) thus the polypeptide chain will be exposed in partially folded states (prone to aggregate) during the synthesis. In addition, the correct folding of a protein can also be compromised by transcription or translation errors, the partial synthesis of the protein or the damage made by reactive oxygen species.[68] When a protein is not able to fold properly, it can be trapped into partially folded or misfolded states. In these states, hydrophobic regions of the protein (that would be buried in the native state) are exposed and thus the possibility to suffer aggregation is raised.[69] Moreover, aggregation events are promoted in the crowded environment of the cell where some proteins also need the binding of specific cofactors or the acquisition of functional quaternary structures to properly fold.

Nascent polypeptide chains sometimes require the action of a complex cellular machinery to reach the native state efficiently.[70] This complex cellular machinery is composed by molecular chaperones and chaperonins, which some of them are energy-dependent.[71] In the same way, as a regulatory system, the life span of intracellular proteins is controlled by degradation pathways, e.g. the proteasome or cellular system of proteolysis.[72,73] Therefore, when a protein is damaged or the biological function is not needed, the protein is removed.

However, the existence of these quality control systems does not guarantee the success of protein folding.

### 3.3. Folding and quality-control mechanisms

In the cell, proteins are constantly degraded into amino acids and replaced by newly synthesized polypeptide chains (protein turnover).[74] The protein turnover is a highly selective process. The turnover rate correlates with protein function and subcellular location and it is controlled by the protein homeostasis (proteostasis) network, which is the central pathway responsible for the generation and maintenance of the proteome in a healthy cell.[75]

Molecular chaperones have a central role in the proteostasis network[65] by refolding misfolded proteins, promoting degradation of misfolded polypeptides by ubiquitin-proteasome system (UPS) pathway[76] or sequestering misfolded proteins into specialized quality control compartments.[77] In the cytosol of eukaryotic cells, there are two distinctly regulated chaperone networks: i) chaperones linked to protein synthesis (CLIPS) that are functionally and physically linked to the translation machinery and assist the folding of newly translated proteins, and ii) the heat shock proteins (HSP), which can be induced by heat shock factors and protect the proteome from different stresses.[78] In addition, certain chaperones have the ability to extract and refold proteins from aggregates such as the small heat shock proteins and chaperones from AAA-ATPases family.[79] Furthermore, when the folding and proteolysis systems are overloaded, the aggresome-autophagy pathway could be another cellular defense system. This system collects and compartmentalizes misfolded or aggregated proteins into specialized inclusions called aggresomes, which are then cleared by autophagy.[80]

In the cytosol, soluble misfolded proteins are cleared through the UPS upon polyubiquitination, which tags them for destruction by the 26S proteasome while the majority of misfolded endoplasmic reticulum (ER) proteins are cleared by the ER-associated degradation (ERAD) pathway.[8,81] The ability of chaperones to recognize misfolded polypeptides is central to protein quality control. In general, chaperones recognition is mediated by hydrophobic interactions[10] while degradation is often controlled by the recognition of an amino-terminal residue according to the N-end rule.[82]

A number of pathological conditions, environmental and metabolic stresses increase the production of misfolded proteins. Accordingly, some transcriptional programs sense stress and enhance the expression of molecular chaperones and degradation components. However, the protein homeostasis capacity is declined through aging[83,84] and can be overwhelmed by the accumulation of misfolded proteins. The contribution of these stresses to burden the cellular quality control machinery is often unknown and they may play an important role in the onset of misfolding diseases.

### 3.4. Misfolding diseases

Conformational or misfolding diseases define a group of disorders in which protein folding fails due to mutations.[85] However, in a misfolding disease at least some of the target protein may be correctly folded by the proteostasis network of the cell, and therefore proteostasis capacity would influence the onset of the disorder.[86]

The proteostasis network is composed by several biological pathways, comprising chaperones, folding enzymes and trafficking and degradation components. This network is controlled by signalling pathways that modify the concentration, distribution and activities of key components. Basically, the function of this network is to re-sculpt the energy landscape of a protein folding by reducing aggregation, promoting folding and assisting protein degradation.[86] Therefore, the proteostasis network capacity influences the relation between the protein folding energetics and the biological folding process.

A minimal proteostasis boundary is defined by combining protein folding energetics and the proteostasis network capacity to determine whether a protein reaches adequate levels of folding for biological function.[36] The location of this boundary depends on the proteostasis capacity and it is defined in a three-dimensional space by protein folding thermodynamics (stability), folding and misfolding kinetics (see Figure 3.4.1).

In a cell, those proteins with energetics below the proteostasis boundary will be properly folded at a given proteostasis network capacity (green nodes in Figure 3.4.1). Alternatively, less stable, slowly folding or rapidly misfolding proteins are closer to the boundary and their ability to fold will be more sensitive to changes in the proteostasis network capacity. Furthermore, folding energetics can be compromised by the action of stresses or mutations and the protein would

move beyond the proteostasis boundary leading to degradation, aggregation and disease (red nodes in Figure 3.4.1).[36]

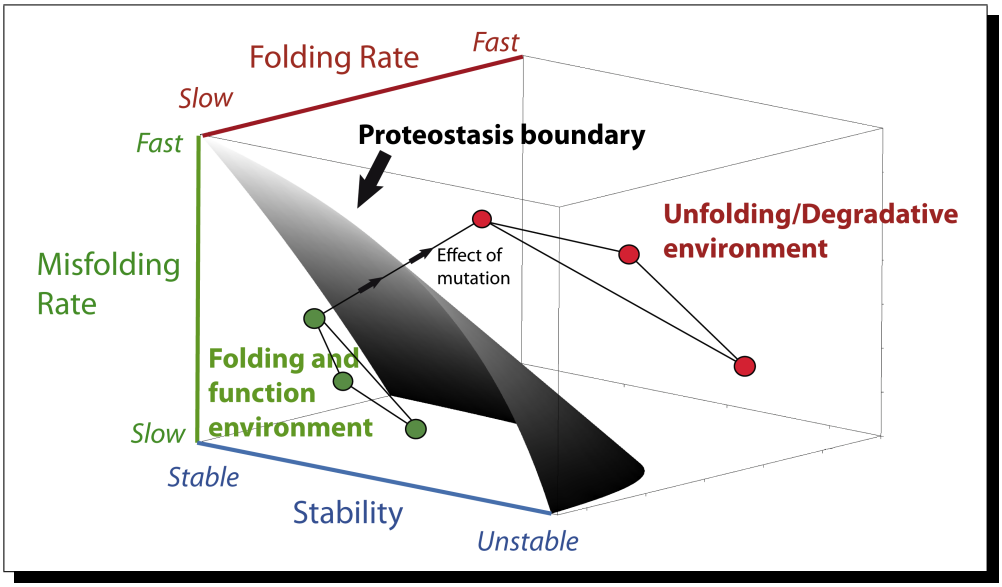


FIGURE 3.4.1. **Illustration of the proteostasis boundary.** The proteostasis boundary is defined by kinetics of folding, misfolding and thermodynamic stability. The proteostasis network copes with protein folding constraints in a healthy cell (green nodes). When a mutation is presented it can result in misfolding leading to loss-of-function or gain-of-toxic-function disease (red nodes).

In summary, a link between folding diseases and the proteostasis network is proposed. When a cell is not able to handle its proteostasis load (due to stress, aging or chronic expression of misfolded proteins) it could lead to a loss-of-function disorder (such as cystic fibrosis[87] or lysosomal storage disorders[88]) or a gain-of-toxic-function disorder (such as neurodegenerative diseases[86,89]). In theory, any misfolding disease (protein folding energetics outside the proteostasis boundary) could be corrected by i) expanding the proteostasis boundary or ii) reducing the constraints to fall below the proteostasis network boundary by stabilizing the native state or reducing the misfolding rate. Pharmacological chaperones and proteostasis regulators could be used for modulation of the proteostasis network. This modulation is becoming a promising therapeutic approach for the design of treatments for misfolding diseases.[13]

## 3.5. Misfolding diseases in model organisms

The knowledge of the interplay between folding energetics and proteostasis network capacity (position and shape of proteostasis boundary) could be used

for the design of new therapeutic approaches for misfolding diseases. The *in vitro* characterization of disease-causing mutants should be complemented by an *in vivo* characterization based on animal model systems of the disease, in which the restoration of the proteostasis could be tested.[90]

Due to its versatility and intrinsic properties as an animal model, the nematode *C. elegans* has become a preferred system for modelling human conformational diseases:[12,89,91–95]

- *C. elegans* is a simple organism that shares essential biological features with humans.[96]
- The nematode has a small size with a short life cycle and lifespan.[97]
- Mutant strains become homozygous by self-fertilization and can be easily grown in the laboratory obtaining a high statistical population.[96]
- The animal body is transparent thus cells and proteins can be tracked *in vivo*.[98,99]
- The genome of the nematode is completely sequenced[100] and it is amenable to forward and reverse genetic screens,[101] particularly to gene inactivation by RNAi.[102]

In regard to PH1, a predicted *open reading frame* (ORF), denoted as *T14D7.1*, orthologous to the human *AGXT* gene was found[103] in this nematode. *C. elegans* could be an appropriate model to study the proteostasis of human AGT enzyme, the relation of disease-causing mutants with different elements of the proteostasis network[104,105] and to identify protein homeostasis components to be pharmacologically targeted.

CHAPTER 4

# Engineering protein stability

The stability of a folded protein must be enough to keep its active conformation during a biologically relevant time-scale in order to satisfy the functional demands of the cell. In the previous chapter, we have seen that there is a relation between the protein folding energetics and the proteostasis network capacity, in which properties such as protein stability, folding kinetics or misfolding kinetics define the position in reference to the proteostasis boundary. Therefore, for a given protein, the increase of protein stability by protein engineering could be a useful tool to improve protein efficiency *in vivo* and could be used in GT or ERT strategies[16] to treat monogenic inborn errors of metabolism such as PH1.[15]

## 4.1. Thermodynamic *vs.* kinetic stability

It is thought that the main forces that stabilize proteins structure are the hydrophobic effect and the hydrogen bonds[106] while the conformational entropy is the main destabilizing force. Nonetheless, the thermodynamic stability of a protein is marginal ($\sim$ 5-10 $kcal/mol$) due to an enthalpy-entropy compensation.

The protein **thermodynamic stability**, in the simplest two-state model ($N \Leftrightarrow U$), is defined by the population of the native and unfolded states in equilibrium conditions at a given temperature (the folding equilibrium constant, $K_T$)[107]. Usually, the thermodynamic stability is expressed in terms of the partial molar Gibbs energy difference ($\Delta G$) between the unfolded (U) and the native state (N) (see Figure 4.1.1-a):

$$\Delta_N^U G\,(T) = \Delta H(T) - T \cdot \Delta S(T) = -R \cdot T \cdot ln\,(K_T)$$

$$K_T = \frac{[U]}{[N]}$$

The thermodynamic stability of a protein, as a function of temperature, is fully specified with the definition of three thermodynamic parameters: enthalpy and entropy changes (at the reference temperature) and the heat capacity change. These three parameters can be accurately measured by differential scanning calorimetry (DSC). The profile of the unfolding Gibbs energy *vs.* the temperature, defines the stability curve of a protein (see Figure 4.1.1-b). The stability curve has one maximum and there is a range of temperatures where the protein is stable ($\Delta G > 0$). In addition, there are two temperatures ($T_{m'}$ and $T_m$) where the stability curve crosses the abscissa ($\Delta G = 0$) constituting the high and low-temperature melting points, respectively (see Figure 4.1.1-b).

However, a complete thermodynamic description of the *in vitro* protein unfolding is not always possible due to the existence of irreversible events that prevent the establishment of the equilibrium. A thermodynamically stable protein will mainly populate the native state, but if the unfolded state is undergoing an irreversible alteration (aggregation or proteolysis) in a relevant timescale, the equilibrium will be shifted to the unfolded state, thus the thermodynamic stability does not guarantee the native state during a given *in vivo* timescale. Alternatively, the **kinetic stability** of a protein, which determines the rate of protein unfolding according to the height of the free energy barrier between native and unfolded states ($\Delta G^{\neq}$), is a mechanism that could keep the native state of a protein in a biologically relevant time (see Figure 4.1.1-a).[19]

According to the transition state theory,[108] the rate constant ($k$) of an irreversible denaturation process can be described as:

$$k = k_0 \cdot exp\left(-\frac{\Delta G^{\neq}}{R \cdot T}\right)$$

where the $k_0$ is the pre-exponential factor and $\Delta G^{\neq}$ is the free activation energy barrier between the native and the transition state.

Evolution may have selected marginal thermodynamic stability of proteins in which this positive, but not too large unfolding Gibbs energy, may have some physiological advantages[109] such as i) to avoid proteins trapped in misfolded structures and/or ii) to allow transient unfolding states during transportation or function. Therefore, it is proposed that many proteins evolved to have a high kinetic stability (low denaturation rate) in order to keep the safety of the native state.[110] However, mutations on amino acid sequence may be translated into a modification of the kinetic stability leading to misfolding diseases.[111]

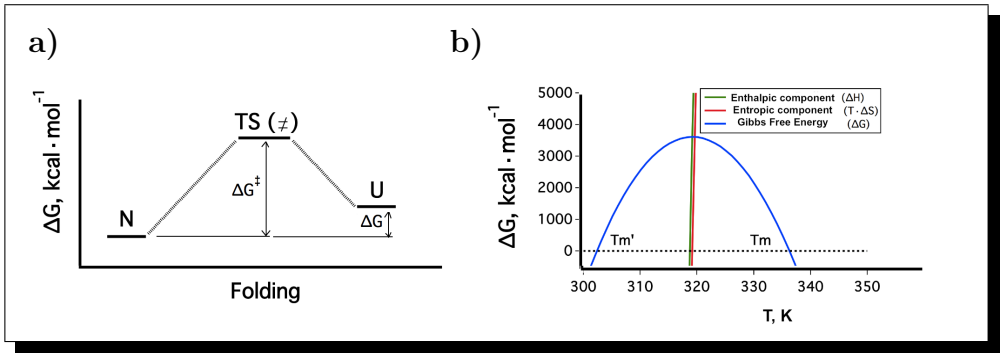


FIGURE 4.1.1. **Thermodynamic *vs* kinetic stability. a)** representation of the thermodynamic stability ($\Delta G$) of a protein as the unfolding Gibbs energy between the unfolded (U) and native (N) state and the kinetic stability as the free-energy barrier ($\Delta G^{\neq}$) between the native (N) and transition state (TS); **b)** representation of a protein stability curve.

## 4.2. Protein engineering

The aim of protein engineering is the modification of the relevant properties of an enzyme, e.g. stability and activity. These properties should be modified in order to be adapted to certain conditions imposed by a technological process or for improving treatments of human diseases.[112] Protein engineering has also become a tool that generates a feedback for structure-function relationships.[113]

Local and long-range interactions are involved in the stability of a protein. Stabilizing mutations can increase the thermodynamic stability (increasing the free energy difference between unfolded and native states) or the kinetic stability (increasing the activation free energy between the transition and native states). There are many strategies that can be used to stabilize proteins. Rational approaches use the available structural information to correlate a given property with patterns of the structure. These approaches imply the knowledge of the tridimensional structure of the target protein. Alternatively, directed evolution implies the acceleration of timescale by application of a selective pressure to a large set of variants that allows the modulation of a selected property.

The **consensus concept**[114] for protein engineering, can be classified as a semi-rational approach because it does not need 3D-structural information and it is only based in the amino acid sequence (primary structure) of proteins.

According to the neutral theory of molecular evolution, in a protein those portions that are functionally less important evolve faster than the more important ones, considering a constant rate of amino acid or nucleotide substitution through evolution.[115] Thus, from a sequence alignment of homologous proteins,

the most probable amino acid at a given position (consensus amino acid) contributes more than average to protein stability than the non-consensus amino acid.[14,116] Moreover, specific activity and thermostability could be, at least, partially independent properties (incomplete trade-off).[117] Thus, by using the consensus approach, the protein thermostability is expected to be improved while the specific activity is maintained.[118,119] As a result, the consensus concept seems to be a simple, effective and inexpensive approach for protein engineering in order to design an improved AGT protein for treatment of PH1.

# Part 2

# Objectives

## Objectives

The present thesis belongs to a multidisciplinary project focussed on primary hyperoxaluria type 1 as a conformational disease. The main goal of this project is to explore new potential therapeutic approaches to treat PH1.

To achieve this long-term objective, different approaches have been considered in order to get knowledge about molecular mechanisms underlying this disease and the results are partially collected in this thesis. Consequently, the present thesis is composed of different chapters with the following individual objectives:

(1) Improve our knowledge on the mechanisms underlying PH1 disease-causing mutations by analyzing the complex mutational effects on the AGT folding and stability *in vitro* and in cells.

(2) Study whether the nematode *C. elegans* may be used as a model for PH1. As a first approach, a predicted orthologous protein of the human AGT in the nematode has been characterized.

(3) Improve the efficiency of the gene and enzyme replacement therapy for PH1 by enhancing the human AGT protein using the consensus-based approach.

(4) Obtain information about the biophysical basis of AGT stability *in vitro vs.* in cells by characterizing the thermal and chemical denaturation of a set of AGT proteins composed by non-disease polymorphisms, disease-causing and consensus-based variants.

# Part 3

# Results and Discussion

CHAPTER 5

# *In vitro* characterization of PH1 pathogenic variants

## 5.1. Brief introduction

The knowledge of PH1 disease-causing mechanisms is a powerful tool for designing new therapeutic approaches. In order to obtain more information about the effects of PH1 mutations on AGT folding, activity and stability, we have performed a biochemical and biophysical characterization of most common non-pathogenic and disease-causing AGT variants.

This set is composed by 4 polymorphic variants (the *major* allele or AGT-Ma, the *minor* allele or AGT-Mi, and the two point mutations that form the *minor* allele: P11L-Ma and I340M-Ma), 6 known PH1 mutations that are found in the *minor* allele background (H83R-Mi, F152I-Mi, G170R-Mi, I244T-Mi, P319L-Mi and A368T-Mi) and 2 variants with uncertain pathogenicity that are also found in the minor allele background (R197Q-Mi and A295T-Mi).

In this chapter the *in vitro* characterization is presented and discussed.

## 5.2. Results.

**Analyzed mutations are spread over the three dimensional structure of AGT.** Analyzed mutations are located in different elements of the protein structure in which native residues are mainly buried in the folded state (see Figure 5.2.1). Mutated residues are shown in the crystal structure of AGT-Ma in the Figure 5.2.2-a.

The **P11L** and **I340M** polymorphisms are found in the dimer interface. The sterically constrained Pro11 is changed by an aliphatic residue while the I340M substitution keeps the hydrophobic nature of the residue. The rest of the mutations are located far from the dimerization interface, and only the **H83R** mutation is found in the active site. The His83 makes hydrogen-bonding interaction with an oxygen atom of the phosphate group of PLP (see Figure 5.2.2-b). Therefore, the substitution by an arginine should modify PLP binding due to steric and electrostatic effects.

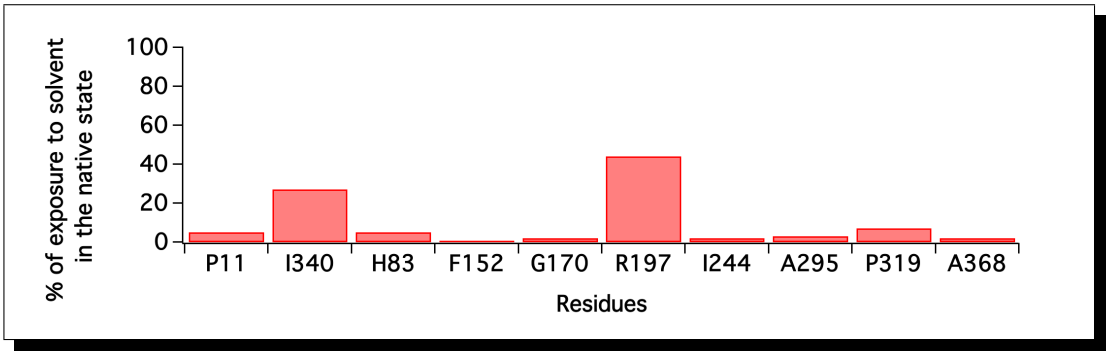


FIGURE 5.2.1. **Percentage of accesible surface area (ASA) in the native state.** Data obtained from the crystal structure of AGT-Ma[50] using a home-built software kindly provided by Prof. Jose Manuel Sanchez-Ruiz.

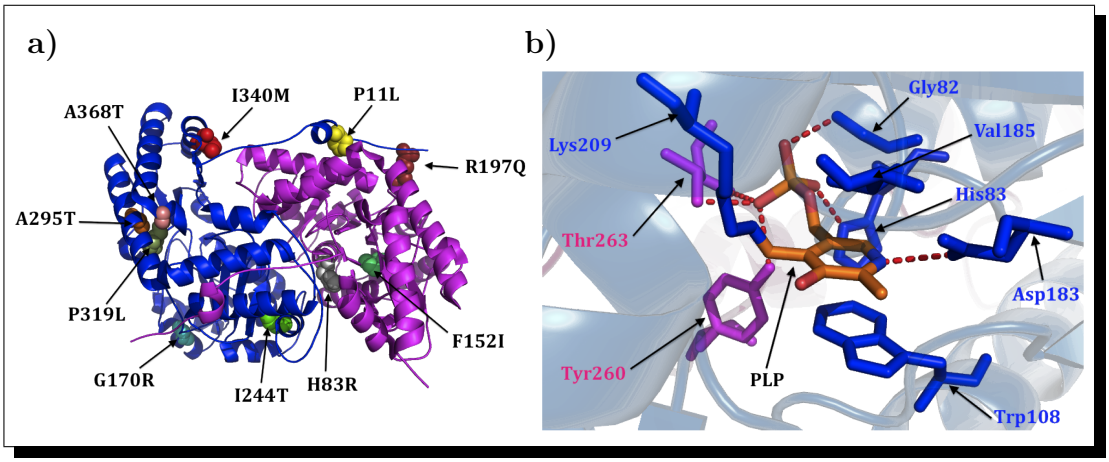


FIGURE 5.2.2. **Mutations and polymorphisms on AGT structure (PDB ID 1h0c). a)** AGT dimer structure where PH1 mutations and polymorphisms are labeled and pointed by arrows; **b)** representation of the active site of AGT where key residues are labelled and important hydrogen bonds are represented by dotted lines. Residues from each monomer are marked as blue or purple. *Images were created with Pymol Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.*

The **A295T**, **P319L** and **A368T** mutations are found in the vicinity of a domain:domain interface within the AGT monomer. The buried Pro319, located in a turn, is changed to an aliphatic leucine while the A295T and A368T mutations change a hydrophobic alanine to a polar uncharged threonine. The **G170R** and **I244T** mutations imply the change of a buried aliphatic residue to a positively charged residue with a longer side chain and a polar uncharged residue, respectively. The **F152I** substitution keeps the aliphatic nature while removes an aromatic residue. Finally, the **R197Q** mutation is located in a turn

and it involves the change of a large positively charged to a polar uncharged residue.

**PH1 mutations reduce the purification yield.** The polymorphisms and disease-causing variants have been expressed and purified from BL21 strains of *E. coli* following the same procedure. In all cases, a soluble and highly pure and homogeneous protein was obtained. Nonetheless, the average yield of purification was very different among the analyzed variants, ranging from 3 to 15 mg of protein per liter of cell culture, in which the lower average yields correspond to PH1 mutations (see Table 5.2.1).

TABLE 5.2.1. **Purification yields and retention volumes for AGT variants.**

| *Variants* | *Purification yield* [1] $mg \cdot l^{-1}$ *cell culture* | *Retention volume* [2] *ml* |
|---|---|---|
| AGT-Ma | 16.08 (100%) | 82.7 |
| AGT-Mi | 15.38 (96%) | 82.8 |
| P11L-Ma | 3.76 (23%) | 83.2 |
| I340M-Ma | 25.97 (162%) | - |
| H83R-Mi | 6.97 (43%) | 81.3 |
| F152I-Mi | 5.52 (34%) | - |
| G170R-Mi | 6.48 (40%) | 83.2 |
| R197Q-Mi | 3.91 (24%) | 83.0 |
| I244T-Mi | 3.79 (24%) | 83.2 |
| A295T-Mi | 13.01 (81%) | 82.6 |
| P319L-Mi | 4.83 (30%) | - |
| A368T-Mi | 5.59 (35%) | 82.8 |

[1] Purification yield percentage over AGT-Ma is shown between parentheses
[2] Using a HiLoad$^{TM}$ 16/60 Superdex$^{TM}$ 200 SEC column

All samples have been subjected to size exclusion chromatography (SEC) during the purification process. Chromatograms show a single peak with an average retention volume of 82.6 ml. This retention volume corresponds to a molecular weight of around 90 kDa, consistent with the AGT dimer (see Figure 5.2.3).

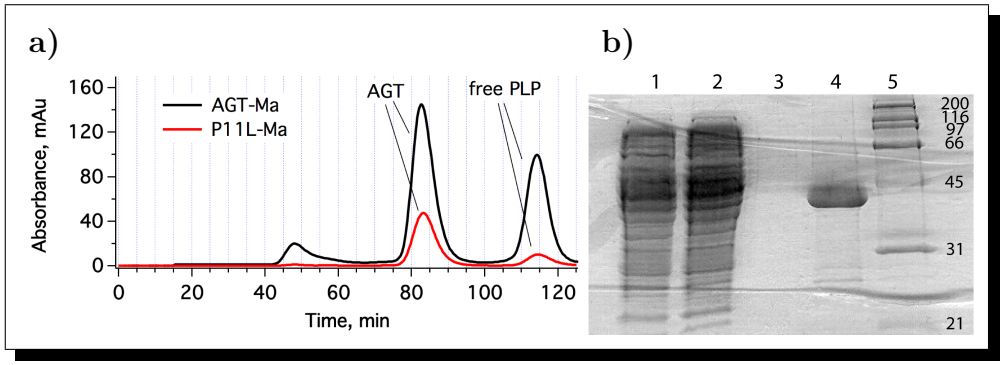


FIGURE 5.2.3. **Purification of AGT variants. a)** Typical size exclusion chromatograms for AGT. Data corresponds to AGT-Ma (black) and P11L-Ma (red) variants using a HiLoad$^{TM}$ 16/60 Superdex$^{TM}$ 200 SEC column; **b)** SDS-PAGE of AGT-Ma samples during purification: 1) soluble extract, 2) flow-through after IMAC, 3) wash sample and 4) elution sample; 5) molecular weight marker.

**The oligomerization state of the AGT is not altered by PH1 mutations.** The oligomerization state has been further evaluated for these variants. Dynamic light scattering (DLS) studies were performed to obtain more information about their quaternary structure (see Table 5.2.2). The diameters obtained for the two polymorphisms, AGT-Ma and AGT-Mi, are 8.1 ± 2.3 nm and 7.8 ± 0.1 nm respectively. The average diameter for the pathogenic variants is 8.4 ± 1.3 nm. These results are consistent with SEC chromatograms (see Table 5.2.1), suggesting that there is no alteration of the oligomerization state of these variants. Considering the uncertainty associated with these measurements, the obtained hydrodynamic diameter of AGT-Ma is in agreement with the value that has been previously reported.[120]

**Most PH1 mutations have little impact on catalysis.** The enzyme kinetics of these variants have been studied using a double displacement mechanism (also known as a ping-pong mechanism).[62] Using this model, the kinetic parameters are obtained from a global fit of data at different alanine and glyoxylate concentrations. The kinetic parameters for AGT variants are compiled in the Table 5.2.3. These variants display similar specific activities ($V_{max}$) in comparison with AGT-Ma. In addition, there are no large changes in the values obtained for the Michaelis-Menten constant of L-alanine ($K_{M,Ala}$) and glyoxylate ($K_{M,Glyox}$) suggesting similar affinities for both substrates. The only exception for this behaviour is found in the variant H83R-Mi, which shows a reduced $V_{max}$ of around 15-fold while keeps similar $K_M$ values for substrates as it is shown in the Figure 5.2.4.

TABLE 5.2.2. **Hydrodynamic diameter of holoenzyme AGT variants determined by DLS**.

| *Variants* | *Hydrodynamic diameter*[1] *nm* |
|---|---|
| AGT-Ma | 8.1 ± 2.3 |
| AGT-Mi | 7.8 ± 0.1 |
| P11L-Ma | 9.8 ± 2.8 |
| I340M-Ma | 8.2 ± 0.1 |
| H83R-Mi | 8.9 ± 2.3 |
| F152I-Mi | 8.9 ± 0.2 |
| G170R-Mi | 6.6 ± 0.9 |
| R197Q-Mi | 9.9 ± 0.7 |
| I244T-Mi | 7.1 ± 0.7 |
| A295T-Mi | 9.5 ± 0.4 |
| P319L-Mi | 7.3 ± 1.0 |
| A368T-Mi | 8.2 ± 0.2 |

[1] *Data from 3 to 9 independent measurements.*

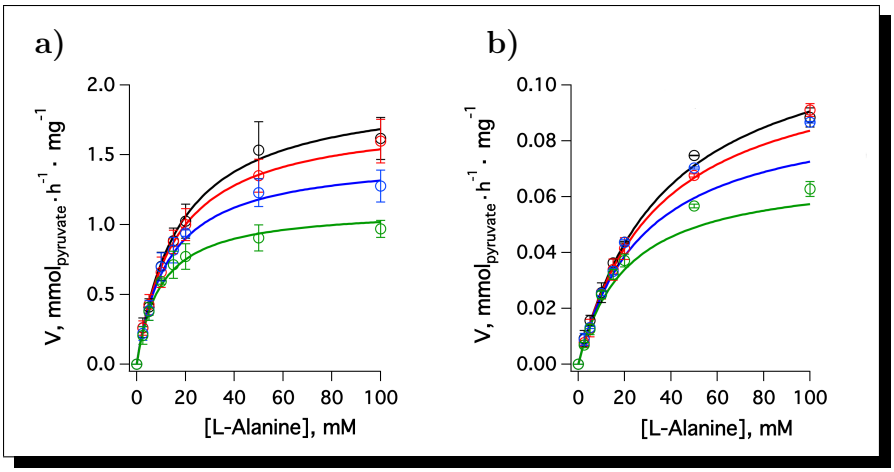


FIGURE 5.2.4. **Enzyme activity measurements at 37°C for a) AGT-Ma and b) H83R-Mi.** Results obtained for different L-alanine concentration in presence of 0.25 mM (green), 0.5 mM (blue), 1 mM (red) and 2 mM (black) glyoxylate in 4 different independent measurements. Lines are the fit of data.

TABLE 5.2.3. **Enzyme kinetic parameters at 37°C obtained by using a double-displacement mechanism**.

| *Variants* | $V_{max}$ [1] $m\,mol_{pyr} \cdot h^{-1} \cdot mg^{-1}$ | $K_{M,Ala}$ [1] $mM$ | $K_{M,Glyox}$ [1] $\mu M$ |
|---|---|---|---|
| AGT-Ma | 2.22 ± 0.09 | 19.5 ± 1.4 | 245 ± 29 |
| AGT-Mi | 2.26 ± 0.09 | 19.8 ± 1.5 | 197 ± 27 |
| P11L-Ma | 2.43 ± 0.08 | 18.3 ± 1.2 | 185 ± 22 |
| I340M-Ma | 3.03 ± 0.10 | 22.3 ± 1.3 | 277 ± 27 |
| H83R-Mi | 0.141 ± 0.005 | 42.4 ± 2.4 | 156 ± 19 |
| F152I-Mi | 1.90 ± 0.19 | 15.7 ± 3.1 | 201 ± 68 |
| G170R-Mi | 2.54 ± 0.13 | 21.9 ± 2.0 | 293 ± 42 |
| R197Q-Mi | 1.84 ± 0.06 | 18.1 ± 1.2 | 168 ± 21 |
| I244T-Mi | 2.82 ± 0.20 | 16.8 ± 2.4 | 193 ± 49 |
| A295T-Mi | 2.10 ± 0.10 | 18.5 ± 1.3 | 292 ± 30 |
| P319L-Mi | 1.54 ± 0.13 | 16.2 ± 2.7 | 212 ± 60 |
| A368T-Mi | 2.12 ± 0.13 | 23.0 ± 2.5 | 297 ± 50 |

[1] *Data from global fittings of 2-4 independent experiments.*

**The microenvironment of the active site is different in the H83R-Mi variant.** Because of the low enzymatic activity, the microenvironment of the coenzyme in the active site has been studied for H83R-Mi in comparison with AGT-Ma by near-UV visible absorption and circular dichroism (CD) spectra. These spectra were collected with the coenzyme forming the internal (PLP) and external (PMP) Schiff base (see Figure 5.2.5).

When the **coenzyme is bound by an internal Schiff base**, the AGT-Ma variant has a maximum in the absorption spectrum at 425 nm and a shoulder around 340 nm, were both peaks belong to different tautomeric forms of the bound cofactor through an internal Schiff base. In addition, the CD spectrum shows a strong positive peak centered at 430 nm and a weak negative peak around 340 nm. The CD spectrum also shows weak signals from 270 to 300 nm probably due to aromatic residues. On the other hand, the absorption and CD spectra of the H83R-Mi variant have peaks with similar intensity to those

of AGT-Ma but 10-15 nm blue shifted. These spectroscopic differences suggest that the coenzyme as PLP is bound and maintained in the active site of H83R-Mi but in a *different* microenvironment.

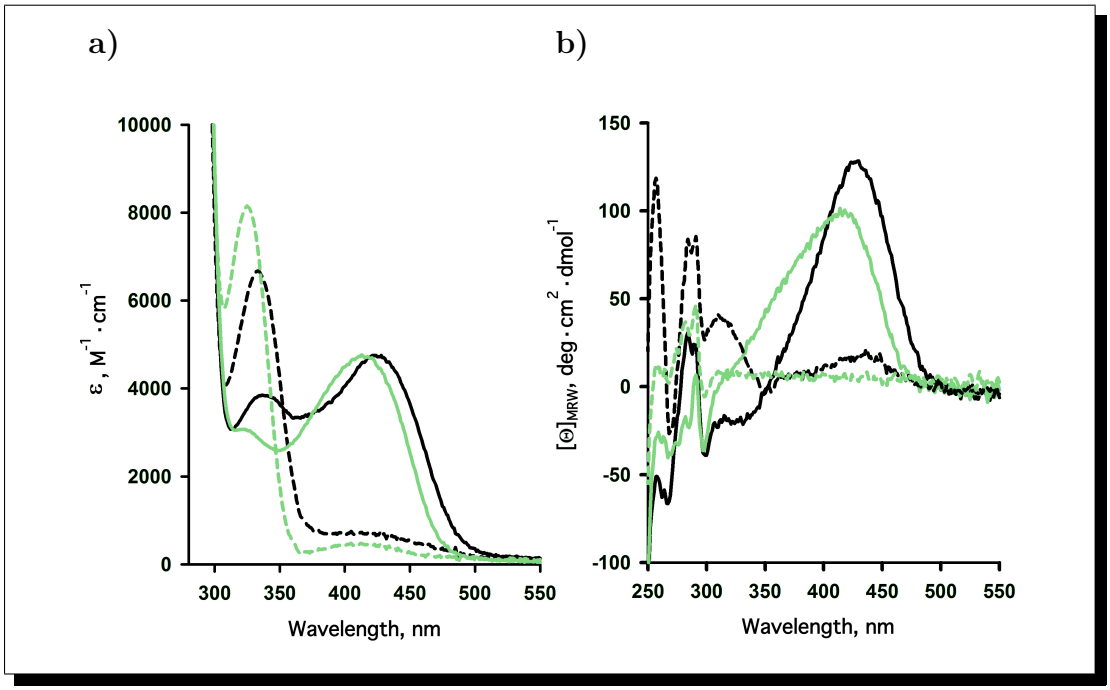


FIGURE 5.2.5. **Near-UV and visible spectroscopic characterization by a) absorbance and b) circular dichroism for AGT-Ma (black) and H83R-Mi (green).** The spectra for each variant is shown for PLP (continuous lines) and PMP bound (dashed lines) after incubation with 500 mM of L-alanine.

When both enzymes were incubated in the presence of 500 mM of L-alanine, the coenzyme **formed the external Schiff base** corresponding to half reaction of transamination. In this case, the AGT-Ma absorption spectrum has a band around 320-325 nm and the CD spectrum has also a positive dichroic band at 320 nm because of the PMP. Again, the H83R-Mi variant shows the same peak 10 nm blue shifted in the absorption spectra, and no CD signal from 300 to 550 nm. Therefore, these results support a distorted orientation of the coenzyme in the active site or maybe even the partial release of the PMP.

Because of PMP remains tightly bound during the catalytic turnover of AGT-Ma,[62] the ability of PMP to remain in the active site was tested for the H83R-Mi variant. By incubating both enzymes (36 µM in monomer) with L-alanine for 1 h at 25°C, the PLP was converted into PMP. The amount of PMP released was followed by UV absorption ($\epsilon_{325} = 8300\ M^{-1} \cdot cm^{-1}$) after

filtration (cut-off of 10 kDa). Under these conditions, only 3% of the PMP was released for AGT-Ma while for H83R-Mi this value reaches 44%. These results are consistent with a PMP tightly bound during the transamination reaction of AGT-Ma[62] and a weaker affinity of H83R-Mi variant for PMP, supporting its distorted active site micro-environment.

**PH1 variants show similar binding affinities for PLP.** The ability to bind the coenzyme PLP was measured by quenching of intrinsic Trp fluorescence. The affinity of the coenzyme was measured using two approaches: a kinetic and an equilibrium approach. The kinetic approach was used for those variants that underwent aggregation during the incubation with PLP, such as H83R-Mi, F152I-Mi, G170R-Mi, I244T-Mi, P319L-Mi and P11L-Ma. The rest of variants were analyzed under the equilibrium approach.

With both approaches, the main objective is to obtain an equilibrium dissociation constant value for the coenzyme PLP ($K_{D,PLP}$). The equilibrium approach allow us to determine $K_{D,PLP}$ directly, while in the kinetic approach this value is calculated indirectly from the kinetic rate constants of dissociation ($k_{off}$) and association ($k_{on}$) of the coenzyme under pseudo-first order conditions. The results obtained are summarized in the Table 5.2.4.

The results obtained for equilibrium dissociation constants differ in magnitude according to the approach used (see Table 5.2.4). Since the kinetic approach implies an indirect evaluation of $K_{D,PLP}$ , the uncertainty linked to the low value of the kinetic rate constants of dissociation ($k_{off}$) determines the reliability of the $K_{D,PLP}$ value. For this reason, the comparisons of $K_{D,PLP}$ values between variants have been made using AGT-Ma as a reference according to the approach used in each case (see Figure 5.2.6).

In general, considering the uncertainty of data, all variants show similar $K_{D,PLP}$ values. However, we can highlight some differences:

i): the H83R-Mi and F152I-Mi variants show an incremented kinetic rate constants ($k_{on}$ and $k_{off}$) of around 700% and 1500% *vs.* AGT-Ma, respectively (see Figure 5.2.6-a). This result suggests that although these variants show similar affinities for the coenzyme, the events of association and dissociation are more dynamic;

ii): the I244T-Mi variant shows a 2-fold higher $K_{D,PLP}$ value than AGT-Ma, which seems to exceed the uncertainty limits and is due to a

4-fold decrease in the $k_{on}$ rate constant, suggesting a lower affinity for the coenzyme;

iii): the P319L-Mi variant shows an apparent increase in the coenzyme binding affinity (its $K_{D,PLP}$ is 20% of the AGT-Ma value), however this fact cannot be confirmed because of the high uncertainty propagated from the dissociation constant rate to the $K_{D,PLP}$ value.

TABLE 5.2.4. **PLP binding affinities for the AGT variants obtained by kinetic and equilibrium approaches.**

| *Variants* | *Equilibrium* | *Kinetic* | | |
|---|---|---|---|---|
| | $K_{D,PLP}$ [1] | $K_{D,PLP}$ [2] | $k_{on}$ | $k_{off}$ |
| | $nM$ | $nM$ | $M^{-1} \cdot s^{-1}$ | $s^{-1}\,(\cdot 10^5)$ |
| AGT-Ma | 100 ± 30 | 1228 ± 273 | 189 ± 4 | 23 ± 5 |
| AGT-Mi | 157 ± 15 | N/D | N/D | N/D |
| P11L-Ma | N/D | 1069 ± 88 | 150 ± 1 | 16 ± 1 |
| I340M-Ma | 172 ± 42 | N/D | N/D | N/D |
| H83R-Mi | N/D | 1184 ± 383 | 1470 ± 160 | 174 ± 53 |
| F152I-Mi | N/D | 490 ± 298 | 2987 ± 68 | 146 ± 89 |
| G170R-Mi | N/D | 1378 ± 169 | 125 ± 2 | 17 ± 2 |
| R197Q-Mi | 156 ± 34 | N/D | N/D | N/D |
| I244T-Mi | N/D | 3058 ± 237 | 51 ± 1 | 16 ± 1 |
| A295T-Mi | 71 ± 16 | N/D | N/D | N/D |
| P319L-Mi | N/D | 240 ± 361 | 268 ± 7 | 6 ± 10 |
| A368T-Mi | 134 ± 26 | N/D | N/D | N/D |

*N/D, not determined*

[1] *$K_{D,PLP}$ data obtained from a single titration except AGT-Ma that was evaluated in 3 different titration experiments.*

[2] *$K_{D,PLP}$ estimated from kinetic experiments at different PLP concentrations. Errors were obtained from the fit of data ($K_{D,PLP} = k_{off}/k_{on}$).*

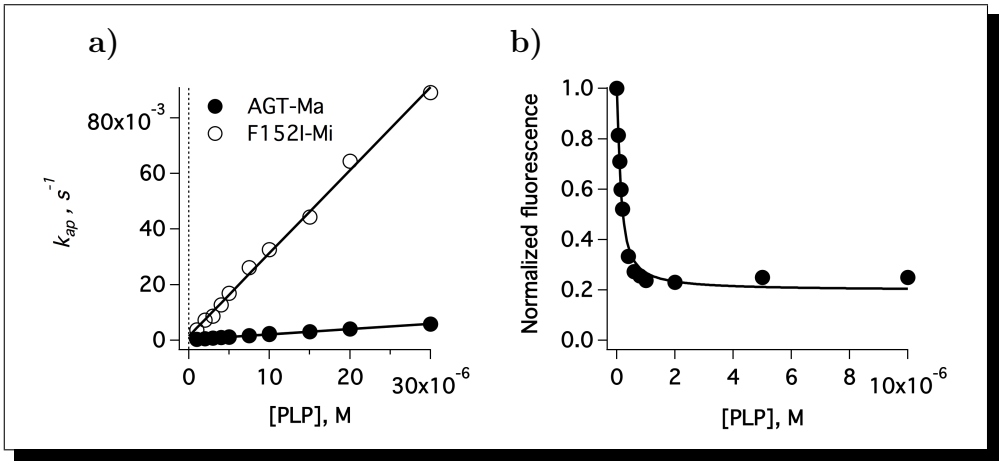


FIGURE 5.2.6. **Binding affinities of the AGT for PLP. a)** kinetics of PLP binding obtained by the kinetic approach in pseudo-first order conditions for the AGT-Ma and F152I-Mi; **b)** titration data obtained by the equilibrium approach for the AGT-Ma.

**PH1 mutations preferentially destabilize the apo-form of the AGT enzyme.** The thermal stability of the AGT variants has been evaluated by DSC. Under the experimental conditions used, thermograms obtained show a single endothermic transition (the system absorbs energy) for all variants in the holo and apo forms. The thermal unfolding of the AGT is well described by a two-state irreversible process[7] that can be represented as:

$$N \xrightarrow{k} F$$

where the native dimer ($N$) is unfolded to a final aggregated state ($F$) that cannot fold back and the process is characterized by a kinetic constant ($k$).[121] This model was fitted to the experimental data and it was also validated by consistency tests[121] of the activation energy values ($E_a$) (see Figure 5.2.7) for all AGT variants.

With this model we evaluated the $E_a$, the temperature where the maximum of the transitions occurs ($T_m$ parameter) and the enthalpy of the thermal unfolding ($\Delta H$). Alternatively, the order of the reaction ($1/\mu$) of the thermal unfolding was confirmed as a first order,[122] i.e., the transition state is a dimer ($\mu \approx 1$) and only for some variants in the apo state the order of the reaction displayed a small deviation from 1 (see Table 5.2.5).

TABLE 5.2.5. **Thermal denaturation parameters for AGT variants.**

| | *Variants* | $T_m$ [1] $^oC$ | $\Delta H_{cal}$ [2] $kcal \cdot kmol^{-1}$ | $E_a$ [3] $kcal \cdot kmol^{-1}$ | $\mu$ |
|---|---|---|---|---|---|
| HOLOENZYMES | AGT-Ma | 82.1 | 548 ± 4 | 109 ± 5 | 0.95 ± 0.02 |
| | AGT-Mi | 76.6 | 374 ± 14 | 101 ± 5 | 1.00 ± 0.01 |
| | P11L-Ma | 73.8 | 400 ± 4 | 95 ± 3 | 0.98 ± 0.01 |
| | I340M-Ma | 84.1 | 530 ± 24 | 115 ± 4 | 0.90 ± 0.03 |
| | H83R-Mi | 58.2 | 236 ± 16 | 55 ± 3 | 0.90 ± 0.03 |
| | F152I-Mi | 73.1 | 454 ± 18 | 97 ± 4 | 0.92 ± 0.04 |
| | G170R-Mi | 75.5 | 458 ± 18 | 100 ± 2 | 0.99 ± 0.01 |
| | R197Q-Mi | 77.9 | 402 ± 6 | 110 ± 6 | 0.96 ± 0.01 |
| | I244T-Mi | 75.8 | 404 ± 10 | 103 ± 2 | 0.95 ± 0.02 |
| | A295T-Mi | 77.5 | 500 ± 10 | 105 ± 8 | 0.91 ± 0.01 |
| | P319L-Mi | 76.5 | 394 ± 4 | 101 ± 3 | 0.98 ± 0.04 |
| | A368T-Mi | 76.5 | 440 ± 6 | 102 ± 7 | 0.98 ± 0.04 |
| APOENZYMES | AGT-Ma | 58.4 | 254 ± 16 | 111 ± 10 | 1.30 ± 0.02 |
| | AGT-Mi | 51.8 | 204 ± 10 | 121 ± 6 | 1.15 ± 0.06 |
| | P11L-Ma | 49.9 | 200 ± 14 | 105 ± 3 | 1.14 ± 0.01 |
| | I340M-Ma | 61.1 | 370 ± 22 | 143 ± 11 | 1.04 ± 0.03 |
| | H83R-Mi | 46.9 | 156 ± 6 | 62 ± 6 | 0.91 ± 0.04 |
| | F152I-Mi | 44.7 | 136 ± 18 | 87 ± 7 | 1.13 ± 0.04 |
| | G170R-Mi | 48.5 | 200 ± 12 | 92 ± 9 | 1.00 ± 0.04 |
| | R197Q-Mi | 52.3 | 200 ± 4 | 110 ± 10 | 0.96 ± 0.06 |
| | I244T-Mi | 47.3 | 162 ± 10 | 63 ± 1 | 0.88 ± 0.02 |
| | A295T-Mi | 49.4 | 244 ± 14 | 92 ± 9 | 0.89 ± 0.01 |
| | P319L-Mi | 47.1 | 194 ± 16 | 62 ± 2 | 0.88 ± 0.01 |
| | A368T-Mi | 51.8 | 200 ± 2 | 112 ± 6 | 1.15 ± 0.06 |

[1] *$T_m$ for a scan rate of 3°C/min (fitting errors < 0.1°C).*

[2] *mean±s.d. expressed per mol of AGT dimer from 3 different scan rates.*

[3] *mean±s.d. from 4 consistency test.*[121]

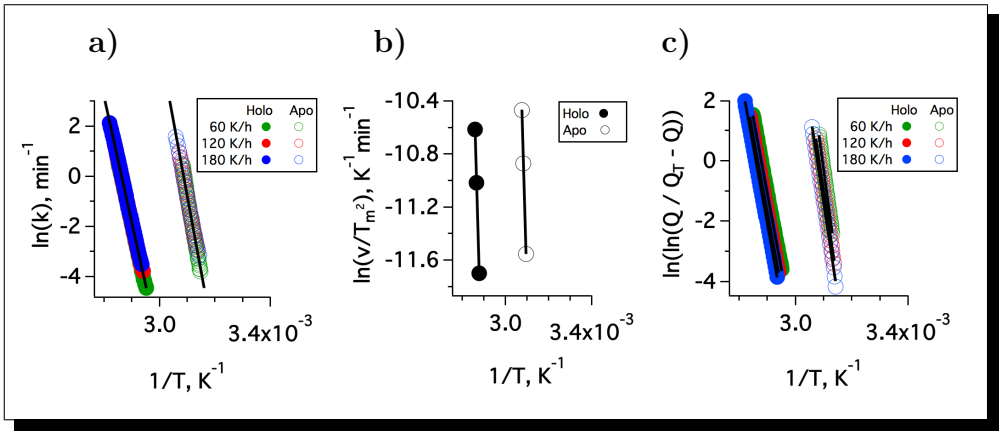


FIGURE 5.2.7. **Consistency tests that support the thermal unfolding of AGT follows a two-state irreversible process** (data for AGT-Mi). **a)** *Test A*: Arrhenius plot including $k$ data obtained from 3 scan rates. **b)** *Test B*: plot of $ln(\nu/T_m^2)$ *vs.* $1/T_m$ (in the plot the coordinate $1/T$ actually represents $1/T_m$). **c)** *Test C*: values of $ln[ln(Q_t/(Q_t - Q))]$ *vs.* $1/T$. Lines correspond to the fit of the data.

All thermal transitions for AGT variants share some common features: transitions are **irreversible**, as shown by the lack of signal in the second run for all samples; transitions are **kinetically controlled** as shown by the scan rate dependence of the $T_m$ parameter (see 5.2.8-a); and finally the reaction displays **first-order** kinetics as shown by the lack of the protein concentration dependence in the $T_m$ parameter (see 5.2.8-b) and values of $\mu \approx 1$ (see Table 5.2.5).

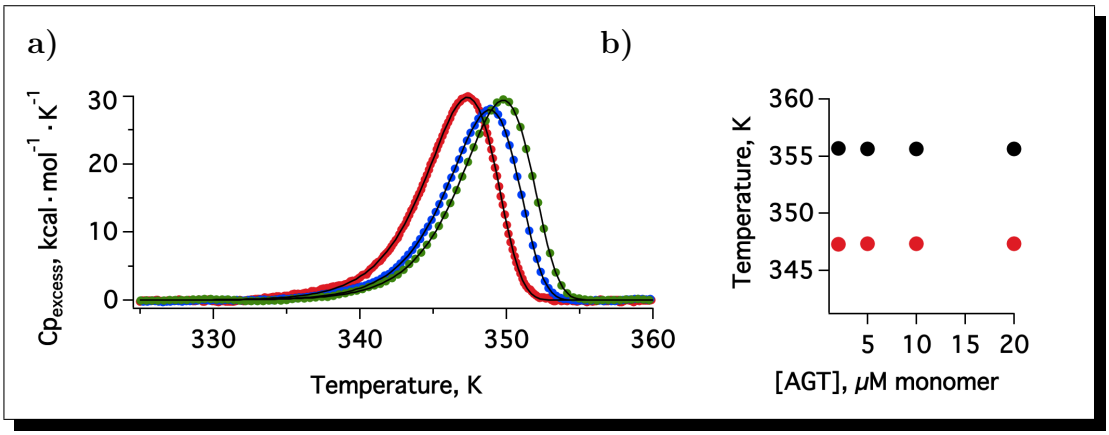


FIGURE 5.2.8. **Thermal transition at different scan rates and protein concentration dependence on $T_m$. a)** thermograms of holo AGT-Mi at different scan rates: 1°C/min (red), 2°C/min (blue) and 3°C/min (green). Lines are the best fit of data to a two-state irreversible model; **b)** lack of protein concentration dependence of the $T_m$ parameter. Data correspond to holo AGT-Ma (black) and P11L-Ma (red).

If we consider the $T_m$ as a simple operational parameter of the **thermal stability** we can make a preliminary comparison of the stability. Firstly, if we compare the stability for each variant between its holo and apo form, we observe that for all variants the removal of PLP has a dramatic effect on stability with a $T_m$ reduction of 20-25°C (see Table 5.2.5 and Figure 5.2.9).

Secondly, if we focus on holo enzymes we can observe that AGT-Ma variant has a $T_m$ that is 5.5°C higher than AGT-Mi, while the polymorphisms P11L-Ma and I340M-Ma have a difference of -2.8°C and +7.5°C compared to the $T_m$ of AGT-Mi. Moreover, PH1 holoenzymes have a similar $T_m$ (±1°C) than AGT-Mi and there are two exceptions, H83R-Mi and F152I-Mi variants that have a reduction of 17.8°C and 3.5°C respectively.

Finally, if we now compare the apo enzymes, we have a similar trend with the polymorphisms but almost all of PH1 variants have a reduction in the $T_m$ compared to AGT-Mi, such as F152I-Mi (-7.1°C), H83R-Mi (-4.9°C), P319L-Mi (-4.7°C) or I244T-Mi (-4.5°C).

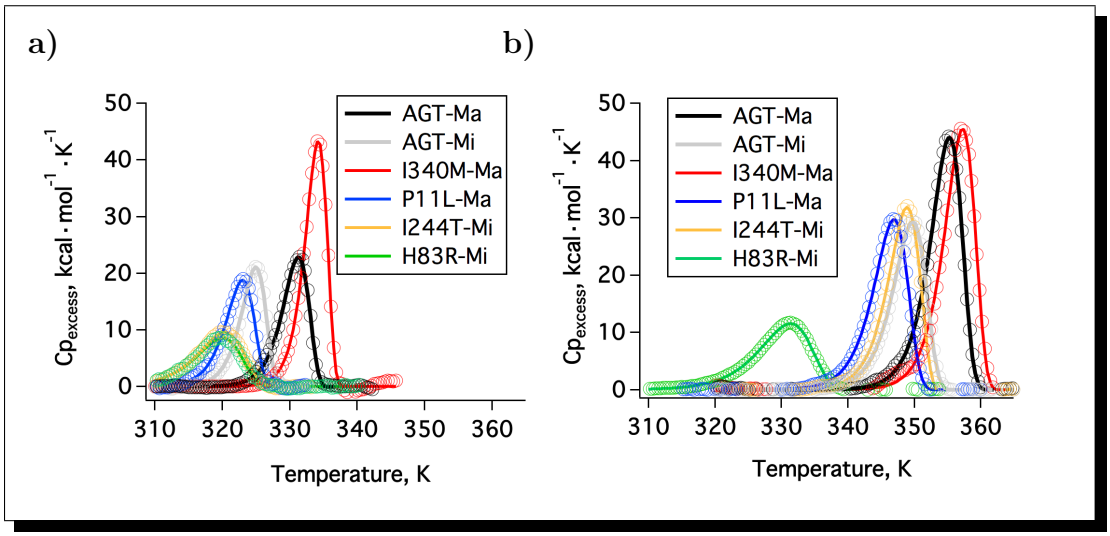


FIGURE 5.2.9. **Thermograms for the thermal unfolding, obtained by differential scanning calorimetry.** Different variants are shown (see color code) in the **a)** apo form and the **b)** holo form. Lines are the best fit of data to a two-state irreversible model.

In addition, we can go beyond the thermal stability information that is given by the $T_m$ parameter because the DSC allows us to obtain information about the **kinetic stability**, i.e., the time scale of denaturation. The kinetic constant of the unfolding as a function of the temperature can be extracted from

thermograms for each variant. Thus, the $E_a$ can be evaluated from the Arrhenius plots of the kinetic constants for AGT variants and we can calculate the half life ($t_{1/2}$) at different temperatures (being the half life inversely proportional to the denaturation kinetic constant). In Table 5.2.6 values of the kinetic constants extrapolated at the physiological temperature of 37°C are shown for holo and apo enzymes. Because, for some variants, the kinetic constant has to be extrapolated more than 40°C (from the range of temperatures where experimental data were collected to 37°C) values obtained for apo enzymes are expected to be more accurate, not even been necessary to extrapolate in some cases (see Figure 5.2.10).

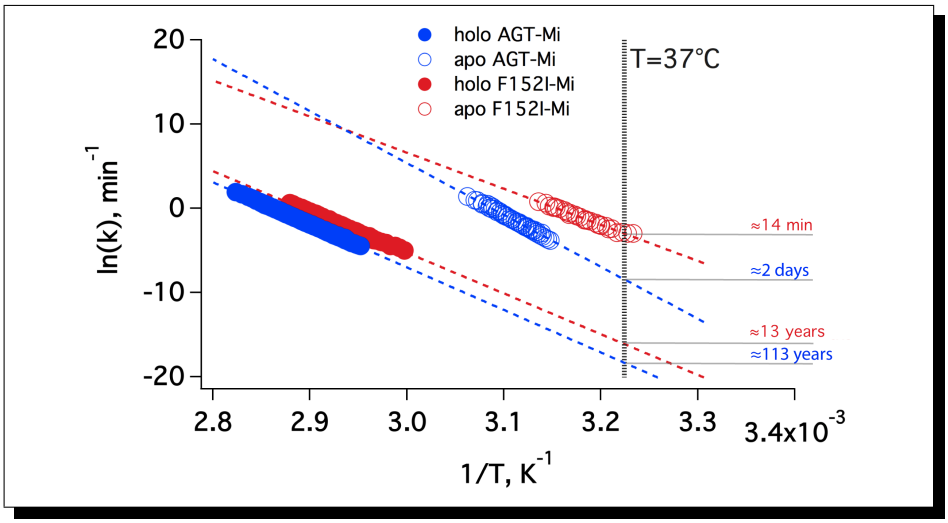


FIGURE 5.2.10. **Arrhenius plot for two AGT variants as holo and apo forms.** Lines correspond to the fit of experimental data (markers) and allow us to obtain the half life parameter at physiological temperature of 37°C (vertical dotted line).

According to data in Table 5.2.6, if we focus on the **kinetic stability of holo enzymes**, we can observe that all variants are kinetically stable, with half life at 37°C from 12-14 years (P11L-Ma and F152I-Mi respectively) to almost 80,000 years (for the most stable variant I340M-Ma). The only exception found is once again the H83R-Mi variant that shows a dramatic reduction in kinetic stability with a half-life of around 4 hours. Furthermore, the stability of disease-causing holoenzymes is almost similar if we compare with the AGT-Mi, with ratios that range from 0.2 to 6.0 (see Table 5.2.6), where only polymorphisms I340M-Ma and AGT-Ma are more stable (3 and 1 orders of magnitude respectively). On the other hand, if we focus now on **kinetic stability of apo**

**enzymes** we have a different scenario. Now, PH1 mutations dramatically reduce the kinetic stability of the enzyme that is translated into half life values from some minutes to a few hours. The H83R-Mi, I244T-Mi, F152I-Mi and P319L-Mi show a reduction on kinetic stability (at 37°C) of around 2 orders of magnitude in comparison with AGT-Mi (see Table 5.2.6). Moreover, in the same comparison other variants show a reduction in the kinetic stability of 1 order of magnitude (G170R-Mi and A295T-Mi and P11L-Mi). The polymorphisms AGT-Ma and I340M-Ma exhibit again high stability as apo enzymes.

TABLE 5.2.6. **Kinetic stability parameters for AGT variants.**

| | HOLOENZYMES | | | APOENZYMES | | |
|---|---|---|---|---|---|---|
| *Variants* | $k_{37}$ [1] $min^{-1}$ | $\frac{k_{AGT-Mi}}{k_{variant}}$ [2] | $t_{1/2}$ [3] *years* | $k_{37}$ [1] $min^{-1}$ | $\frac{k_{AGT-Mi}}{k_{variant}}$ [2] | $t_{1/2}$ [3] *hours* |
| AGT-Ma | $6.4{\cdot}10^{-10}$ | 34 | 2089 | $1.6{\cdot}10^{-5}$ | 16 | 722 |
| AGT-Mi | $2.2{\cdot}10^{-8}$ | 1.0 | 61 | $2.6{\cdot}10^{-4}$ | 1.0 | 44 |
| P11L-Ma | $1.1{\cdot}10^{-7}$ | 0.2 | 12 | $2.1{\cdot}10^{-3}$ | 0.12 | 6 |
| I340M-Ma | $1.7{\cdot}10^{-11}$ | 1300 | 78652 | $2.0{\cdot}10^{-7}$ | 1300 | 57762 |
| H83R-Mi | $2.9{\cdot}10^{-3}$ | $7.6{\cdot}10^{-6}$ | $4.6{\cdot}10^{-4}$ | $3.8{\cdot}10^{-2}$ | $6.8{\cdot}10^{-3}$ | 0.3 |
| F152I-Mi | $9.9{\cdot}10^{-8}$ | 0.2 | 14 | $4.9{\cdot}10^{-2}$ | $5.3{\cdot}10^{-3}$ | 0.2 |
| G170R-Mi | $3.7{\cdot}10^{-8}$ | 0.6 | 36 | $6.8{\cdot}10^{-3}$ | 0.04 | 2 |
| R197Q-Mi | $3.7{\cdot}10^{-9}$ | 6.0 | 361 | $5.5{\cdot}10^{-4}$ | 0.5 | 21 |
| I244T-Mi | $1.1{\cdot}10^{-8}$ | 2.0 | 122 | $3.6{\cdot}10^{-2}$ | $7.2{\cdot}10^{-3}$ | 0.3 |
| A295T-Mi | $1.2{\cdot}10^{-8}$ | 1.8 | 111 | $5.3{\cdot}10^{-3}$ | 0.05 | 2 |
| P319L-Mi | $9.4{\cdot}10^{-9}$ | 2.3 | 142 | $3.9{\cdot}10^{-2}$ | $6.6{\cdot}10^{-3}$ | 0.3 |
| A368T-Mi | $2.1{\cdot}10^{-8}$ | 1.0 | 64 | $5.9{\cdot}10^{-4}$ | 0.4 | 20 |

[1] *Kinetic constant rates for irreversible denaturation extrapolated to 37° C (fitting errors < 10%).*
[2] *Ratio of kinetic constant rates compared to AGT-Mi at 37° C.*
[3] *Kinetic stability at 37° C expressed as half life ($t_{1/2}$) from Arrhenius plots.*

As mentioned above, the calculation of the rate constant at 37°C for some variants, requires a long extrapolation in which the uncertainty of the Arrhenius plot slope is propagated as we move far away from experimental data.

Nonetheless, experimental data of variants in the apo form are close enough to the physiological temperature to provide an accurate value of the kinetic constant (see Figure 5.2.10). In order to prove this accuracy, we have tested the stability of variants in apo form by thermal inactivation experiments after incubation at 37°C (see Figure 5.2.11). Kinetic constant and half life values measured at 37°C by this procedure are consistent with those obtained by DSC thermograms (see inset in Figure 5.2.11).

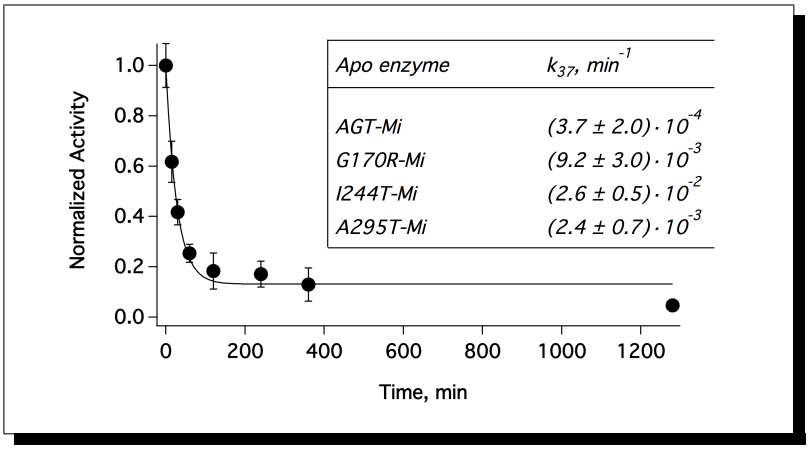


FIGURE 5.2.11. **Thermal inactivation of the apoenzyme I244T-Mi.** Enzyme activity was measured at different times of incubation at 37°C. In the table are shown the constant rates obtained by thermal inactivation for AGT-Mi and some PH1 variants. Errors are from 3 independent measurements.

## 5.3. Discussion

At the beginning of this project, only some common alleles associated to PH1, such as G170R-Mi,[7,123] I244T-Mi[7,59,124] and F152I-Mi,[125] had been characterized. During this work, we have characterized a set of twelve polymorphic variants where six of them are known PH1 mutations. To date, this work represents one of the most exhaustive and complete *in vitro* characterization of PH1 mutations.

Few AGT mutations have been found to affect the enzymatic activity[62] and among the PH1 mutations characterized in this work, only H83R-Mi variant has manifested a clear catalytic defect. According to the AGT structure, the substitution of the histidine 83 by an arginine is expected to modify the orientation of the coenzyme in the active site due to steric and electrostatic effects. We have shown that the H83R-Mi variant is not able to undergo an efficient transamination showing a low affinity for PMP. In addition, H83R-Mi is not well

stabilized upon the binding of the coenzyme and is highly instable especially as holo form. Therefore, molecular defects of this variant may imply a distortion of the proper orientation of the coenzyme in the active site,[62] that could influence the optimal and required protonation state of PLP for the transamination.[52]

In our study, we have shown that PH1 mutations mainly affect the protein stability. We found that PLP binding to AGT enzyme improves the kinetic stability of native state in 4-5 orders of magnitude thus holo enzymes are kinetically stable at physiological temperature. Also, in a comparison with the non-pathogenic AGT-Mi variant, PH1 mutations such as F152I-Mi, G170R-Mi, I244T-Mi, A295T-Mi and P319L-Mi are highly destabilized as apo proteins with fast rates of denaturation at physiological temperature.

The apo state of AGT enzyme could be transiently populated *in vivo* because given the high reactivity of free PLP,[126] its concentration is tightly regulated by *pyridoxal kinase* (PLK) and *pyridoxine-5'-phosphate oxidase* (PNPOx) enzymes. The PLP transference is thought to be mediated by a complex between PLK or PNPOx and the $B_6$ enzyme.[127] Therefore, the low kinetic stability of apo disease-causing variants could have important implications in protein homeostasis of AGT. The apo state might be susceptible to intracellular irreversible alterations, such as mitochondrial mistargeting, aggregation or degradation.[7] Therefore, pharmacological therapies based on stabilization of native state (such as pyridoxine supplementation) may recover AGT function.[5,31]

When we started this project, only the thermal stability of F152I-Mi and G170R-Mi had been reported earlier by circular dichroism and inactivation experiments in which the thermal destabilization was mainly shown for the apo state.[123,125] Our detailed DSC analysis provides information about AGT kinetic stability, denaturation mechanism and structural/energetic features of its denaturation free energy barrier previously unexplored. Our data suggest that mutational effects must concern mainly the structure and energetics of the dimeric native and/or transition state for denaturation. Thus, those effects on dimer dissociation and full monomer unfolding do not contribute to the relevant kinetic stability since they occur after the rate-limiting step.

Some of the characterized variants (F152I-Mi, G170R-Mi, I244T-Mi and P319L-Mi) show evident signs of misfolding and possible accelerated turnover upon expression in CHO (*Chinese hamster ovary*) cells (see Figure 5.3.1-a and -b). Total immunoreactive AGT protein levels are reduced in most of mutations

compared to AGT-Ma suggesting an enhanced AGT turnover.[59,60] The correlation between protein levels and specific activities in soluble extracts suggests that soluble fraction is composed by active dimeric AGT. While AGT-Ma is mainly found in soluble fraction (90%), mutations decrease the protein fraction in soluble extracts suggesting a reduced intracellular *foldability* by enhancing aggregation. Moreover, two mutations (G170R-Mi and F152I-Mi) also cause protein mitochondrial mistargeting in CHO cells (see Figure 5.3.1-c). Taken together, these results suggest that there is some degree of correlation between the low stability of the apo state and protein mitochondrial mistargeting and/or aggregation (see Table 5.3.1 and Appendix A).

TABLE 5.3.1. **Summary of mutational effects on molecular properties of AGT protein.**

| *Variants* | *Kinetic Stability* | | *Eukaryotic Cell Expression* | | | | *Hsc*70 *interaction* |
|---|---|---|---|---|---|---|---|
| | *Holo* | *Apo* | *Total AGT* | *Soluble AGT* | *Activity* | *Targeting* | |
| AGT-Ma | ↑ | ↑ | ↑↑ | ↑ | ↑ | Peroxisomes | ↓↓ |
| AGT-Mi | = | = | = | = | = | Peroxisomes | = |
| P11L-Ma | ↓ | ↓ | = | = | = | Peroxisomes | = |
| I340M-Ma | ↑↑ | ↑↑ | ↑↑ | ↑ | ↑ | Peroxisomes | ↓↓ |
| H83R-Mi | ↓↓↓ | ↓↓ | = | = | ↓↓ | Peroxisomes | =↑ |
| F152I-Mi | ↓ | ↓↓ | ↓ | ↓ | ↓ | Mitochondria | =↑ |
| G170R-Mi | = | ↓↓ | = | ↓↓ | ↓↓ | Mitochondria | =↑ |
| R197Q-Mi | ↑ | = | ↑↑ | ↑ | ↑ | Peroxisomes | =↑ |
| I244T-Mi | = | ↓↓ | ↓ | ↓ | ↓ | Peroxisomes | =↑ |
| A295T-Mi | = | ↓↓ | ↑ | ↑ | = | Peroxisomes | ↑ |
| P319L-Mi | = | ↓↓ | ↓ | ↓↓ | ↓↓ | Peroxisomes | ↑ |
| A368T-Mi | = | = | = | = | = | Peroxisomes | =↑ |

*Semiquantitative comparison taking AGT-Mi as reference. Symbols indicate: (↑) increase, (↓) decrease, (=↑) slightly increase and (=) unchanged (see Appendix A).*

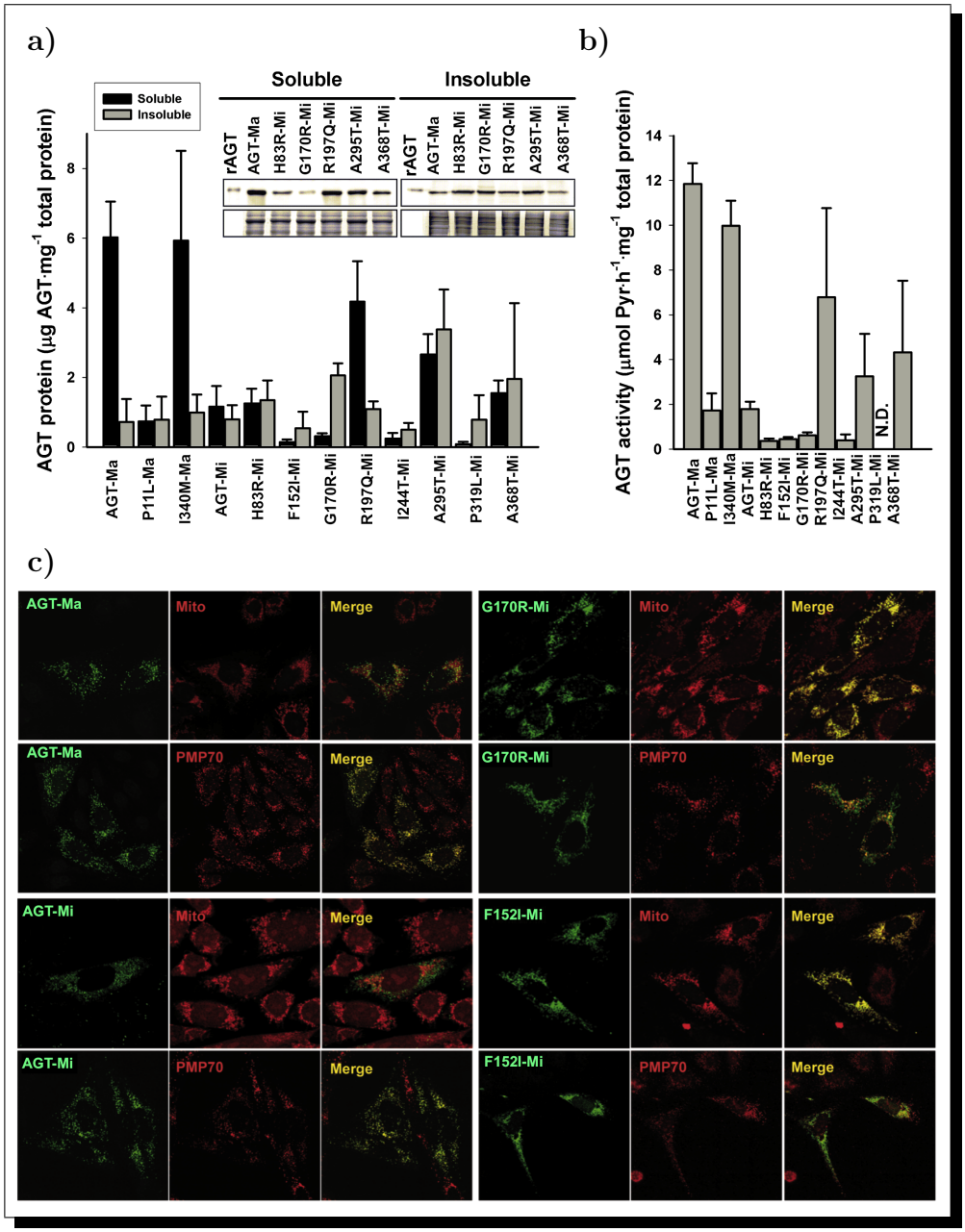


FIGURE 5.3.1. **Expression studies of AGT variants in CHO cells. a)** levels of immunoreactive AGT in soluble (black) and insoluble (grey) extracts. Inset shows representative immunodetection experiments (upper) and loading controls (bottom); rAGT: recombinant His-tagged AGT; **b)** AGT activities in soluble extracts. Data are means±s.d. of 2-5 independent experiments, N.D. not detectable; **c)** immunolocalization in the mitochondria (upper) and peroxisome (bottom) for AGT-Ma, AGT-Mi, G170R-Mi and F152I-Mi. *Data were obtained by Prof. Eduardo Salido from Centre for Biomedical Research on Rare Diseases at University of La laguna through the collaboration on this project (see Appendix A).*

The AGT-Mi is a non-pathogenic natural polymorphism that is known to sensitize AGT towards deleterious mutations.[125] In this respect, AGT-Mi has shown a reduction in protein kinetic stability and enhanced protein misfolding and degradation (see Figure 5.3.1-a and -b). The disease-causing mutations over the AGT-Mi background appear to exacerbate some of these defects. Strong interactions with Hsc90, Hsc70[7,59] and bacterial Hsp60[124] have been reported previously for I244T-Mi and G170R-Mi variants. Furthermore, all PH1 mutations analyzed in the present work have shown strong interactions with Hsc70 chaperones upon expression in a cell-free system and immunoprecipitation using antibodies anti-Hsc70 (see Figure 5.3.2).

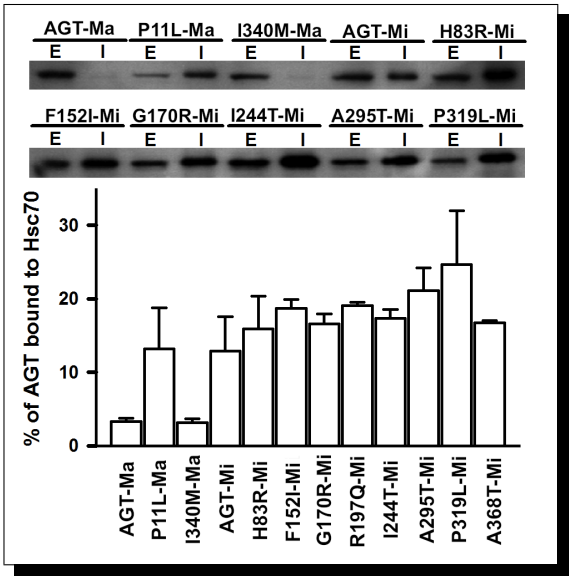


FIGURE 5.3.2. **Interaction of AGT variants with Hsc70 chaperones in a cell-free system (TnT).** Upper) representative autoradiograms of AGT proteins labeled with $^{35}S-Met$. E: total AGT synthesized in extracts, I: AGT immunoprecipitated using anti-Hsc70 antibodies. Note that 1 µl of the TnT reaction was loaded in E lanes while 6 µl from TnT lysate was loaded in I lanes; Bottom) percentage of immunoprecipitated AGT compared to total AGT synthesized. Mean±s.d. from 3 independent experiment. *Data were obtained by Prof. Eduardo Salido from Centre for Biomedical Research on Rare Diseases at University of La laguna through the collaboration on this project (see Appendix A).*

Therefore, we propose that acquisition of the tertiary and dimeric quaternary structure is crucial for proper folding of AGT. Enhanced interactions of PH1 mutants with molecular chaperones may suggest a rougher folding landscape with higher population of kinetic/equilibrium intermediates.[11] This complexity in the proteostasis of AGT together with differences in the proteostasis

network may explain inter-individual variability[128] in clinical presentation for patients sharing a given phenotype.[20]

Chaperones are key components to protein quality control by promoting folding or degradation.[65,71,76] We have shown that PH1 mutations mainly affect protein folding and stability suggesting that an alteration in protein homeostasis could be involved in defective function of AGT protein. Despite our findings, those specific events and interactions responsible for the partition between protein mitochondrial mistargeting, aggregation and degradation remain elusive.[7] Consequently, molecular chaperones emerge as an important checkpoint in the folding of PH1 mutants, likely by partitioning AGT folding intermediates into native dimers, peroxisomal import, mitochondrial mistargeting, aggregation or degradation. Therefore, a detailed characterization of PH1 mutants interactions during protein folding and misfolding could open new approaches for therapeutic intervention in PH1.

CHAPTER 6

# *C. elegans* as an animal model for PH1

## 6.1. Brief introduction

In the previous chapter, we have shown that the alterations of AGT proteostasis by PH1 mutations are complex and largely unexplored. *In vivo* studies using an animal model would help us to further develop our preliminary observations. Thus, we selected *C. elegans* as model organism because there are many advantages in the use of this nematode as an animal model for protein-misfolding diseases.[89,105,129,130]

Two goals were set in our approach to this model organism: i) demonstrating whether *C. elegans* may be used as a model for PH1 i.e., whether we could be able to generate a hyperoxaluric phenotype in this nematode; ii) the use of the nematode as a test bench to identify and modulate key proteostasis elements and thus design new therapeutic approaches to PH1 by targeting critical interactions. Due to complications during the nematode strains construction, only the former goal is addressed in this chapter.

It has been predicted, by an *in silico* search,[103] an orthologous gene (ORF named *T14D7.1*) to the human *AGXT* gene. We have compared this orthologous protein with the human AGT, in order to have more knowledge about the physiological role of an AGT orthologous in the nematode (where glyoxylate is not a metabolic end product), the potential functional divergence with human AGT and the possibility to create a hyperoxaluric phenotype in the nematode.

In this chapter the *in vitro* characterization of the nematode orthologous protein, in a comparison with human AGT, is presented and discussed.

## 6.2. Results

**The predicted protein sequence of T14D7.1 appears to be an orthologous of peroxisomal human AGT.** The predicted ORF of *T14D7.1* was found with a expectation value (E-value i.e., alignments expected to occur in a database search by chance) of $3.1 \cdot 10^{-71}$ in comparison to human *AGXT* gene.[103] Accordingly, a protein sequence comparison of the T14D7.1 with the human AGT and AGT2 show that the protein of the nematode is much more similar to the peroxisomal human AGT (see Figure 6.2.1 and Table 6.2.1). This comparison shows that identity between T14D7.1 and human AGT sequences is around 41% (167 identical positions) with an additional similarity of 30% (121 similar positions). The *T14D7.1* gene is composed by 3557 bp that are divided in 11 exons in the chromosome 2 of the nematode.[103] Translation of this gene results in a protein with 405 residues, which is 13 residues longer than human AGT. The extra residues of the T14D7.1 protein are located in the N-terminal domain (17 residues longer than human AGT) and the sequence lacks the C-terminal tripeptide required in human AGT for peroxisomal import (see Figure 6.2.1).

```
1    -----------------MASHKLLVTPPKALLKPLSIPNQLLLGPGPSNLPPRIMAAGGL   43  humanAGT
1    MISTRFLRPSVSIFGFGIKSSMSSRAPPKALLQDMVVPPRQLFGPGPSNMADSIAETQSR   60  T14D7.1
                     : *     :******: : :* : *:******:   *  : .

44   QMIGSMSKDMYQIMDEIKEGIQYVFQTRNPLTLVISGSGHCALEAALVNVLEPGDSFLVG  103  humanAGT
61   NLLGHLHPEFVQIMADVRLGLQYVFKTDNKYTFAVSGTGHSGMECAMVNLLEPGDKFLVV  120  T14D7.1
     :::* :  :: *** ::: *:****:* *  *:.:**:**..:*.*:**:*****.***

104  ANGIWGQRAVDIGERIGARVHPMTKDPGGHYTLQEVEEGLAQHKPVLLFLTHGESSTGVL  163  humanAGT
121  EIGLWGQRAADLANRMGIEVKKITAPQGQAVPVEDIRKAIADYKPNLVFVCQGDSSTGVA  180  T14D7.1
       *:*****.*:.:*:* .*: :*    *     ::::.:.:*::** *:*: :*:*****

164  QPLDGFGELCHRYKCLLLVDSVASLGGTPLYMDRQGIDILYSGSQKALNAPPGTSLISFS  223  humanAGT
181  QPLETIGDACREHGALFLVDTVASLGGTPFAADDLKVDCVYSATQKVLNAPPGLAPISFS  240  T14D7.1
     ***: :*: *:.: .*:***:*******:  *   :* :**.:**.****** : ****

224  DKAKKKMYSRKTKPFSFYLDIKWLANFWGCDDQPRMYHHTIPVISLYSLRESLALIAEQG  283  humanAGT
241  DRAMEKIRNRKQRVASFYFDAIELGNYWGCDGELKRYHHTAPISTVYALRAALSAIAKEG  300  T14D7.1
     *:* :*: .** :  ***:*   *.*:**** : :  ****  *: ::*:** :*: **::*

284  LENSWRQHREAAAYLHGRLQALGLQLFVKDPALRLPTVTTVAVPAGYDWRDIVSYVIDHF  343  humanAGT
301  IDESIQRHKDNAQVLYATLKKHGLEPFVVDEKLRLPCLTTVKVPEGVDWKDVAGKMMT-N  359  T14D7.1
     :::* ::*:: *  *:. *:  **: ** *  **** :*** ** * **:*:.. ::

344  DIEIMGGLGPSTGKVLRIGLLGCNATRENVDRVTEALRAALQHCPKKKL  392  humanAGT
360  GTEIAGGLGATVGKIWRIGTFGINSNSTKIENVVELLSKSIGEKSK---  405  T14D7.1
       ** **** :.**: *** :* *:.  ::: .*.* *  :: .  *
```

FIGURE 6.2.1. **Sequence alignment of the human AGT (GI code: 134855) and the predicted protein sequence of T14D7.1 (GI code: 5824614).** Identical residues are marked as (*), conserved substitutions as (:), substitutions with similar shape as (.) and gaps are marked by (-). The residues of the active site are highlighted in red.

TABLE 6.2.1. **Protein sequence comparison between peroxisomal human AGT, mitochondrial human AGT2 and *C. elegans* T14D7.1.**

| | *AGT vs AGT2* | *AGT vs T14D7.1* | *AGT2 vs T14D7.1* |
|---|---|---|---|
| Identical positions | 75 (19%) | 167 (41%) | 77 (19%) |
| Similar positions | 112 (28%) | 121 (30%) | 103 (25%) |

Comparison made by using Clustal Omega algorithm.[131]

Moreover, T14D7.1 protein should be able to bind the coenzyme PLP because the main residues involved in the binding and stabilization of the coenzyme in the active site are conserved in both sequences (see highlighted residues in Figure 6.2.1).

**T14D7.1 is a soluble protein with a dimeric quaternary structure.** We have expressed and purified the T14D7.1. protein (for more details see M&M section) with an average yield of 4 mg of soluble product per liter of LB medium. The size exclusion chromatography, performed during the purification, gives a single peak centered around 84.9 ml (see Figure 6.2.2). The retention volume obtained for T14D7.1 was similar but slightly higher than the one of dimeric human AGT (82.6 ml as was shown in the previous chapter). This result suggests that T14D7.1 also seems to be a dimer with similar size than human AGT.

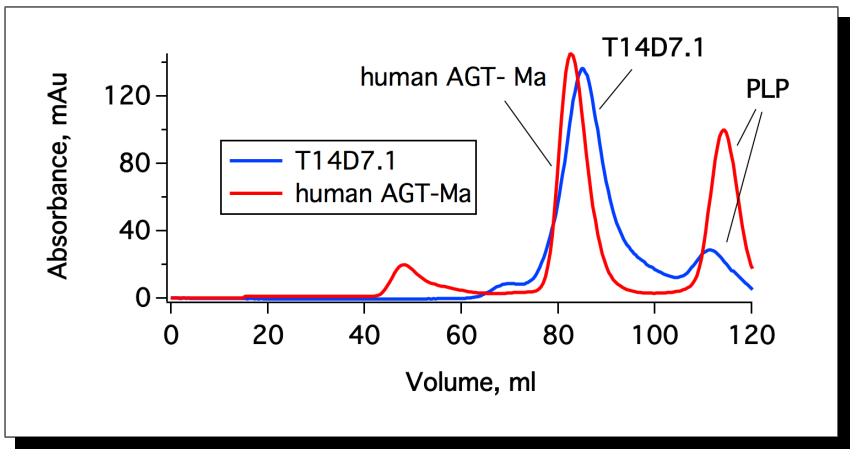


FIGURE 6.2.2. **Size exclusion chromatogram for T14D7.1.** SEC was performed using a HiLoad$^{TM}$ 16/60 Superdex$^{TM}$ 200 size exclusion chromatography column. Data correspond to T14D7.1 (blue) and human AGT-Ma (red).

The functional assembly of T14D7.1 was further characterized by DLS. The hydrodynamic diameter obtained for T14D7.1 is 8.8 ± 0.4 nm, which is slightly larger than the human AGT-Ma obtained in a control experiment (8.1 ± 0.1). Polydispersity of the measurements is 15 ± 1 % and 7 ± 6 % respectively, which are consistent with the narrow particle size distribution of a monodisperse sample. Therefore, the results from DLS are consistent with a T14D7.1 protein folded as a dimer but slightly larger than the human AGT.

**The spectroscopic characterization of T14D7.1 confirms its ability to bind PLP and PMP.** The residues of the active site involved in the PLP binding are conserved according to Figure 6.2.1. Therefore, the ability to bind the coenzyme was measured by near-UV visible absorbance and circular dichroism spectra for T14D7.1, using the human AGT-Ma as a reference (see Figure 6.2.3). The spectra were measured with the coenzyme PLP and also PMP after incubation of the protein with L-alanine.

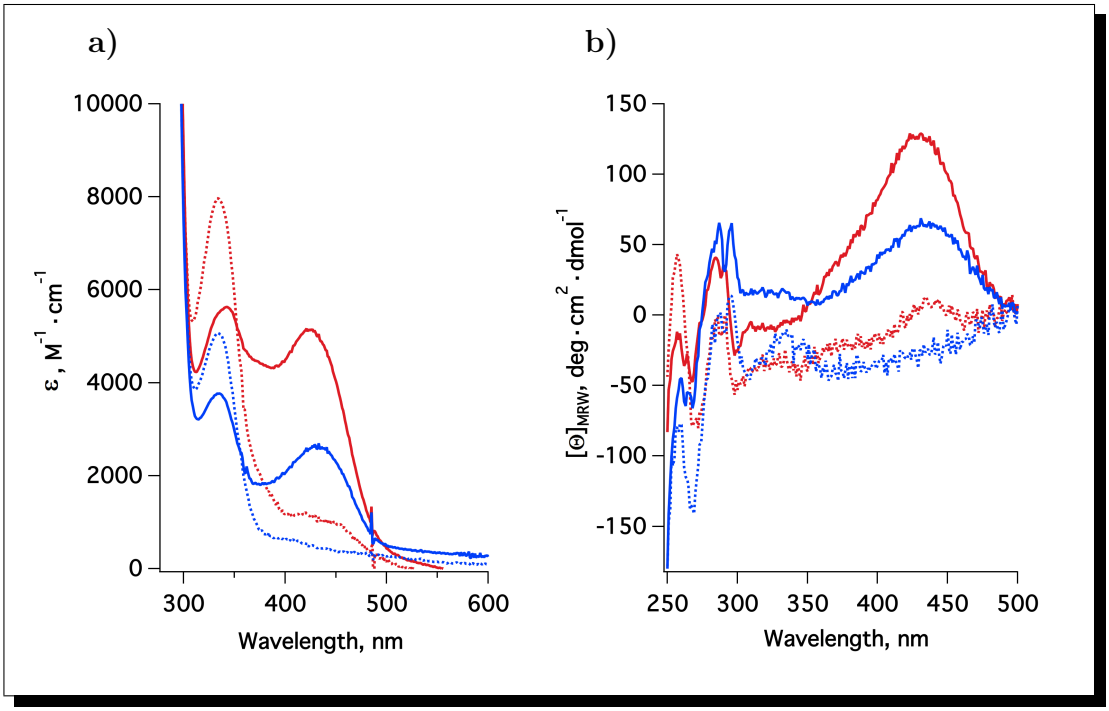


FIGURE 6.2.3. **Near-UV visible spectroscopic characterization by a) absorbance and b) circular dichroism for T14D7.1 (blue) and AGT-Ma (red).** The spectra for each variant is shown for PLP (continuos lines) and PMP bound (dashed lines) after incubation with 500 mM of L-alanine.

In the **absorbance spectra** of the holo forms (coenzyme as PLP) we observe a peak centered around 420 nm that corresponds to the ketoenamine tautomer of the internal Schiff base of PLP and another peak centered around 340 nm that is assumed to belong to the corresponding enolimine tautomer[123] of the coenzyme in the active site. After incubation with L-alanine, spectra of both proteins show a main peak at around 340 nm that correspond to the PMP form of the coenzyme, with almost no absorption in the region of 420 nm.

In addition, **CD spectra** also show the ability of T14D7.1 to bind PLP with a positive dichroic signal centered at 420 nm having lower intensity than human AGT. Again, when both proteins are found in the presence of the substrate, the formation of PMP is displayed by a reduction of the peak at 420 nm and the appearance of a CD signal at 340 nm. These results support that T14D7.1 protein is able to bind PLP through an internal Schiff base. In addition, the coenzyme in the active site of T14D7.1 is properly aligned and activated to take the amino group from the substrate (L-alanine) producing PMP, which is kept in the active site as an external Schiff base.

Attempts to obtain the apo form of the T14D7.1 protein have failed so far. The procedure applied was similar to the human AGT. After incubation with L-alanine, the internal Schiff base is broken giving the external Schiff base. Then, the external Schiff base could be removed from the active site probably due to a relaxation of the quaternary structure by the pH reduction.

The application of this procedure was not enough to remove the external Schiff base (PMP) from the active site of the T14D7.1. The use of even lower pH than the one we used for human AGT leads to protein aggregation before the release of the coenzyme. Therefore, any measurable fraction of the apo form was obtained.

To confirm that PMP remained bound, the treated protein was incubated in the presence of glyoxylate in order to complete the transamination reaction. As a result, the spectrum of T14D7.1 as holo-PLP was obtained, in which the absorbance peak at 420 nm was restored while the peak at 340 nm got reduced (see Figure 6.2.4).

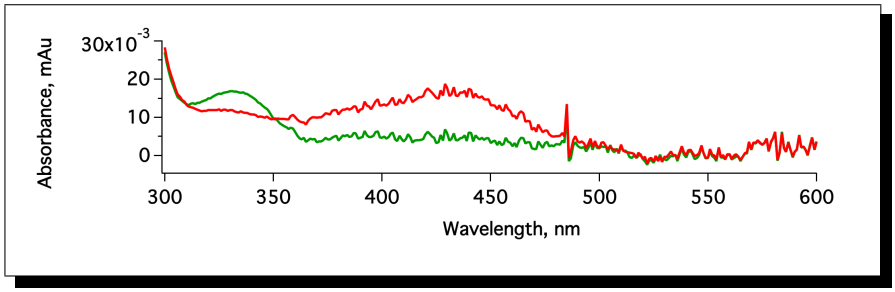


FIGURE 6.2.4. **Reconstitution of the holo-PLP T14D7.1 spectrum by incubation with L-alanine**. The spectrum of the protein, after several treatments to remove the PMP from the active site (green), returns back to the spectrum of the holo-PLP protein (red), supporting that PMP remained bound.

**T14D7.1 is more active than human AGT.** We have measured the activity of the T14D7.1 on the natural substrates of the human AGT. Therefore, L-alanine has been used as an amino donor and glyoxylate as keto acid. Global fit of data at different substrate concentrations allows us to obtain the kinetic parameters for the overall transamination. Data are summarized in Figure 6.2.5 and Table 6.2.2 .

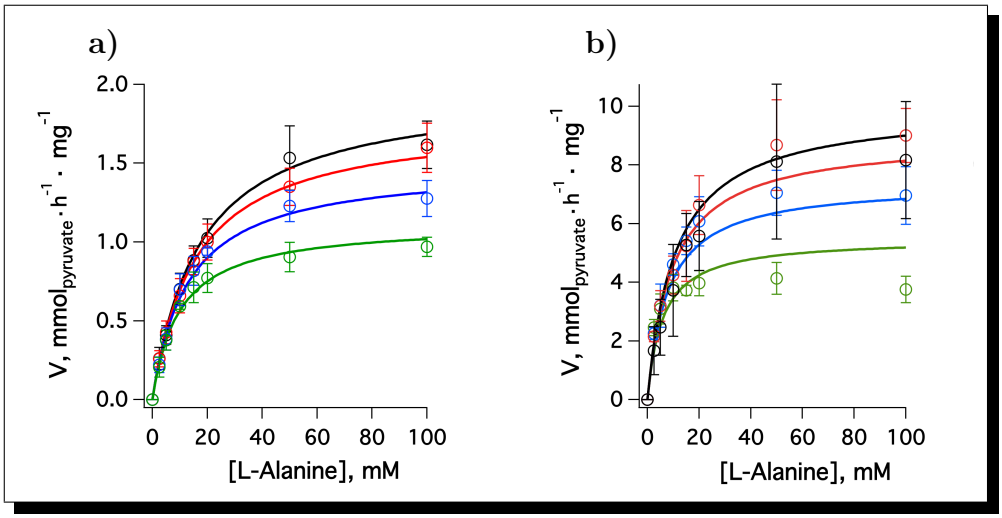


FIGURE 6.2.5. **Enzyme activity measurements for a) AGT-Ma and b) T14D7.1 at 37°C.** Results obtained for different L-alanine concentration in the presence of 0.25 mM (green), 0.5 mM (blue), 1 mM (red) and 2 mM (black) glyoxylate in 3-4 different independent measurements.

As we can observe, the specific activity of T14D7.1 is 5-fold higher than the human AGT-Ma at 37°C. However, affinities for both substrates are kept in a similar range and only the affinity for alanine is slightly increased in T14D7.1 (see Figure 6.2.5 and Table 6.2.2).

TABLE 6.2.2. **Kinetic parameters of transamination for T14D7.1 *vs.* human AGT-Ma protein at 37°C**. Data were obtained using L-alanine and glyoxylate.

| *Variant* | $V_{max}$ $m\,mol_{pyr} \cdot h^{-1} \cdot mg^{-1}$ | $K_{M,Ala}$ $mM$ | $K_{M,Glyox}$ $\mu M$ |
|---|---|---|---|
| *C. elegans* T14D7.1 | 11.28 ± 0.73 | 12.2 ± 1.7 | 263 ± 53 |
| *Human* AGT-Ma | 2.22 ± 0.09 | 19.5 ± 1.4 | 245 ± 29 |

In addition, we have evaluated the dependence of the specific activity of T14D7.1 with temperature and pH changes. On one side, through the studied range of temperature, both proteins have shown the same trend with an increase of specific activity on temperature (see Figure 6.2.6-a). If we consider that activity rate follows an Arrhenius temperature dependence, we can infer information about the activation energy ($E_a$) of the overall activity process (see Figure 6.2.6-b). Both proteins show a similar and small $E_a$ (see Table 6.2.3).

Moreover, the activity dependence on pH is also similar for both proteins (see Figure 6.2.6-c and -d). The sigmoidal fit of the data gives for both proteins an identical apparent $pK_a$ (see Table 6.2.3) suggesting that both proteins have a similar pattern of protonation states of the active site, which could be a sign of similar reaction specificities[52] and environmental pH in their intracellular location[132] (e.g., mitochondrion *vs* peroxisome).

TABLE 6.2.3. **Analysis of the temperature and pH dependence on specific activity.** Evaluation of the activation energy and the apparent $pK_a$ of the specific activity according to temperature and pH respectively, for T14D7.1 and AGT-Ma.

| *Variant* | *Temperature Dependence* | *pH Dependence* |
|---|---|---|
| | $E_a$ $kcal \cdot mol^{-1}$ | $pK_a$ |
| *C. elegans* T14D7.1 | 4.02 ± 0.9 | 6.03 ± 0.12 |
| *Human* AGT-Ma | 2.2 ± 1.8 | 5.96 ± 0.22 |

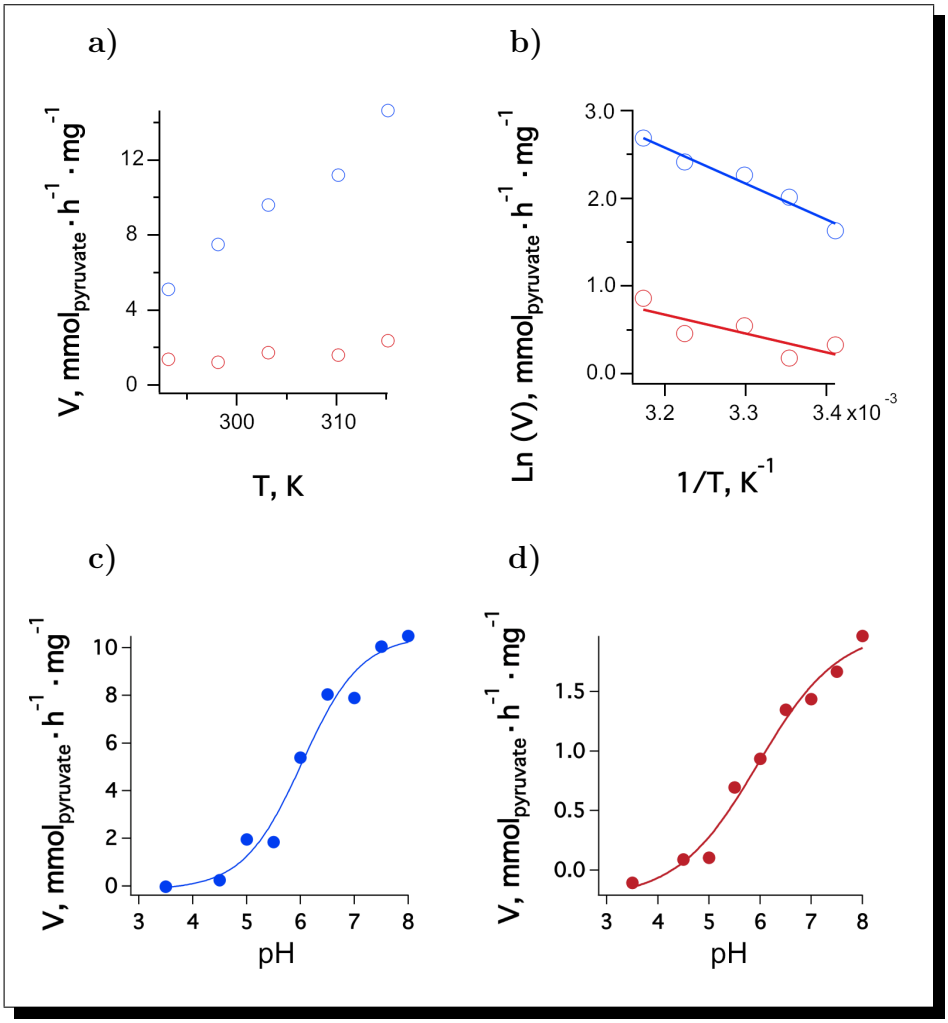


FIGURE 6.2.6. **Influence of the temperature (a and b) and pH (c and d) in the specific activity of T14D7.1 (blue) and AGT-Ma (red).**

**T14D7.1 protein is less stable than human AGT.** The stability of the T14D7.1 upon thermal unfolding has been characterized using DSC. As we proved for human AGT, thermograms obtained for the T14D7.1 are also well described using a two state irreversible model. Moreover, values of the activation energy obtained with the application of the consistency tests[121] are in good agreement. As shown in the previous chapter, the dependence of the $T_m$ parameter on scan rate and the independence on protein concentration support that the thermal unfolding of the holo T14D7.1 is a kinetically controlled process with first order kinetics.

In terms of **thermal stability**, T14D7.1 is less stable than human AGT. Although both proteins show similar $E_a$, the $T_m$ parameter of T14D7.1 is 12.3°C

lower than the human protein (see Table 6.2.4 and Figure 6.2.7-a). As a result, these parameters are translated into an important reduction of the kinetic stability for T14D7.1.

In terms of **kinetic stability**, expressed as a half-life extrapolated at 37°C, the nematode protein is 190-fold less stable than human protein (see Figure 6.2.7-b). However, if we compare the kinetic stability of human AGT at 37°C with the kinetic stability of T14D7.1 at 15 or 25°C (range of growth temperatures for *C. elegans*) we find that the protein of the nematode is from 2000-fold to 5-fold more stable, respectively (see Table 6.2.4).

In addition, there is a difference of about 200 kcal per mol of dimer between the **experimental enthalpy of unfolding** ($\Delta H$) for the T14D7.1 and the human AGT at their corresponding $T_m$ (see Table 6.2.4). However, if the enthalpy of unfolding is corrected by temperature and both values are evaluated at the denaturation temperature of human AGT-Ma, we obtain an enthalpy of unfolding for T14D7.1 of 504 kcal per mol of dimer. Now, this value does not differ largely of the AGT-Ma denaturation enthalpy (548 kcal per mol of dimer).

TABLE 6.2.4. **Energetic parameters of the thermal unfolding for T14D7.1 and human AGT-Ma.**

| *Variant* | $T_m$ [1] | $\Delta H$ [2] | $E_a$ [3] | $k_{37}$ ($k_{20}$) [4] |
|---|---|---|---|---|
| | $^oC$ | $kcal \cdot mol^{-1}$ | $kcal \cdot mol^{-1}$ | $min^{-1}$ |
| *C. elegans* T14D7.1 | 69.8 | 366 ± 11 | 112 ± 15 | $1.2{\cdot}10^{-7}$ ($7.1{\cdot}10^{-12}$) |
| *Human* AGT-Ma | 82.1 | 548 ± 4 | 109 ± 5 | $6.4{\cdot}10^{-10}$ |

[1] determined at scan rate of 180 °C/h.
[2] enthalpy of unfolding at $T_m$ expressed per mol of dimer.
[3] mean ± sd from the consistency test.
[4] kinetic constant rates for irreversible denaturation extrapolated to 37°C (and 20°C).

The experimental enthalpy of unfolding is also lower than the expected enthalpy for a protein of the size of T14D7.1. The total changes in accessible surface area ($\Delta ASA$), the number of residues ($N_{res}$), the heat capacity change upon unfolding ($\Delta C_p$) and the enthalpy of unfolding ($\Delta H_U$) are highly correlated[133] for a globular protein. The theoretical enthalpy of unfolding ($\Delta H_U$) and the heat capacity change upon unfolding ($\Delta C_p$) of T14D7.1 can be calculated with the use of empirical structure-energetic relationships as a function of the number of residues[133] (see Table 6.2.5).

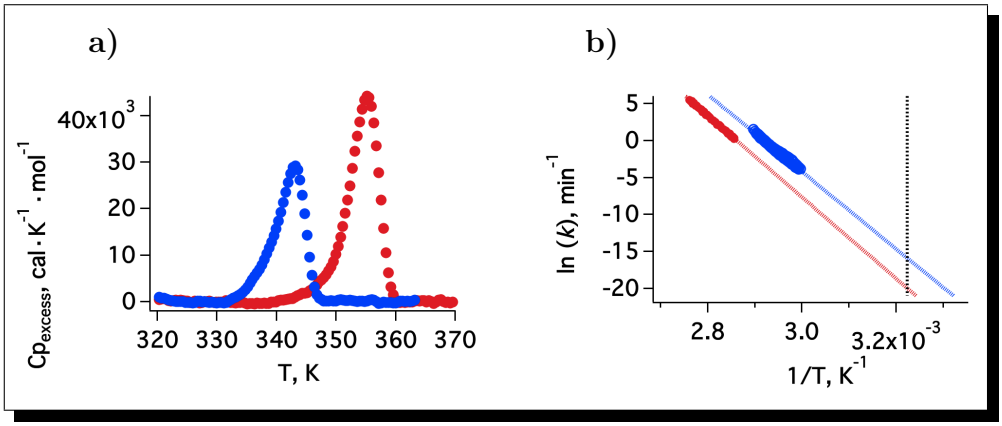


FIGURE 6.2.7. **Thermal and kinetic stability of T14D7.1 (blue) and human AGT-Ma (red). a)** thermograms for thermal unfolding obtained at scan rate of 180 K/h by DSC; **b)** Arrhenius plots where markers correspond to experimental data. Lines correspond to the fit of data and allow us to obtain the kinetic stability at the physiological temperature of 37°C (vertical dotted line).

TABLE 6.2.5. **Theoretical parameters of the unfolding of the T14D7.1 and human AGT as a function of the number of residues[133]**

| *Variant* | $\Delta C_p$ $kcal \cdot mol^{-1} \cdot K^{-1}$ | $\Delta H_U(T_m)$ [1] $kcal \cdot mol^{-1}$ |
|---|---|---|
| *C. elegans* T14D7.1 | 11.23 ± 0.19 | 675 (366) |
| *Human* AGT-Ma | 10.87 ± 0.19 | 787 (548) |

[1] Theoretical enthalpy of the unfolding evaluated at its corresponding $T_m$. Experimental values are expressed in parentheses

The **theoretical** value obtained for the denaturation enthalpy of T14D7 at its denaturation temperature ($T_m$) is 675 kcal per mol of dimer with a heat capacity change of 11.2 ± 0.2 $kcal/K \cdot mol$. The difference between theoretical and experimental enthalpy of unfolding suggests that thermal unfolding of T14D7.1 does not involve a complete loss of the native structure. This is a slightly different situation compared to AGT-Ma, which seems to involve the loss of more native structure (see Table 6.2.5).

We have also evaluated the behaviour of T14D7.1 during **urea induced unfolding**. The process was monitored by measuring the CD signal at 222 nm after overnight incubation of the samples at 25°C, in the presence of different urea concentrations, until sample equilibration was reached (see Figure 6.2.8).

To obtain an estimation of the urea unfolding resistance and the transition midpoint ($C_m$ value) a two-state equilibrium unfolding model was used.

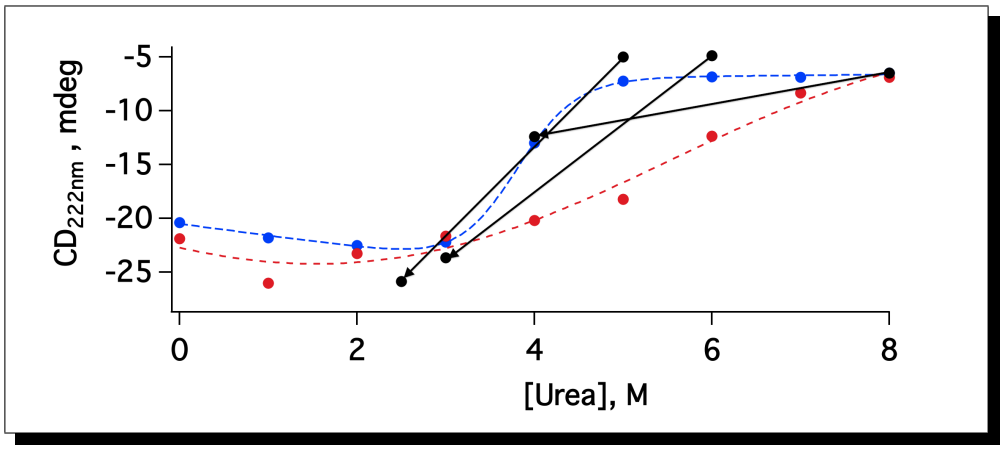


FIGURE 6.2.8. **Urea-induced unfolding equilibrium of T14D7.1 (blue) and human AGT-Ma (red).** The unfolding was monitored by CD at 222 nm. The dashed line respresent the fit of the data to a two-state model. The reversibility is shown by the black dots and arrows.

The human AGT-Ma unfolds irreversibly in the presence of urea showing a broad curve. Due to aggregation of AGT-Ma under some urea concentrations (intermediate values), the two-state reversible model does not hold because the equilibrium was not reached[123] (see Figure 6.2.8). In addition, the maximum urea concentration of 8 M appears to be insufficient to reach the post-transition region. Nonetheless, it has been published previously that urea induced unfolding of the AGT-Ma has a $C_m$ value of around 6.9 M.[123] Alternatively, T14D7.1 shows a narrow unfolding curve with well-defined pre- and post- transition baselines. The concentration of urea at the inflection point (the $C_m$ parameter) is 3.73 M, with a high value of the slope of the curve. Moreover, the urea induced unfolding of the T14D7.1 protein is reversible (see Figure 6.2.8). During the reversibility measurements the refolding of the protein was found to take place rapidly, where the native signal was recovered within the dead time of the experiments (few seconds).

## 6.3. Discussion

In this work we have confirmed that T14D7.1 is an alanine:glyoxylate aminotransferase protein, which resembles that of the peroxisomal human AGT. The T14D7.1 protein displays an increased *in vitro* activity (5-fold higher at 37°C)

and a lower kinetic stability (190-fold lower at 37°C) than human AGT. As human AGT, the substrate specificity of T14D7.1 is highly selective for L-alanine as amino donor and glyoxylate as amino acceptor (see Table 6.3.1).

TABLE 6.3.1. **Specific activity of T14D7.1 in comparison with AGT-Ma.**

| *Amino donor* | *Human AGT − Ma* | *C.elegans T14D7.1* |
|---|---|---|
| | *Activity* $mmol \cdot h^{-1} \cdot \mu g^{-1}$ | *Activity* $mmol \cdot h^{-1} \cdot \mu g^{-1}$ |
| L-alanine | 1.405 ± 0.043 | 6.405 ± 0.075 |
| L-serine | 0.061 ± 0.02 | 0.00465 ± 0.00045 |
| L-arginine | 0.027 ± 0.004 | N.D. |
| L-glutamate | 0.0082 ± 0.004 | N.D. |
| L-aspartate | N.D. | N.D. |
| L-phenylalanine | 0.0145 ± 0.005 | 0.01085 ± 0.00015 |

*N.D., not detectable. Data were obtained by Dr. Barbara Cellini and Dr. Elisa Oppici at the department of Life and Sciences and Reproduction, Section of Biological Chemistry at University of Verona.*

While the physiological temperature of humans is kept constant at 37°C, the nematode *C. elegans* is an ecthotherm* animal that can survive between 8-27°C and whose physiology is highly affected by the environmental temperature. The low denaturation temperature of T14D7.1 (12.3°C lower than human AGT-Ma) could be explained as a temperature adaptation due to a correlation between protein thermal/kinetic stability and environmental/physiological temperature.[134]

In order to be a good regulator of metabolism, enzymes must change the reaction velocity in a substrate concentration range according to its $K_M$ value. It is usually found that $K_M$ values of orthologous enzymes from organisms adapted to different temperatures are strongly conserved (assuming inter-species similarities in substrate concentration or $K_M : [substrate]$ ratio).[135] In this case we have found that both enzymes display similar $K_M$ values for substrates. Moreover, if we compare the activity of T14D7.1 and human AGT-Ma at its physiological temperature (i.e., 20°C and 37°C respectively), T14D7.1 activity

*Animal whose physiological temperature tracks ambient temperature

is only 2-fold higher than AGT-Ma. Accordingly, it has been found that cold adaptation is usually accompanied by increases in the activities of enzymes involved in major pathways of metabolism.[136]

T14D7.1 is expressed from embryo to adult stage mainly in the intestinal and body wall muscle cells[99] but there is no information about its intracellular location. Nonetheless, there are two evidences that would discard a peroxisomal location: i) T14D7.1 lacks the consensus sequence of the PTS1 pathway in the C-terminal tripeptide that is conserved between humans and nematodes[137] where the protein C34C6.6 is the orthologous for human Pex5p in the nematode;[94] ii) the *C. elegans* genome lacks genes encoding protein specific for the PTS2 pathway to target proteins to peroxisomes, and proteins that are imported to peroxisome by the PTS2 pathway in other organisms (such as humans) contain a PTS1 sequence in *C. elegans*.[138] On the other side, mitochondria appear to be a plausible location for T14D7.1 due to the extra residues at its N-terminal domain. Although some mitochondrial proteins contain an internal targeting signal, the information required for translocation to mitochondria is usually presented as a cleavable sequence at the N-terminal domain (mitochondria targeting sequence or MTS). There is no consensus MTS but it usually consists of 10-80 residues that prone to form amphipathic helices. Bioinformatic tools such as MitoProt (http://ihg.gsf.de/ihg/mitoprot.html) or TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) suggest that the N-terminal domain of T14D7.1 could form a cleavable amphipathic helix with a probability of 87% and 85% respectively. In collaboration with Dr. Ana Calvo (Department of Cell Biology, Yale University, CT, USA) the intracellular location of T14D7.1 protein is being determined and preliminary results support the hypothesis of a mitochondrial location.

While in humans glyoxylate must be eliminated by the AGT protein, nematodes contain an active glyoxylate cycle[139] where *isocitrate lyase* (EC 4.1.3.1) and *malate synthase* (EC 4.1.3.2) activities allow to bypass the decarboxylation steps of the TCA cycle. The glyoxylate cycle is regulated in a developmentally specific manner[140,141] and is performed by a single bi-functional protein (encoded by *icl-1* gene) with different activities in different domains (isocitrate lyase in N-terminal and malate synthase in C-terminal) that is located in intestinal and body wall muscle cells.[142] Therefore, in nematodes, glyoxylate is

a key metabolite that can be used to produce energy through TCA cycle or to create a net production of carbohydrates through gluconeogenesis.[143]

Although the T14D7.1 protein may be associated to the inter-conversion of amino acids, the biological role of T14D7.1 remains unknown. There are numerous strains reported in the WormBase with mutated alleles on the *T14D7.1* gene but there are no phenotypes associated. However, in collaboration with Dr. Ana Calvo, we are currently characterizing the phenotype of a strain with a knockout *T14D7.1* gene (*allele tm6307* from National Bioresource Project for the Experimental Animal «Nematode *C. elegans*»). The bi-functional glyoxylate cycle protein of *C. elegans* is involved in the intermediary metabolism of the dauer larva (arrested developmental stage of *C. elegans* to survive harsh conditions)[144] and it is essential for embryonic morphogenesis and influence lifespan.[145] According to this, in order to create a hyperoxaluric phenotype in the nematode, at least the gene expression of T14D7.1 and the bi-functional glyoxylate cycle protein should be inhibited by RNAi. In this scenario glyoxylate is expected to be oxidized to oxalate that can lead to the formation of CaOx stones. Nonetheless, the key role of the glyoxylate cycle protein in the normal development of the nematode may represent a challenge in the creation of a viable hyperoxaluric model in *C. elegans*.

In any case, the nematode *C. elegans* could still be a useful tool to study the interactions of human AGT protein during the folding process. Therefore, the creation of a test bench model is an ongoing project.

CHAPTER 7

# Engineering an improved human AGT enzyme

## 7.1. Brief introduction

An alternative therapeutic approach to the rescue of PH1 disease-causing variants, is the replacement of the deficient protein or gene. Enzyme replacement therapy (ERT) and gene therapy (GT) are two potential treatments for monogenic inborn errors of liver metabolism such as PH1.[146]

The liver is the center of metabolism for proteins, lipids and glucose and therefore it is a target for treatment of a wide variety of diseases. Due to the stable environment of hepatocytes, enriched vasculature, great regenerative capacity and the central role in drug metabolism, the liver is one of the most appropriate organs for gene therapy.[147] Nonetheless, gene delivery is an essential first step that is crucial for the success of GT.[148] In addition, GT requires the transduction of a large percentage of the hepatocytes of patients with a wild type copy of the gene,[149] thus the modification of the coding sequence delivered with the GT vector could be an attractive strategy to improve its efficiency. Similarly, the use of an engineered protein with higher stability and activity could also enhance the success in ERT.[16]

The consensus approach is based on statistical information contained in sequence alignments of proteins with moderate to high homology and it is a simple strategy to improve protein stability.[150] The procedure of how this enhanced enzyme has been engineered and its biochemical and biophysical characterization are presented and discussed in this chapter.

## 7.2. Results

**Eight single point mutations were selected by the consensus-based approach.** The consensus amino acid for each position in human AGT protein sequence was obtained (see Figure 7.2.1) by the comparison of human AGT-Ma with twenty-one sequences of AGT orthologous from vertebrates (see Figure 7.2.2). According to the consensus approach, the ratio of frequencies between the most common amino acid found at a certain position in the multiple alignment and the frequency of the amino acid found in the human sequence was evaluated.[116] Those mutations with ratios higher than 1 should increase protein stability. Moreover, the selected mutations were also found in a wider sequence alignment using human AGT and eighty-nine sequences from eukaryotes (see Appendix B).

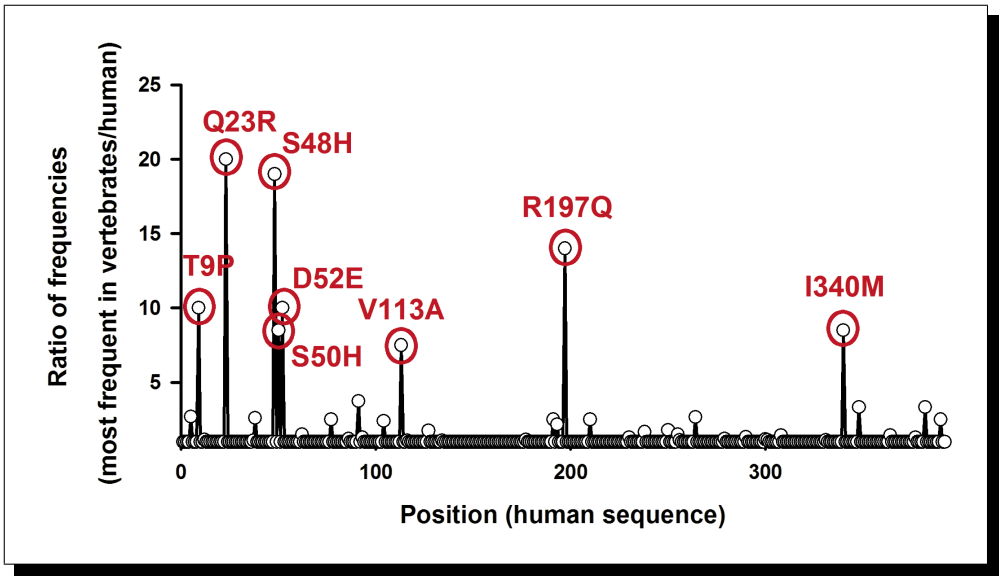


FIGURE 7.2.1. **Ratio of frequencies according to the sequence alignment of human AGT and 21 sequences of AGT orthologous from vertebrates.** The consensus-based mutations are marked in red.

Eight mutations were selected from the sequence alignment. Nonetheless, only six of these single mutations were studied. Mutations T9P and R197Q could not be expressed at proper yields under different conditions and therefore these mutations were taken out from our analysis. As a result, only six single mutations were selected to be expressed under the background of the human major allele (that encodes for AGT-Ma protein).

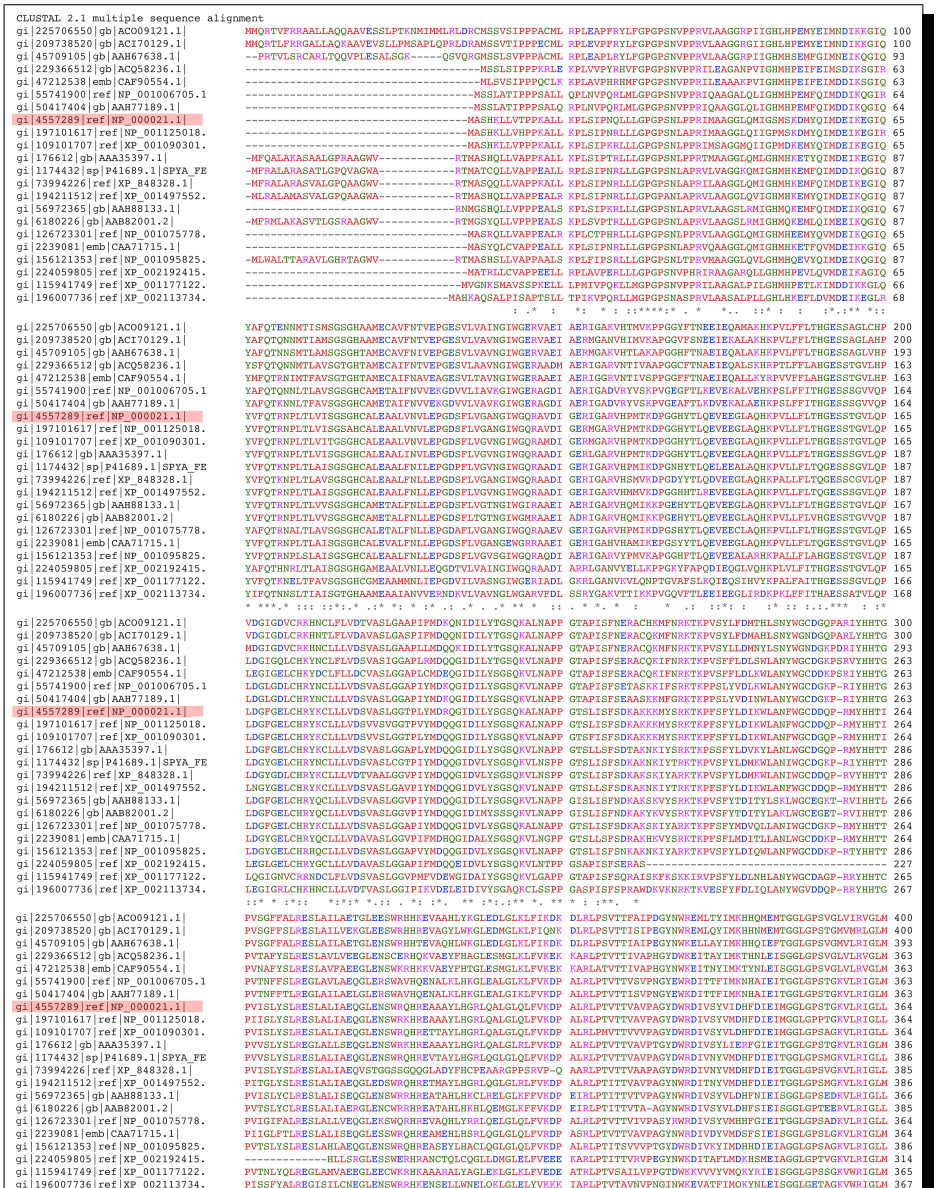


FIGURE 7.2.2. **Sequence alignment of AGT proteins from vertebrates (Clustal Omega).** GI codes from top to botton are 225706550, 209738520, 45709105, 229366512, 47212538, 55741900, 50417404, 4557289 (human AGT is highlighted), 197101617, 109101707, 176612, 1174432, 73994226, 194211512, 56972365, 6180226, 126723301, 2239081, 156121353, 224059805, 115941749 and 196007736.

These six substitutions were mainly located in the first third of the primary sequence of human AGT: Q23R, S48H, S50H, D52E, V113A and I340M (see Figure 7.2.1). One of these mutations involves a change from a neutral to a basic residue (Q23R) in which a larger side chain is also included as in S48H and S50H mutations. The other three substitutions (D52E, V113A and I340M) are more conservative. The D52E substitution keeps the negative charge of the side chain whereas the V113A and I340M substitutions keep the hydrophobic nature. Interestingly the mutation I340M obtained by the consensus approach is actually one of the polymorphisms that form the minor allele, together with the P11L substitution (see Appendix A).

According to the three dimensional structure of AGT, residues Gln23, Ser48, Ser50, Asp52 and Ile340 are clustered and located close to the dimerization interface while the residue Val113 is far from this cluster and partially exposed to the solvent (see Figure 7.2.3).

In this document, the six single consensus-based variants under the major allele background are named as Q23R-Ma, S48H-Ma, S50H-Ma, D52E-Ma, V113A-Ma and I340-Ma.

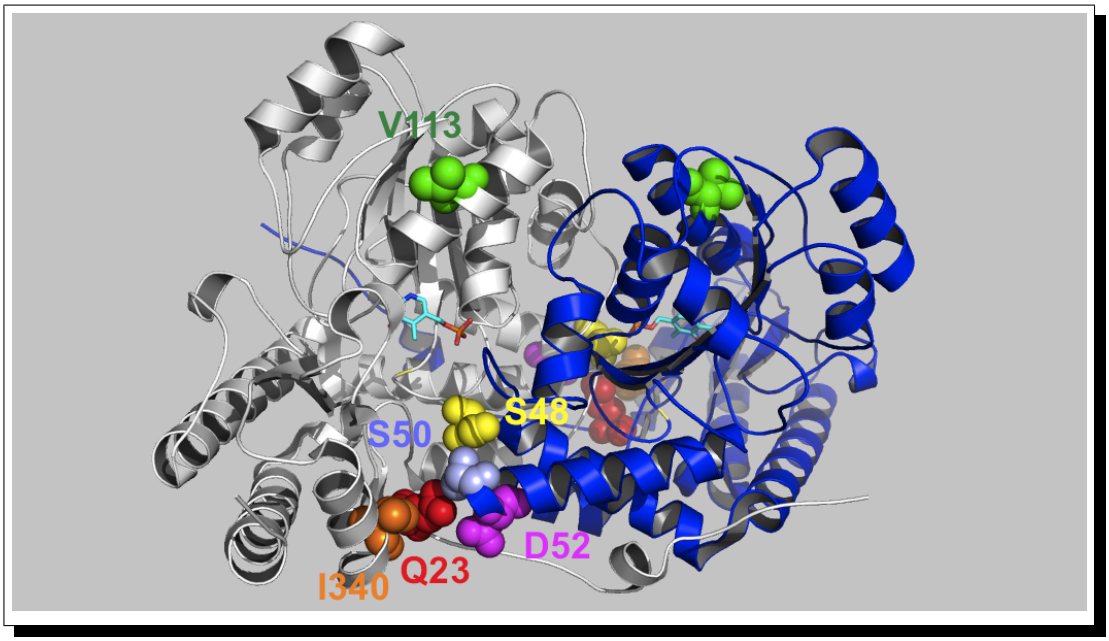


FIGURE 7.2.3. **Consensus-based mutations over the AGT structure.** Five consensus-based positions are clustered close to the dimerization interface: residues Ser48, Ser50 and Asp52 from one monomer and residues Gln23 and Ile340 from the other monomer. *Image was created with Pymol Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.*

**The consensus-based approach improves the *in vitro* stability of the human AGT enzyme.** The thermal unfolding of the six single point mutations was studied by DSC. Like the human AGT-Ma, the thermal unfolding of these consensus-based variants was found to be irreversible and scan rate dependent. Thermal denaturation followed a two state kinetic model with first order kinetics, where the first order of the thermal unfolding was also tested ($\mu$).[122] The energetic data obtained from thermal unfolding of these single mutants are summarized in Table 7.2.1.

Five of the six consensus-based variants show an enhanced thermal stability, with a higher $T_m$ parameter (≈2-3°C) than AGT-Ma (see Figure 7.2.4-a). Only the substitution S50H was found to be destabilizing with a reduction in the $T_m$ of 15°C.

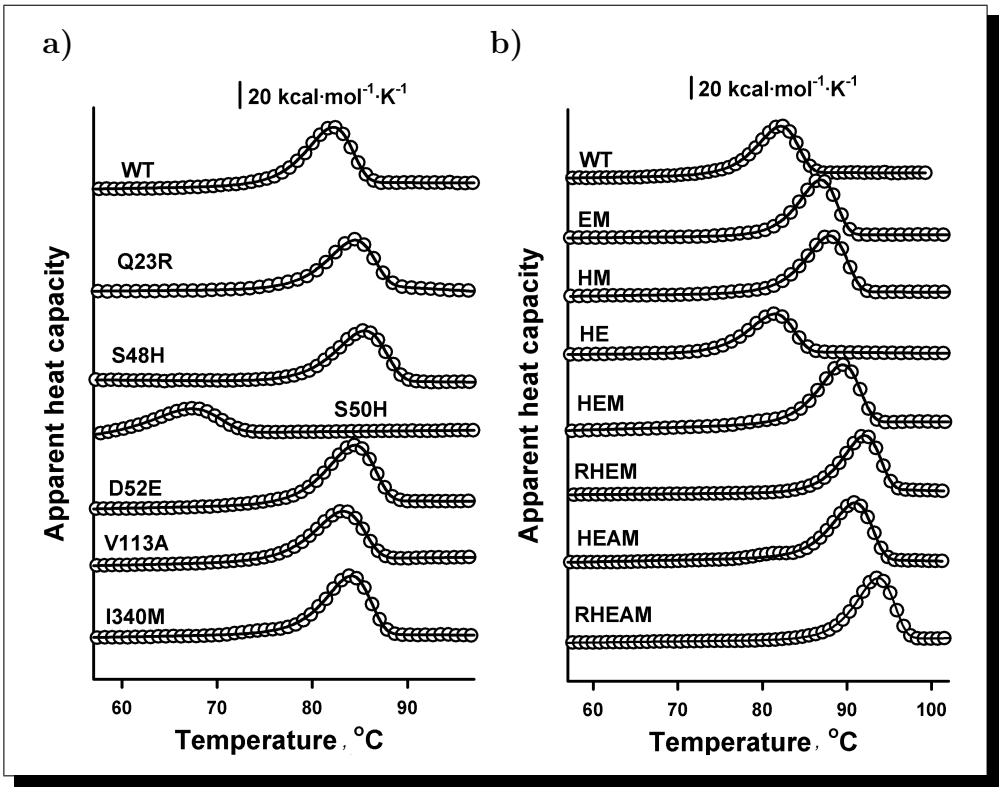


FIGURE 7.2.4. **Thermograms of the thermal denaturation of AGT-Ma and consensus-based variants as holo proteins for a) single mutants and b) multiple mutants.** Markers are the experimental data obtained at scan rate of 180°C/h, while lines are fits to a two state irreversible model.

TABLE 7.2.1. **Thermal and kinetic stability of consensus-based variants compared to AGT-Ma.**

| *Variant* | $T_m$ [1] $^oC$ | $E_a$ [2] $kcal \cdot mol^{-1}$ | $k_{37}$ $min^{-1}$ | $\mu$ |
|---|---|---|---|---|
| AGT-Ma | 82.1 | 109 ± 5 | $6.4{\cdot}10^{-10}$ | 0.95 ± 0.02 |
| Q23R-Ma | 84.5 | 104 ± 3 | $1.1{\cdot}10^{-10}$ | 0.91 ± 0.02 |
| S48H-Ma | 85.5 | 67 ± 2 | $2.3{\cdot}10^{-10}$ | 0.94 ± 0.02 |
| S50H-Ma | 67.1 | 104 ± 6 | $6.8{\cdot}10^{-5}$ | 0.99 ± 0.02 |
| D52E-Ma | 84.4 | 101 ± 1 | $1.3{\cdot}10^{-10}$ | 0.93 ± 0.01 |
| V113A-Ma | 83.2 | 116 ± 9 | $2.0{\cdot}10^{-9}$ | 0.97 ± 0.01 |
| I340M-Ma | 84.1 | 101 ± 1 | $1.7{\cdot}10^{-11}$ | 0.94 ± 0.02 |
| HM-Ma | 88.0 | 121 ± 8 | $1.1{\cdot}10^{-11}$ | 0.93 ± 0.02 |
| HE-Ma | 81.5 | 118 ± 10 | $7.2{\cdot}10^{-10}$ | 0.90 ± 0.01 |
| EM-Ma | 86.9 | 122 ± 5 | $6.4{\cdot}10^{-12}$ | 0.94 ± 0.02 |
| HEM-Ma | 89.5 | 125 ± 5 | $4.3{\cdot}10^{-12}$ | 0.95 ± 0.04 |
| HEAM-Ma | 91.0 | 125 ±6 | $9.5{\cdot}10^{-14}$ | 0.92 ± 0.07 |
| RHEM-Ma | 92.0 | 112 ± 5 | $1.7{\cdot}10^{-13}$ | 0.87 ± 0.03 |
| RHEAM-Ma | 93.7 | 104 ± 3 | $1.1{\cdot}10^{-13}$ | 0.92 ± 0.04 |

[1] $T_m$ values obtained at scan rate of 180°C/h.
[2] $E_a$ values are mean±s.d. from the 4 different tests proposed[121]
[3] Obtained from Arrhenius plots
[4] $\mu$ values obtained from fittings to a non first-order kinetic model[122]

**The stabilizing effect of the single consensus-based mutations is additive.** In order to evaluate the stabilizing effects of these five single point mutations, we combined them to create double (HM-Ma, HE-Ma and EM-Ma), triple (HEM-Ma), quadruple (RHEM-Ma and HEAM-Ma) and quintuple (RHEAM-Ma)* mutants (see Table 7.2.1).

The characterization of these new variants shows that the stabilizing effects is additive (see Figure 7.2.4-b) with an increase in the thermal stability (up to

*The multiple mutants are named with the letters of new mutations introduced. For example the double mutant of S48H and I340M is termed as HM-Ma

12°C in the quintuple mutant) that is translated in a higher kinetic stability extrapolated at physiological temperature (see Table 7.2.1). The kinetic stability is increased in the quadruple and quintuple mutants in 3-4 orders of magnitude compared to AGT-Ma. These results are coherent with the existence of a high correlation between thermal stability (higher $T_m$) and the kinetic stability at physiological temperature (lower $k_{37}$) (see Figure 7.2.5-a and Table 7.2.1).

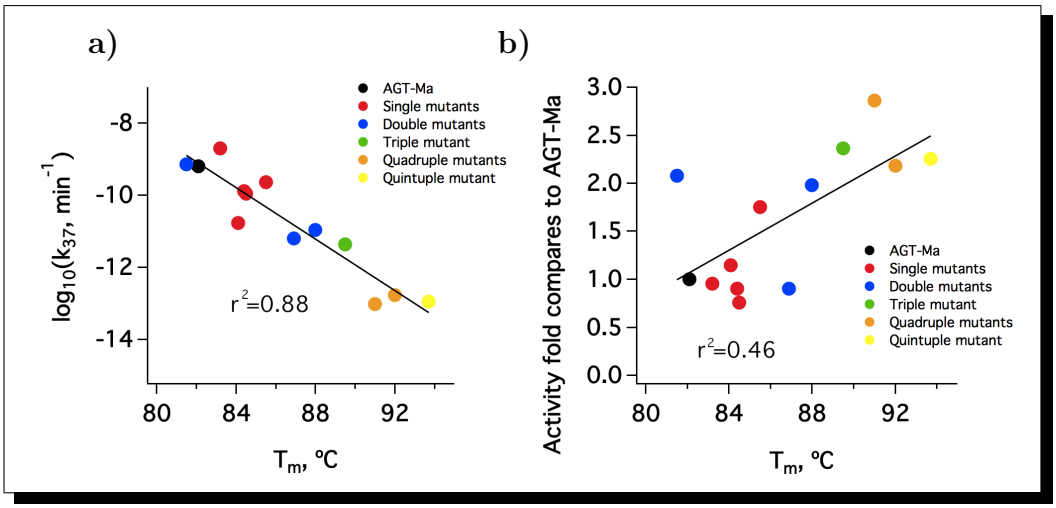


FIGURE 7.2.5. **Correlation between a) kinetic stability and b) enzyme activity with the thermal stability ($T_m$) for the consensus-based variants as a comparison with AGT-Ma.** Lines correspond to a linear fit.

**Consensus-based mutations also increase the activity of the human AGT enzyme.** We have measured the specific activity of the consensus-based variants. This activity was measured using L-alanine and glyoxylate under saturation conditions for both substrates. The results obtained are shown in Figure 7.2.6.

As we can observe, three single variants (Q23R-Ma, D52E-Ma and V113A-Ma) show similar activity than AGT-Ma, while only one substitution (S50H-Ma) causes a 2.5-fold lower activity. There are also two single consensus-based variants (S48H-Ma and I340M-Ma) that have increased the activity (75% and 14% respectively) compared to AGT-Ma. According to this, the multiple mutants show a high specific activity mainly due to the presence of the S48H substitution (see Figure 7.2.6) while there is a weak correlation between thermal stability and activity (see Figure 7.2.5-b).

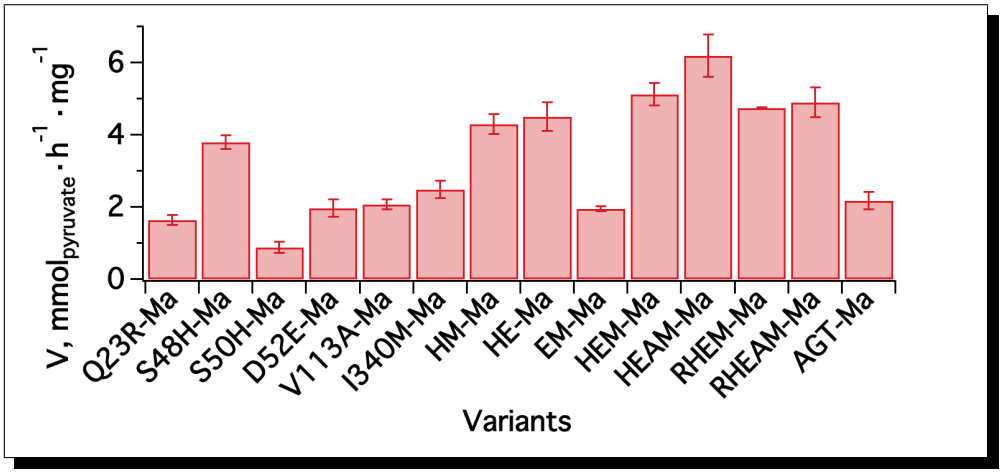


FIGURE 7.2.6. **Specific activity for the consensus-based variants compared to AGT-Ma.** Data from 3 different independent measurements.

**Functional characterization of the highly stable and active quintuple mutant.** According to the previous results, we have studied the functional properties of RHEAM-Ma variant with more detail, in comparison with AGT-Ma (see Table 7.2.2):

*Thermal stability.* The increase of the thermal stability that has been shown for the holo RHEAM-Ma in comparison with the holo AGT-Ma, is also kept in the apo forms, and both proteins have a reduction of the thermal stability through the release of the cofactor (see Table 7.2.2-a and Figure 7.2.7-a). Moreover, the kinetic stability at physiological temperature is also increased in the apo state of the quintuple mutant (see Table 7.2.2-a and Figure 7.2.7-b).

*Enzyme activity.* The consensus-based RHEAM-Ma variant has shown an increased activity of around 125% compared to human AGT-Ma (see Table 7.2.2-b and Figure 7.2.7-c). A complete analysis of the influence of substrates concentration (L-alanine and glyoxylate) on enzyme activity, reveals that the quintuple mutant has a slightly lower $K_M$ constant for the L-alanine while the $K_M$ constant for glyoxylate is 1.6-fold higher (see Table 7.2.2-b).

*Microenvironment of the cofactor.* The microenvironment of the cofactor in the active site was studied by spectroscopic techniques (see Figure 7.2.7-d and -e). When the coenzyme is located in the active site as PLP, both proteins show the same strong absorption and dichroic signal around 420 nm. In addition, both proteins have the same spectroscopic behaviour when the coenzyme is forming the external Schiff base as PMP (peak centered at 320 nm and lack

of signal at 420 nm). This result suggests a similar microenvironment of the active site.

TABLE 7.2.2. **Thermal denaturation and functional properties of RHEAM-Ma variant as a comparison with AGT-Ma.**

*a) Thermal and kinetic parameter of the thermal unfolding*

| *Variant* | HOLOENZYME | | APOENZYME | |
|---|---|---|---|---|
| | AGT-Ma | RHEAM-Ma | AGT-Ma | RHEAM-Ma |
| $T_m$ [1], $^oC$ | 82.1 | 93.7 | 58.4 | 70.2 |
| $\Delta H$ [2], $kcal \cdot mol^{-1}$ | 548 ± 4 | 638 ± 32 | 254 ± 16 | 392 ± 20 |
| $E_a$ [2], $kcal \cdot mol^{-1}$ | 109 ± 5 | 125 ± 6 | 111 ± 10 | 116 ± 7 |
| $\mu$ [2] | 0.95 ± 0.02 | 0.91 ± 0.04 | 1.30 ± 0.02 | 0.89 ± 0.01 |
| $k_{37}$ [3], $min^{-1}$ | $6.4{\cdot}10^{-10}$ | $1.1{\cdot}10^{-13}$ | $1.6{\cdot}10^{-5}$ | $3.6{\cdot}10^{-8}$ |
| $\frac{k_{AGT-Ma}}{k_{RHEAM-Ma}}$ | 1 | 2100 | 1 | 440 |

[1] $T_m$ for a scan rate of 3°C/min
[2] mean±s.d. of AGT dimer from 3 scan rates
[3] Kinetic constant rates extrapolated at 37°C

*b) Functional properties of the enzyme*

| *Variant* | AGT-Ma | RHEAM-Ma |
|---|---|---|
| $V_{max}$, $m\,mol \cdot h^{-1} \cdot mg^{-1}$ | 2.22 ± 0.09 | 4.99 ± 0.22 |
| $K_{M,Ala}$, $mM$ | 19.5 ± 1.4 | 11.5 ± 1.2 |
| $K_{M,Glyoc}$, $\mu M$ | 245 ± 29 | 412 ± 45 |
| $K_{D,PLP}$, $nM$ | 99 ± 20 | 101 ± 18 |

*PLP affinity.* The coenzyme affinity was also tested for RHEAM-Ma and AGT-Ma. The affinity was measured by monitoring the quenching of the intrinsic Trp fluorescence through the binding of PLP in an equilibrium approach. The results obtained show that the quintuple mutant is able to bind the coenzyme in a similar way than AGT-Ma with almost identical equilibrium constants ($K_{D,PLP}$) (see Table 7.2.2-b and Figure 7.2.7-f).

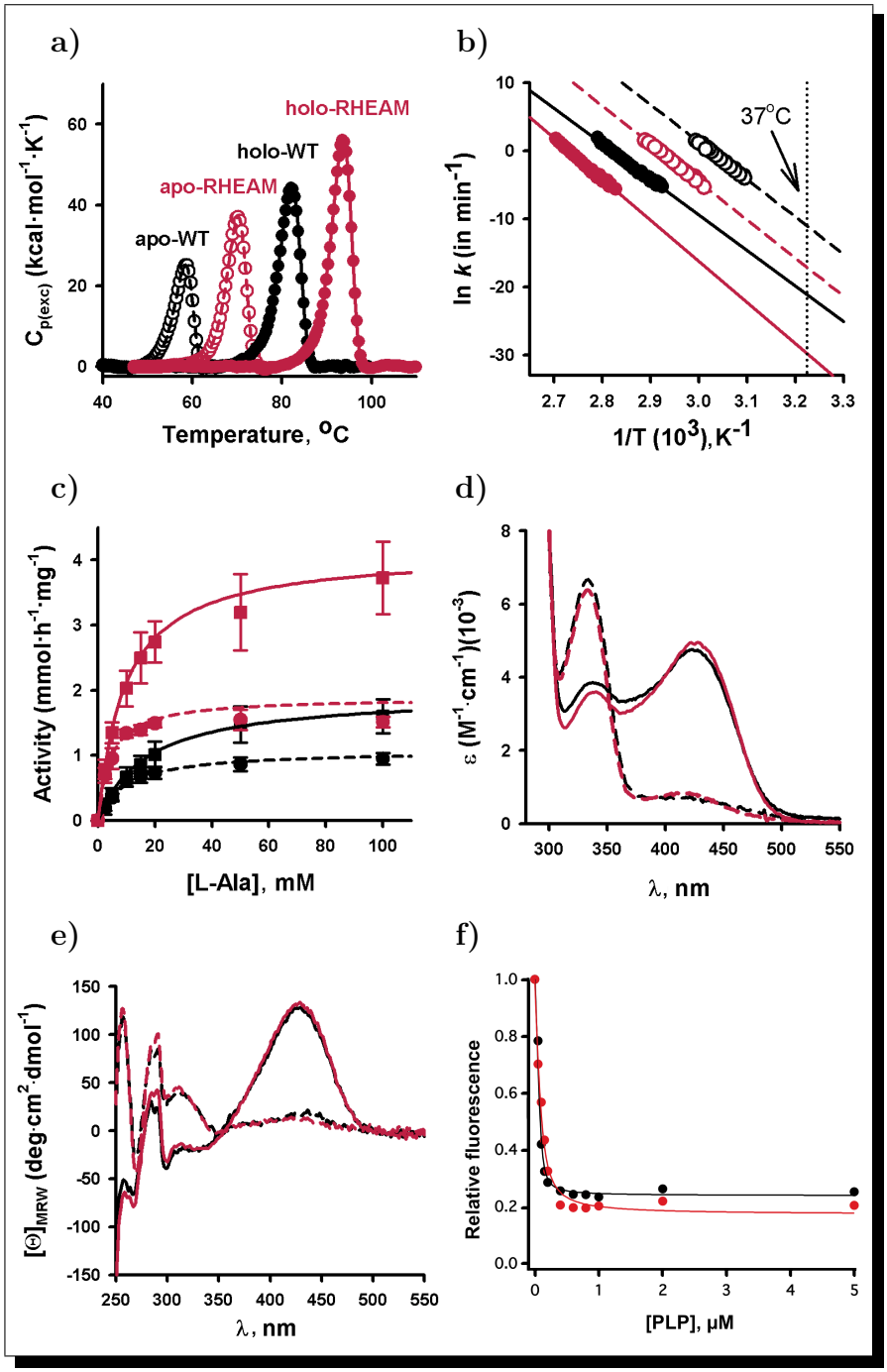


FIGURE 7.2.7. **Comparative stability and functional properties of the consensus-based variant RHEAM-Ma (red) and human AGT-Ma (black).** **a)** thermograms at 3°C/min; **b)** Arrhenius plots for the thermal unfolding of the holo (closed symbols) and apo (open symbols) forms; **c)** enzyme activity dependence on L-alanine at 0.25 mM (circles) and 2 mM (squares) glyoxylate, were lines are the global fitting; **d)** absorption and **e)** circular dichroism spectra of the holo proteins incubated with (dashed lines) or without (continuous lines) L-alanine; and **f)** titration with PLP monitored by fluorescence.

In addition, the kinetic approach was also used to evaluate the dynamics of this process (see Table 7.2.3). Values obtained for the rate constants of association ($k_{on}$) and dissociation ($k_{off}$) of the coenzyme for RHEAM-Ma are ∼3.5-fold lower than those found for AGT-Ma. This result indicates that, although the RHEAM-Ma variant has a similar equilibrium constant than AGT-Ma, the process of binding and release of the coenzyme is slower in RHEAM-Ma.

TABLE 7.2.3. **Kinetic parameters of the PLP affinity for the RHEAM-Ma and AGT-Ma variants.**

| *Variant* | AGT-Ma | RHEAM-Ma |
|---|---|---|
| $k_{on}$, $M^{-1} \cdot s^{-1}$ | 189 ± 4 | 65 ± 1 |
| $k_{off}$, $s^{-1}$ | 23 ± 5 ($\cdot 10^{-5}$) | 5.5 ± 2.5 ($\cdot 10^{-5}$) |

Data were obtained by kinetic measurements

**Electrostatic interactions are involved in the RHEAM-Ma stabilization.** As noted before, according to the structure of human AGT, four of the five consensus-based mutations are located in a cluster at the dimerization interface. The introduction of charges through these mutations may have implications in the kinetic stabilization of AGT enzyme, so the electrostatic contributions were estimated in AGT-Ma, HEM-Ma and RHEAM-Ma variants. This estimation was done using a simple Tanford-Kirwood and Bashford-Karplus (TK-BK) algorithm[151] and the dimeric structural information (atomic coordinates) of those variants[†]. The result of this analysis is given as the energy of charge-charge interactions ($E_{q-q}$) of a given ionizable residue with the rest of the ionizable residues of the dimer.

The comparison of the energy arising from charge-charge interactions for these variants compared to the AGT-Ma ($\Delta E_{q-q}$, see Figure 7.2.8), shows that two mutations (Q23R and D52E) are responsible for the development of new favourable charge-charge interactions (negative $\Delta E_{q-q}$values). Moreover, this analysis suggests that there are four additional ionizable residues (Arg175, Arg197, Arg333 and Asp344) that have created new favourable interactions through the introduction of the consensus-based residues (negative $E_{q-q}$ values, see inset at Figure 7.2.8).

[†]Through the collaboration on this project, the crystallographic structures of I340M-Ma, HEM-Ma and RHEAM-Ma were determined by Dr. Armando Albert at Department of Crystallography and Structural Biology, IQF-Rocasolano, CSIC.

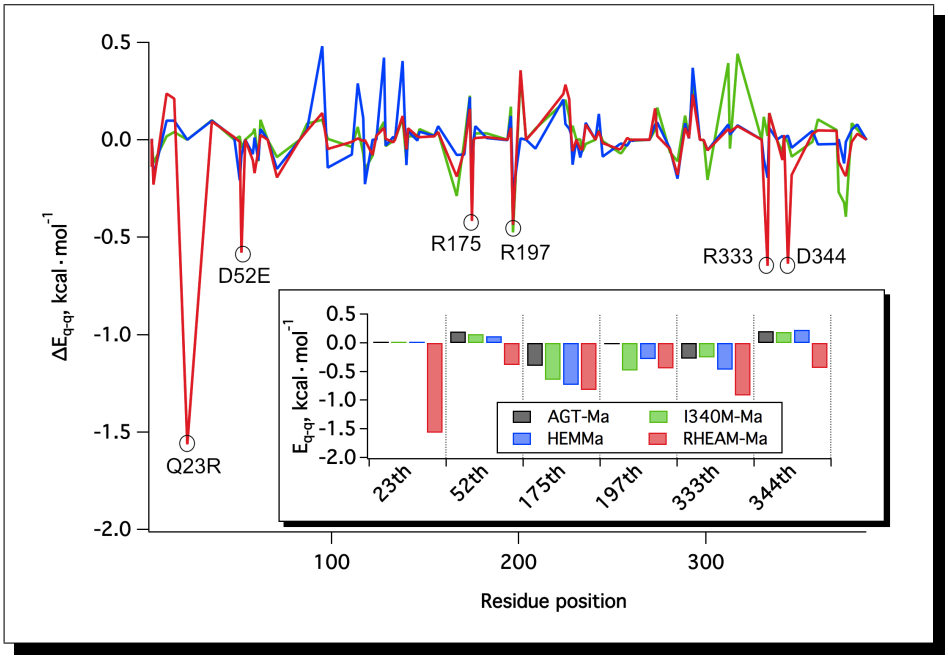


FIGURE 7.2.8. **Electrostatic contributions to the stability I340M-Ma, HEM-Ma and RHEAM-Ma evaluated using a TK-BK model.**[152] The graph shows the differences in the charge-charge interactions ($\Delta E_{q-q}$) of a given ionizable residue for I340M-Ma (green), HEM-Ma (blue) and RHEAM-Ma (red) with regard to AGT-Ma. The inset shows the charge-charge interactions ($E_{q-q}$) at selected positions. Same color code (where black corresponds to AGT-Ma). Negative values indicate that the ionizable groups are mainly involved in stabilizing interactions with groups of the opposite charge.

## 7.3. Discussion

In this work, we have successfully applied the consensus approach to improve a protein of biomedical interest, such as the human AGT. The improved enzyme was obtained through five single point mutations (RHEAM-Ma) and it displays a high kinetic stability (∼1000-fold longer half life than AGT-Ma at 37°C) and an improved catalytic performance (∼125% higher than AGT-Ma).

A low protein stability can lead to degradation, aggregation and lack of biological function *in vivo*[153] and thus stability is an important regulatory trait for proteins. We show that consensus-based mutations target the stability of native AGT dimer (which is the cooperative unit) determining the kinetic stability *in vitro* (i.e., denaturation follows first order kinetics) partly by a remarkable optimization of the electrostatic interactions in the native state. Moreover, this enhanced *in vitro* stability could be also translated *in vivo*[154].

It has been proposed recently that the stability effect of consensus mutations may reflect to some extent the extreme properties of ancestral proteins.[155] Although reconstruction of the ancestral sequence leads to larger protein stability enhancements, the *promiscuity* of the protein is also enhanced by this «travelling back in time».[156] Therefore, this lack of protein specificity may hinder the use of this approach for targeting specific metabolic pathways. By contrast, the consensus approach appears to be a simple and excellent choice to develop robust proteins for GT and ERT with wild type specificity, provided that these changes do not trigger significant immune response.

The consensus approach is an useful tool for different biotechnology fields.[119,157] During the collaboration of this project, the structure of I340M-Ma, HEM-Ma and RHEAM-Ma variants have been obtained and an improved success of the crystallization process and protein solubility have been observed (see Figure 7.3.1-a and Appendix B). Different physical properties of proteins can control the crystallization behaviour and different protein engineering approaches have been used in protein crystallography (such as high purity protocols, enhancing protein solubility,[158,159] engineering crystal contacts[160] or by surface entropy reduction[161]). While it has been proposed that the overall thermodynamic stability is not a major determinant of crystallization propensity,[162] the consensus approach has been also applied successfully to improve the thermostability and crystallizability of the kinase domain of the plant salt-overly-sensitive 2 (CIPK24/SOS2) protein by Dr. Armando Albert (see Appendix B). In addition, the structures of I340M-Ma, HEM-Ma and RHEAM-Ma (determined during this project) and the previously reported structure for AGT-Ma[50] are nearly identical with changes confined at the point mutations (see Figure 7.3.1-b). Therefore, the consensus approach seems to be a simple and generic strategy to increase the success in obtaining high-quality protein crystals for high-resolution structural determination.

The design of an enhanced human AGT protein might be useful to improve the efficiency of GT and ERT for the treatment of PH1 by reducing the dose and/or frequency of vector/protein administration.[16] The *in vivo* stability and activity of the consensus-based RHEAM-Ma was tested upon transiently transfection in CHO cells. The RHEAM-Ma variant was found to be correctly imported to peroxisomes by immunofluorescence confocal microscopy (see Figure

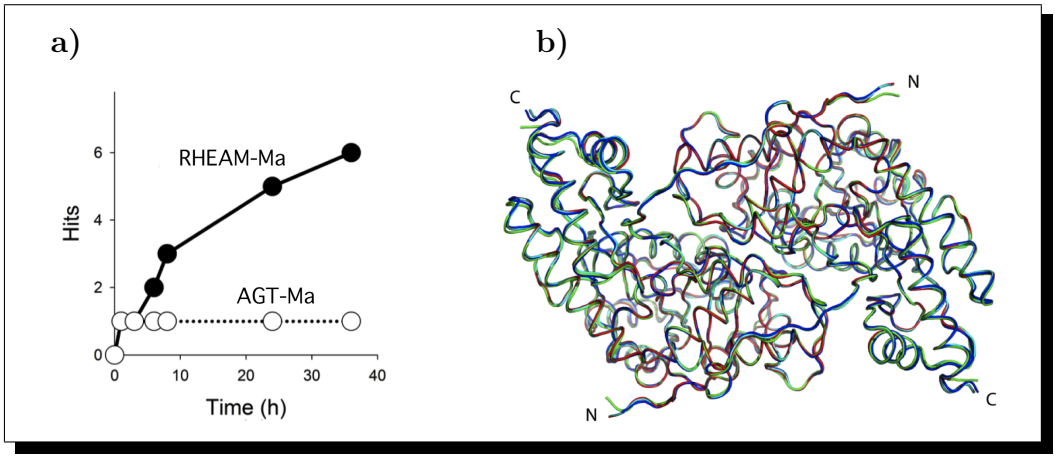


FIGURE 7.3.1. **Application of the consensus approach for crystallization studies. a)** crystallization propensity (number of hits) for RHEAM-Ma in a comparison with AGT-Ma under a variety of conditions; **b)** structural superimposition of the AGT-Ma (green), I340M-Ma (red), HEM-Ma (blue) and RHEAM-Ma (cyan) AGT crystal structures; *Data obtained by Dr. Armando Albert at Department of Crystallography and Structural Biology, IQF-Rocasolano, CSIC.*

7.3.2-a). However, the enhanced *in vitro* stability of RHEAM-Ma was not translated to higher resistance toward degradation or increased steady-state levels or solubility inside eukaryotic cells compared to AGT-Ma (see Figure 7.3.2-b and -c). A possible explanation could be that the kinetically relevant state *in vitro* might differ from the state sensitive to cytosolic and/or peroxisomal degradation by the corresponding quality control systems such as proteasome ATP-dependent mechanism in cytosol[60] or through the activity of a Lon protease in peroxisomes.[163] Alternatively, the conformational stability of AGT-Ma may define an upper threshold from which an enhanced stability against thermal unfolding is not translated in higher resistance toward *in vivo* degradation. That said, the enhanced stability of RHEAM-Ma could lead to higher transduction efficiency upon GT and/or longer half-lives in ERT. Accordingly, the identification of possible advantages of RHEAM-Ma variant in GT and ERT is an ongoing project in collaboration with Prof. Eduardo Salido (University of La laguna, Spain) and Dr. Barbara Cellini (University of Verona, Italy).

In summary, by using the consensus approach we have optimized interactions in the AGT native state increasing the kinetic free energy barrier for protein denaturation while the overall protein fold and enzyme function have been kept or improved. Although the possible advantages of using an enhanced AGT enzyme in GT and/or ERT remain unknown (ongoing project), we have proved

that consensus approach is a simple and generic tool to increase stability and solubility improving the crystallization of hard-to-crystallize proteins.

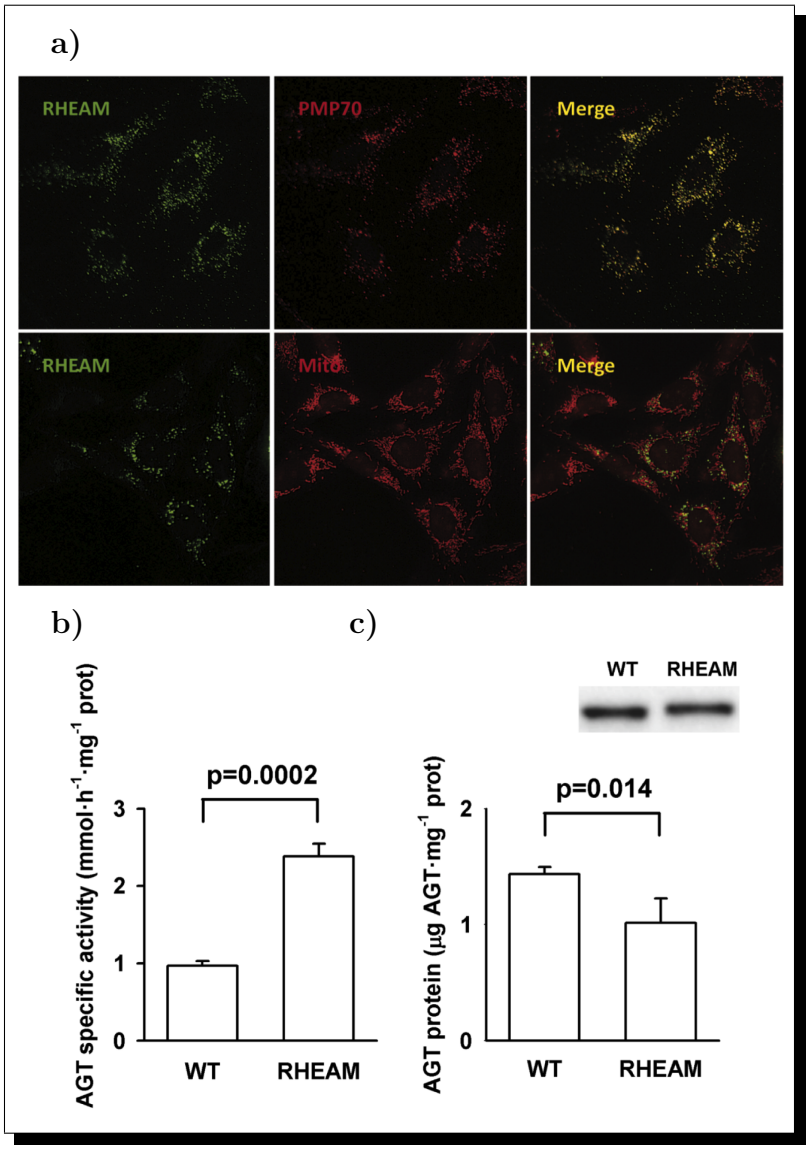


FIGURE 7.3.2. **AGT location, activity and immunoreactive protein levels in transiently transfected CHO cells. a)** subcellular localization by confocal microscopy with anti-AGT (green), anti-PMP70 (peroxisomal marker red) antibodies. Co-localization is detected as a yellow signal; **b)** specific AGT activity (data are mean±s.d. from 6 independent experiments); **c)** immunoreactive protein levels of AGT (data are mean±s.d. from 4 independent experiments). Inset is a representative WB analyses. The statistical significance is calculated using one-tailored unpaired Student's *t* test. *Data were obtained by Prof. Eduardo Salido from Centre for Biomedical Research on Rare Diseases at University of La laguna through the collaboration on this project*

CHAPTER 8

# Energetic insights into the human AGT enzyme stability

## 8.1. Brief introduction

Proteins have a marginal thermodynamic stability *in vitro*.[164] The energy landscape of the protein folding could be modified *in vivo* by the action of chaperones or folding enzymes. The fate of a protein will be influenced by an adequate balance of the proteostasis network. In many genetic diseases, the presence of a destabilizing mutation in a protein is translated in a reduction of the ability to fold in a biological network (folding deficiency) exceeding the capacity of the proteostasis network,[36] and thus protein stability is a key feature in protein misfolding diseases.

We have shown that PH1 variants generally affect the ability of the AGT protein to fold properly, leading to accelerated protein turnover, aggregation or mitochondrial mistargeting (see Appendix A). We have also suggested that the reduced kinetic stability of the apo native state may play an important role in the pathogenic mechanisms of PH1 variants. Alternatively, by using the consensus approach we have enhanced an *in vitro* AGT stability partly by optimization of electrostatic interactions in the native state (see Appendix B).

In this context, we have performed a characterization of the conformational stability of a set of AGT proteins composed of natural (non-pathogenic and disease-causing) and highly stable (consensus-based) variants. The objective is to obtain information about the existence of common patterns in the modulation of the stability *in vitro* and inside cells. In this chapter, these results are presented and discussed.

## 8.2. Results

**Holo naturally-occurring and consensus variants share a common thermal denaturation mechanism.** The characterized set of proteins is composed of ten naturally occurring and twelve consensus-based AGT variants (see Figure 8.2.1), in which all mutations are scattered over the AGT structure and only some consensus-based mutations are found in a cluster close to the dimerization interface. This set of proteins consists of five non-pathogenic polymorphisms (AGT-Ma, AGT-Mi, P11L-Ma, I340M-Ma and R197Q-Mi), five disease-causing variants (H83R-Mi, G170R-Mi, I244T-Mi, A295T-Mi and A368T-Mi), five single consensus-based mutations (Q23R-Ma, S48H-Ma, S50H-Ma, D52E-Ma and V113A-Ma), three double mutants (HE-Ma, HM-Ma and EM-Ma), one triple mutant (HEM-Ma), two quadruple mutants (RHEM-Ma and HEAM-Ma) and one quintuple mutant (RHEAM-Ma)*.

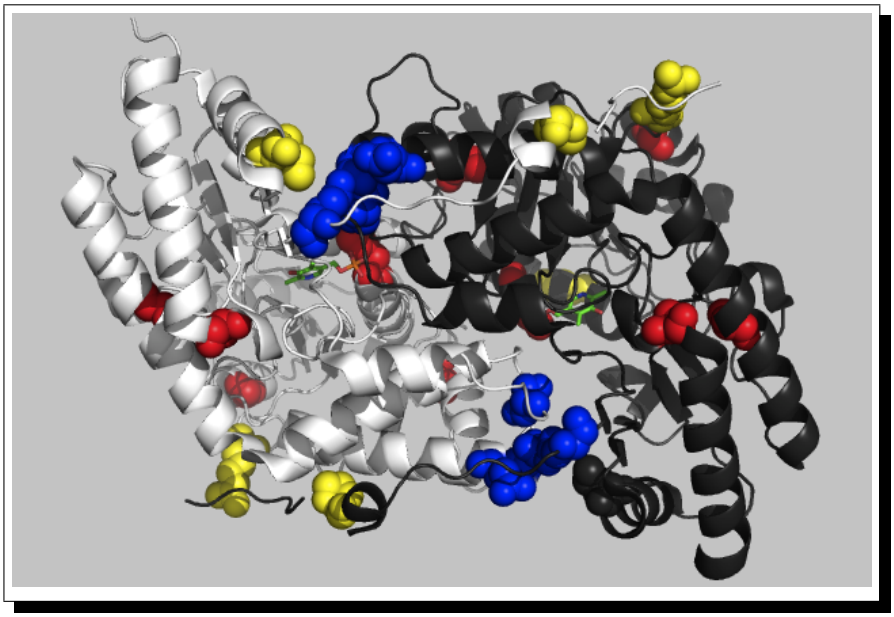

FIGURE 8.2.1. **Localization of the mutations on AGT structure.** Residues shown in red are changed in naturally-ocurring variants and those in blue are involved in consensus mutations.

For all variants, the thermal unfolding in the absence or presence of low-denaturating urea concentration was described by a two state irreversible model, with first order kinetics (see Figure 8.2.2-a and b). In addition, the thermal

*Please, note that the multiple mutants are named with the letters of the new mutations introduced. For example the double mutant of S48H and I340M is termed as HM-Ma

unfolding was also found to be essentially independent on protein concentration (see Figure 8.2.2-c).

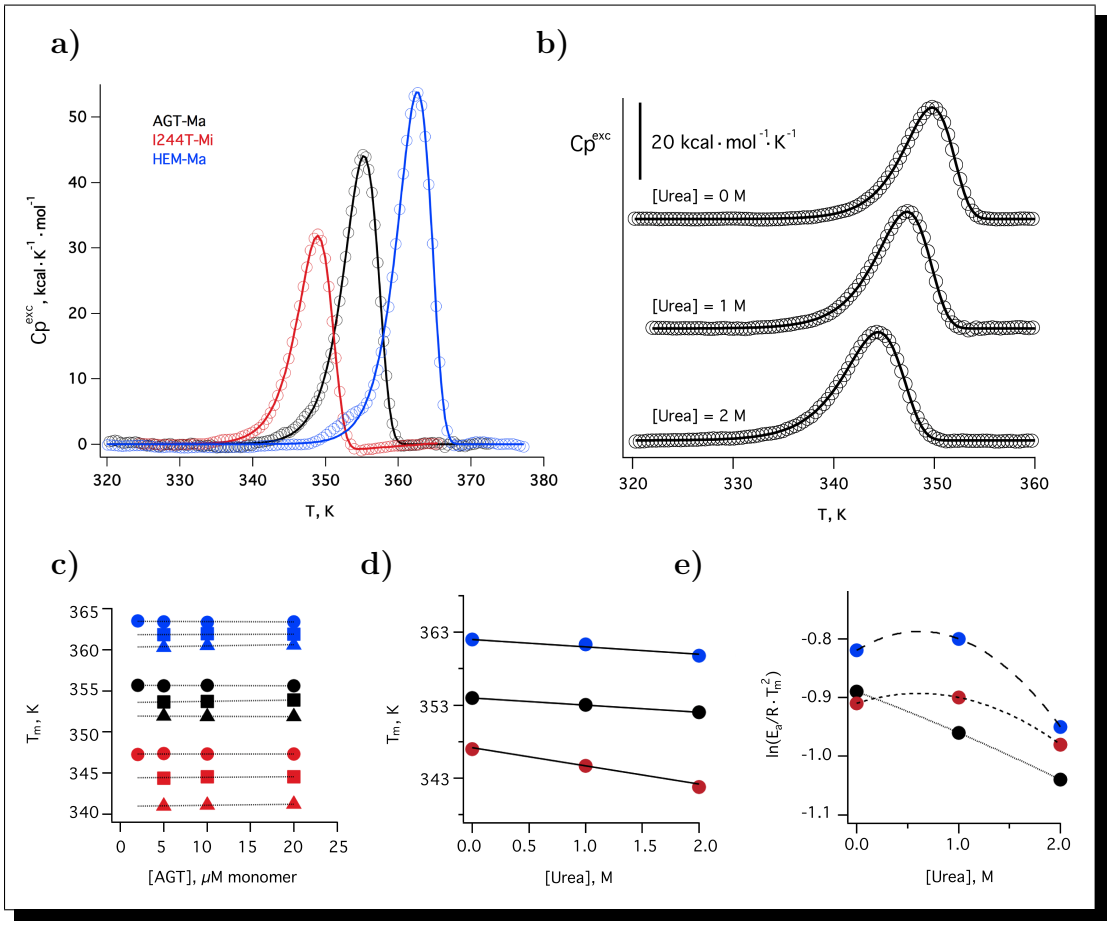


FIGURE 8.2.2. **Thermal unfolding of AGT variants in the absence or presence of urea and determination of kinetic $m$ values. a)** experimental DSC thermograms obtained at 3°C/min for AGT-Ma (black), I244T-Mi (red) and HEM-Ma (blue); **b)** effect of the urea concentration in the stability of AGT-Ma; **c)** dependence of the stability with protein concentration for AGT-Ma (black), I244T-Mi (red) and HEM-Ma (blue) for urea concentration of 0 M (circle), 1 M (squares) and 2 M (triangles); **d)** and **e)** evaluation of the kinetic $m$ values at scan rate of 3°C/min. Lines are the best fit for the data.

**AGT variants share similar degree of unfolding by thermal denaturation.** We have evaluated the changes in accessible surface area ($\Delta ASA$) from the native state[50] and a model of the unfolded states based on Gly-X-Gly tripeptides.[165] We have obtained that 64% of the buried surface upon AGT folding is apolar surface while 36% corresponds to polar surface (see Table 8.2.1) in agreement with contributions to total changes of ASA in a wide set of proteins.[133] It has long been known that buried surface area correlates with

the size of the protein (number of residues), and the value obtained for human AGT correlates with the expected value of a protein of this size.[133] Experimental denaturation enthalpies ($\Delta H$) for the thermal unfolding of this set of AGT enzymes are strongly dependent on $T_m$ values (see Figure 8.2.3-a and Table 8.2.2), yielding a slope of 11.3 ± 0.8 $kcal \cdot K^{-1} \cdot mol^{-1}$, which can be considered as the heat capacity change upon unfolding ($\Delta Cp$). This value correlates well with the theoretical value of $\Delta Cp$ for human AGT of 10.87 $kcal \cdot K^{-1} \cdot mol^{-1}$ that we can obtain using structure-energetics parameterization[133] for globular proteins. These results, suggest that most of the native structure of AGT is disrupted in the thermally denatured state and all variants denature to a similar extent.

TABLE 8.2.1. **Changes in accessible surface area of the human AGT.** The crystal structure of the human AGT-Ma was used for this analysis.[50]

| | $ASA_{total}$, Å$^2$ | $ASA_{apolar}$, Å$^2$ | $ASA_{polar}$, Å$^2$ |
|---|---|---|---|
| $Native$ | 27087 | 15392 | 11695 |
| $Tripeptide$ | 128499 | 80474 | 48025 |
| $\Delta^U_N ASA$ | 101412 | 65082 (64.2%) | 36330 (35.8%) |

**The denaturation transition state of AGT is native-like in terms of area exposed to solvent.** If we focus now in the dependence of the activation energy ($E_a$) values on the $T_m$ parameter (see Figure 8.2.3-b), the correlation yields a slope of 1.85 ± 0.15 $kcal \cdot K^{-1} \cdot mol^{-1}$. If we consider this slope as an activation heat capacity change ($\Delta Cp^{\neq}$) between the transition and native state, it may indicate a native-like transition state in terms of exposure to solvent. Accordingly, the kinetic $m$ values ($m^{\neq}$) have been evaluated from the urea concentration dependence on the $T_m$ parameter, as a measure of the exposure to the solvent of the transition state[166] (see Figure 8.2.2-d and -e and Table 8.2.2). The experimental values obtained were relatively small ($m^{\neq}$ ∼0.4-1.2 $kcal \cdot mol^{-1} \cdot M^{-1}$; see Figure 8.2.3-c) with an average value of $m^{\neq} = 0.77 \pm 0.21$ $kcal \cdot mol^{-1} \cdot M^{-1}$, in comparison with the theoretical equilibrium $m$ value ($m^{eq}$) estimated from the crystal structure of AGT ($m^{eq} = 11.5$ $kcal \cdot mol^{-1} \cdot M^{-1}$).[167] The low value obtained for the ratio $m^{\neq}/m^{eq}$ is also in agreement with a transition state *close* to the native state in terms of solvent exposure. On the other side, there is no clear link between the kinetic $m$ values obtained for the set of AGT variants and the $T_m$ parameter (see Figure 8.2.3-c inset).

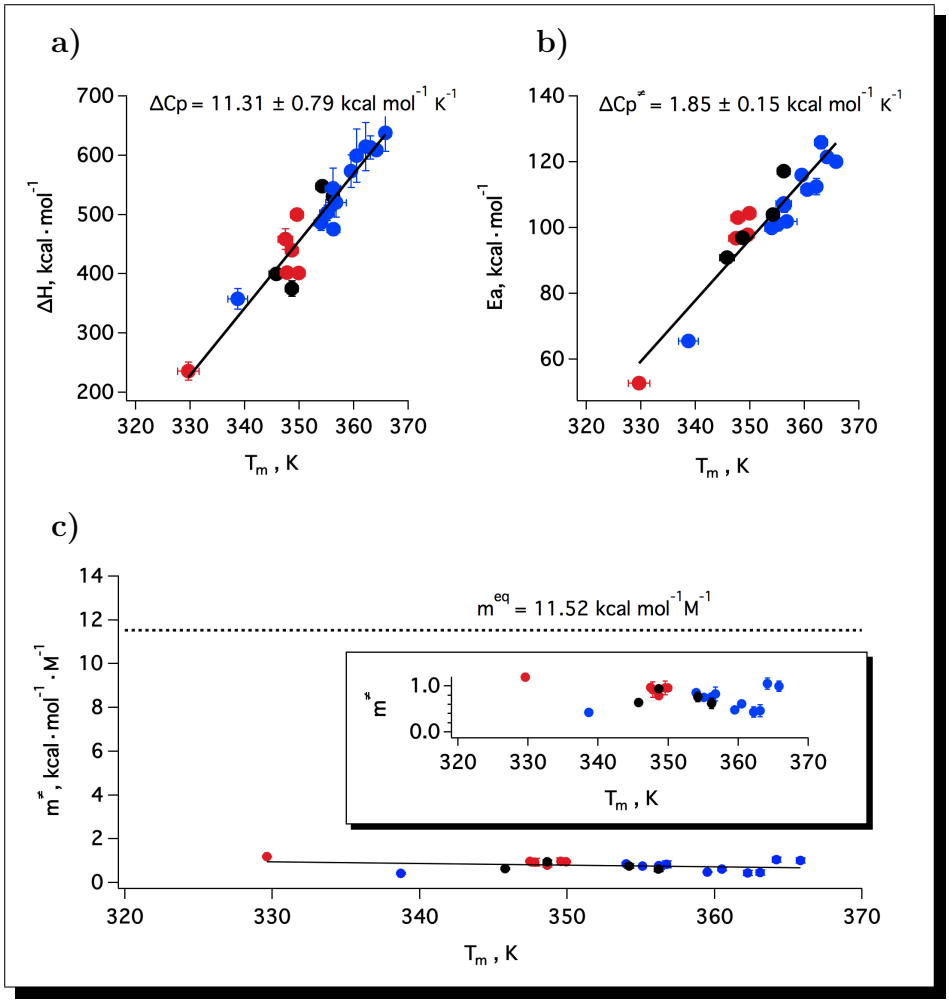


FIGURE 8.2.3. **Correlation of a) calorimetric enthalpies, b) activation energies and c) kinetic $m$ values ($m^{\neq}$) with the $T_m$ parameter.** The calorimetric enthalpies, and kinetic $m$ values are the mean±s.d. from three different scan rates; the activation energies are obtained from Arrhenius plots. Markers correspond to experimental data for polymorphisms (black), disease-causing (red) and consensus-based (blue) variants of AGT protein. Continuous lines correspond to the fit of the data to a straight line. Dashed line in c) shows the equilibrium $m$ value as a reference.

**Mutations on human AGT show an enthalpy-entropy compensation of the denaturation free energy barrier.** The enthalpic ($\Delta H^{\neq}$) and entropic ($\Delta S^{\neq}$) contributions of the unfolding activation free energy ($\Delta G^{\neq}$) have been evaluated for each variant[165] at the physiological temperature of 37°C (see Figure 8.2.4-a). Results obtained support that mutations affect the kinetic stability by producing large changes in the enthalpic and the entropic component of the activation free energy. However, the changes in the activation

TABLE 8.2.2. **Summary of the main parameters for the thermal unfolding and kinetic stability of holo AGT variants obtained by DSC.**

| | *Variant* | $T_m$ [1] | $\Delta H$ [2] | $E_a$ [3] | $k_{37}$ [4] | $m^{\neq}$ [5] |
|---|---|---|---|---|---|---|
| | | $^oC$ | $kcal/mol$ | $kcal/mol$ | $min^{-1}$ | $kcal/mol \cdot M$ |
| NATURAL VARIANTS | AGT-Ma | 82.1 | 547 ± 5 | 109 ± 5 | $6.4{\cdot}10^{-10}$ | 0.76 ± 0.10 |
| | AGT-Mi | 76.6 | 375 ± 13 | 101 ± 5 | $2.2{\cdot}10^{-8}$ | 0.94 ± 0.07 |
| | P11L-Ma | 73.8 | 399 ± 5 | 95 ± 3 | $1.1{\cdot}10^{-7}$ | 0.64 ± 0.04 |
| | I340M-Ma | 84.1 | 530 ± 24 | 115 ± 4 | $1.7{\cdot}10^{-11}$ | 0.62 ± 0.11 |
| | H83R-Mi | 58.2 | 235 ± 16 | 55 ± 3 | $2.9{\cdot}10^{-3}$ | 1.19 ± 0.02 |
| | G170R-Mi | 75.5 | 459 ± 17 | 100 ± 2 | $3.3{\cdot}10^{-8}$ | 0.96 ± 0.05 |
| | R197Q-Mi | 77.9 | 401 ± 6 | 110 ± 6 | $3.7{\cdot}10^{-9}$ | 0.95 ± 0.05 |
| | I244T-Mi | 75.8 | 402 ± 9 | 103 ± 2 | $1.1{\cdot}10^{-8}$ | 0.92 ± 0.15 |
| | A295T-Mi | 77.5 | 500 ± 10 | 105 ± 8 | $1.2{\cdot}10^{-8}$ | 0.96 ± 0.15 |
| | A368T-Mi | 76.5 | 439 ± 6 | 102 ± 7 | $2.1{\cdot}10^{-8}$ | 0.79 ± 0.07 |
| CONSENSUS VARIANTS | Q23R-Ma | 84.5 | 475 ± 10 | 112 ± 5 | $1.1{\cdot}10^{-10}$ | 0.82 ± 0.15 |
| | S48H-Ma | 85.5 | 520 ± 24 | 104 ± 3 | $2.33{\cdot}10^{-10}$ | 0.66 ± 0.06 |
| | S50H-Ma | 67.1 | 357 ± 17 | 67 ± 2 | $6.82{\cdot}10^{-5}$ | 0.42 ± 0.14 |
| | D52E-Ma | 84.4 | 544 ± 34 | 104 ± 6 | $1.30{\cdot}10^{-10}$ | 0.76 ± 0.06 |
| | V113A-Ma | 83.2 | 502 ± 13 | 101 ± 1 | $1.98{\cdot}10^{-9}$ | 0.75 ± 0.01 |
| | HM-Ma | 88.0 | 599 ± 45 | 116 ± 9 | $1.08{\cdot}10^{-11}$ | 0.60 ± 0.07 |
| | HE-Ma | 81.5 | 486 ± 13 | 101 ± 1 | $7.20{\cdot}10^{-10}$ | 0.85 ± 0.06 |
| | EM-Ma | 86.9 | 573 ± 28 | 121 ± 8 | $6.38{\cdot}10^{-12}$ | 0.48 ± 0.01 |
| | HEM-Ma | 89.5 | 315 ± 41 | 118 ± 10 | $4.28{\cdot}10^{-12}$ | 0.43 ± 0.11 |
| | HEAM-Ma | 91.0 | 613 ± 20 | 122 ± 5 | $9.5{\cdot}10^{-14}$ | 0.46 ± 0.13 |
| | RHEM-Ma | 92.0 | 609 ± 5 | 125 ± 5 | $1.7{\cdot}10^{-13}$ | 1.05 ± 0.12 |
| | RHEAM-Ma | 93.7 | 638 ± 31 | 125 ± 6 | $1.1{\cdot}10^{-13}$ | 0.99 ± 0.12 |

[1] Determined at scan rate of 3°C/min.

[2] mean±s.d. of different scan rates; expressed per mole of dimer.

[3] mean±s.d. from four consistency tests.[121]

[4] extrapolated from Arrhenius plots.

[5] mean±s.d. from DSC at 3-5 urea concentrations and 3 different scan rates.

free energy are comparatively small due to an enthalpy-entropy compensation where both components largely cancel each other out.

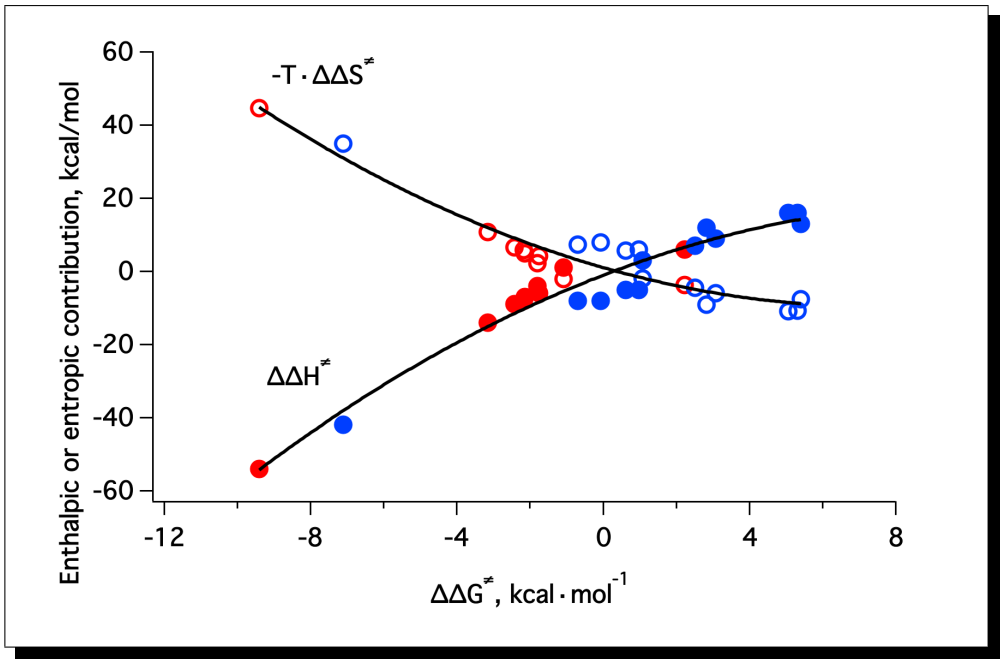


FIGURE 8.2.4. **Enthalpic (closed markers) and entropic (open markers) contributions to the activation free energy for the natural (red) and consensus-based (blue) variants of the human AGT at temperature of 37°C.** Lines are placed to point out the trend and have no theoretical meaning.

**Dissection of the kinetic stability of human AGT: *solvation* and *unfolding* contributions to the activation enthalpy.** During the protein folding process, the water molecules removed from the surface are involved in a water-associated barrier (desolvation barrier) to protein folding. Conversely, during the unfolding process of a protein, the asynchrony between water penetration and breakup of the internal interactions could be described as a solvation barrier.[18,166] Therefore, it has been suggested that the activation energy ($\Delta G^{\neq}$) is composed by: i) contributions from the *unfolding* (unfolding and solvation of regions of the native state in the transition state); and ii) contributions from the *solvation barrier* (networks of broken internal contacts but not yet solvated in the transition state).[165,166]

The activation energy of the protein unfolding can be viewed as the enthalpy difference between the transition and the native state ($E_a \approx \Delta H^{\neq}$).[108] Moreover, the ratio between kinetic urea $m$ value and equilibrium $m$ value ($m^{\neq}/m^{eq}$) can be considered as an estimation of the change in the degree of exposure to the

solvent between transition and native state[133] and the unfolding contribution to activation enthalpy ($\Delta H_{unf}$) can be evaluated from the calorimetric enthalpy derived from DSC. Accordingly, the solvation-barrier ($\Delta H^*$) contribution to activation enthalpy can be evaluated by difference[165] ($\Delta H^{\neq} = \Delta H_{unf} + \Delta H^*$; see Figure 8.2.5).

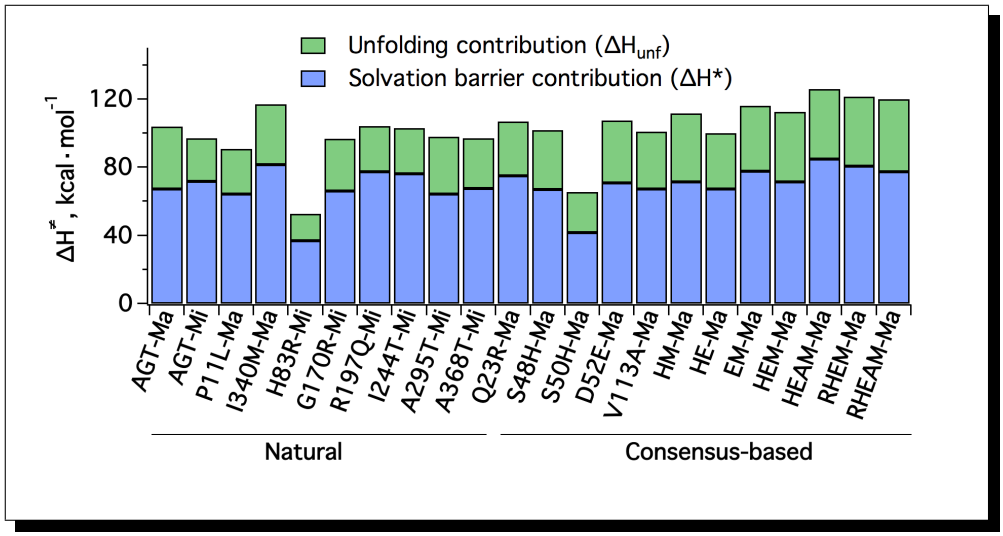


FIGURE 8.2.5. **Unfolding ($\Delta H_{unf}$ - green) and solvation barrier ($\Delta H^*$ - blue) contributions to the activation enthalpies ($\Delta H^{\neq}$) for each AGT variants.**

Our data for this set of AGT variants show that the main contribution to the activation enthalpy is the solvation barrier component (average of 68 ± 3%). Although there are no clear differences between the natural and consensus-based variants, both components of the activation enthalpy correlate with the kinetic stability as can be seen from their moderate increase with the raise of the stability (expressed with the $T_m$ parameter) (see Figure 8.2.6-a).

Structural changes, in terms of accessible surface area between the native and the transition state, can be estimated from the unfolding and solvation barrier contributions[133,165] (see Figure 8.2.6-b). These results show that structural changes between the transition and the native state are always small (average of 18.6%) compared with the total change in the accessible surface area for the complete unfolding. Moreover, the structural changes associated with the solvation barrier are always bigger than those associated with the unfolding contribution. Finally, within the uncertainty associated to these calculations, it seems that the structural changes associated with the solvation barrier increase

as the kinetic stability is raised while the trend for the unfolding contribution seems to be of opposite sign and weaker (see Figure 8.2.6-b).

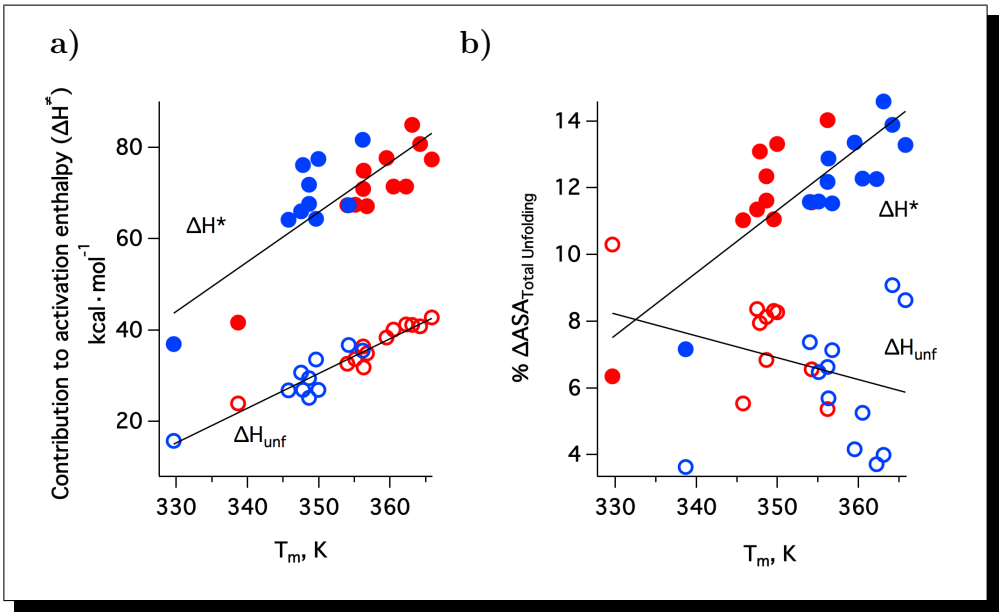


FIGURE 8.2.6. **Dissection of the activation enthalpy of the natural (red) and consensus-based (blue) variants of human AGT.** Temperature dependence of the **a)** energetics contributions and **b)** structural changes (as % of total ΔASA) of the unfolding (open markers) and solvation barrier (closed markers) component.

**Estimation of the *solvation* and *unfolding* contributions to the activation entropy.** In the previous subsection, the activation enthalpy has been deconstructed into its unfolding and solvation barrier contributions. However, the free activation energy ($\Delta G^{\neq}$) will also have an entropic component ($\Delta S^{\neq}$) that can also be deconstructed in *unfolding* and *solvation-barrier* component: $\Delta S^{\neq} = \Delta S_{unf} + \Delta S^{*}$. In order to estimate these contributions, the existence of an enthalpy-entropy compensation is invoked[165] (see more details in M&M section).

This decomposition shows that both components of the activation entropy have a similar correlation with the stability of the variants and also both components contribute to the same extent to the activation entropy (see Figure 8.2.7).

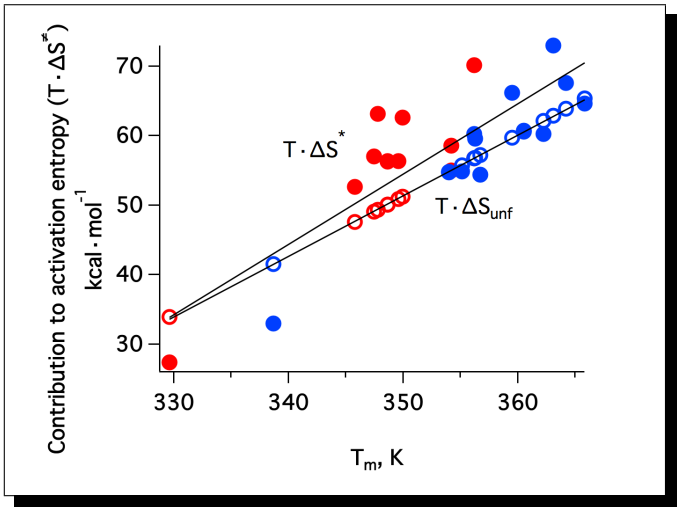


FIGURE 8.2.7. **Dissection of the activation entropy of the natural (red) and consensus-based (blue) variants of human AGT.** Temperature dependence of the energetics contributions of the unfolding (open markers) and solvation barrier (closed markers) components.

**The rate-limiting step in the kinetic stability of human AGT.** We have shown in the previous chapters that the irreversible two-state kinetic model ($N \xrightarrow{k} F$) provides a good description of the thermal denaturation of human AGT variants. We must keep in mind that this two-state kinetic model is a simplification of a more general scheme known as the Lumry-Eyring model:[122]

$$N \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} U \xrightarrow{k_2} F$$

$$K = \frac{k_1}{k_{-1}}$$

in which an unfolded or partially unfolded state (U) is in equilibrium with the native state (N), but it can also suffers irreversible modifications to a non-functional final state (F) that cannot fold back again.

In this model, if the irreversible step ($U \xrightarrow{k_2} F$) does not take place significantly during the time of the DSC transition, it can be assumed that thermograms reflect information about the reversible step so an equilibrium thermodynamic analysis can be done. On the other side, if the irreversible step takes place during the time that protein spends in the DSC transition, the thermogram will be strongly rate limited.[122] Two limiting cases can be presented: i) the irreversible step ($U \rightarrow F$) is fast, therefore the overall denaturation rate

is limited by unfolding kinetics ($N \rightarrow U$); ii) the irreversible step ($U \rightarrow F$) is slow, therefore both steps determine the overall denaturation rate. To discriminate between these two scenarios, the kinetics of global unfolding have been determined for four selected variants:

- AGT-Ma;
- I340M-Ma (highly stable natural polymorphism);
- HEM-Ma (very highly stable consensus-based variant);
- and A295T-Mi (destabilizing disease-causing variant).

It was found that HEM-Ma variant is kinetically resistant to high concentration of urea (9M) at physiological temperature, while A295T-Mi denatures 4-fold faster than AGT-Ma (see Figure 8.2.8-a). We have evaluated the urea denaturation rate constants extrapolated to no denaturant at different temperatures (see Figure 8.2.8-b and c). As for thermal unfolding, the urea denaturation rate constant is independent of protein concentration supporting that no dimer dissociation occurs prior to the rate limiting step (see Figure 8.2.8-d)

The kinetic $m$ values derived from urea-induced unfolding kinetics are comparatively small and nearly temperature independent (similar to those obtained by DSC), with an average $m^{\neq}$ values at different temperatures of: AGT-Ma: 0.55 ± 0.05; A295T-Mi: 0.32 ± 0.02; I340M-Ma: 0.69 ± 0.04 and HEM-Ma: 0.51 ± 0.04 $kcal \cdot mol^{-1} \cdot M^{-1}$. On the other side, the activation energy values obtained for I340M-Ma and HEM-Ma (130 ± 5 and 96 ± 6 $kcal \cdot mol^{-1}$ respectively) are similar to those obtained by thermal unfolding, while the activation energies obtained for AGT-Ma and I295T-Mi are 4-fold lower (22 ± 12 and 22 ± 4 $kcal \cdot mol^{-1}$ respectively) compared to thermal unfolding (see Figure 8.2.8-e).

The kinetic stability (kinetic constant extrapolated at 37°C) of the irreversible denaturation by temperature is found to be higher than that obtained by urea-induced unfolding (see Figure 8.2.8-e). This result suggests that in the context of the $N \leftrightarrow U \rightarrow F$ mechanism, the $N \rightarrow U$ step is not rate limiting. Therefore the overall denaturation rate may depend on either, the thermodynamic stability or the rate of the irreversible step. As a consequence, the changes in kinetic stability in natural and consensus-based mutants may arise from changes in the thermodynamic stability and the aggregation rate constant of unfolded states.

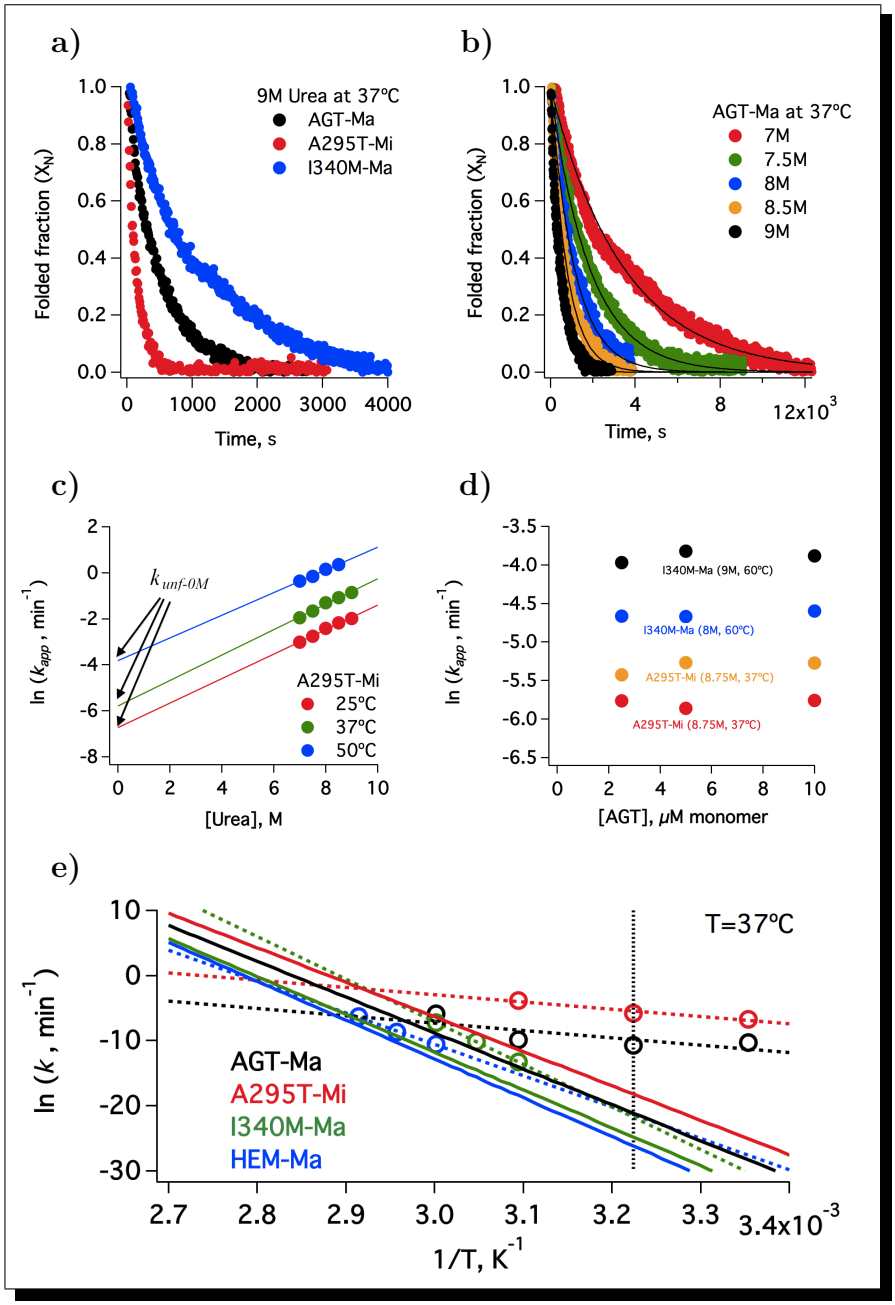


FIGURE 8.2.8. **Kinetics of the urea-induced unfolding. a)** Urea denaturation kinetics of three holo AGT variants at 37°C measured by far-UV CD spectroscopy; **b)** urea denaturation kinetics of holo AGT-Ma at 37°C at different urea concentrations; **c)** linear extrapolation of unfolding rate constants at different urea concentration and temperatures to the absence of denaturation ($k_{unf-0M}$); **d)** dependence of $k_{unf-0M}$ at different temperatures and urea concentration with the AGT concentration; **e)** comparison of denaturation rates obtained from thermal unfolding (DSC - solid lines) and global unfolding rates in the absence of urea (open markers and dashed lines). Physiological temperature is indicated by a vertical dotted line.

In addition, the aggregation propensity has been evaluated by the use of two algorithms, which model aggregation from unfolded regions:[168,169] $Z_{agg}$, which measures the propensity to form high molecular weight aggregates (see Figure 8.2.9); and $Z_{tox}$, which measures the tendency to form low molecular weight aggregates (see Appendix C). This bioinformatic tools may provide an estimation of the rate constant of the irreversible step ($U \rightarrow F$).

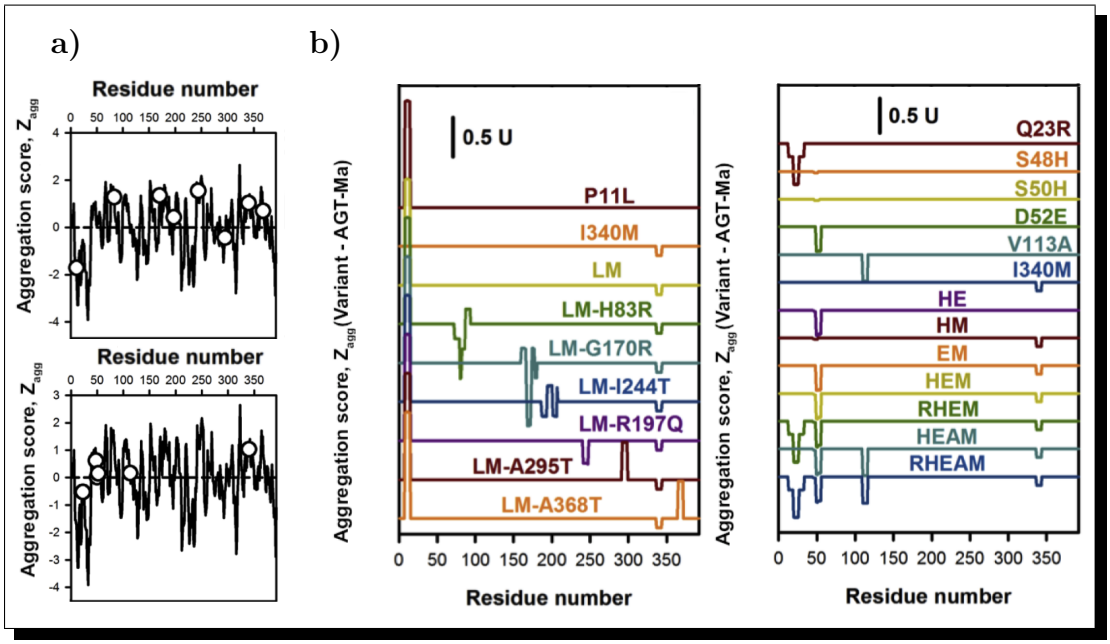


FIGURE 8.2.9. **Sequence-based prediction of aggregation propensities from the unfolded states ($Z_{agg}$ score) using the *Zyggregator* algorithm[168,169]. a)** Propensities per residue for naturally-occurring (upper) and consensus-based (bottom) variants; **b)** difference of propensities between each variant and AGT-Ma sequence for naturally-occurring (left) and consensus-based (right) variants.

Both methods show similar results. Mutations of natural variants seem to occur on regions with some aggregation propensity (positive $Z_{agg}$, see Figure 8.2.9-a) while the consensus based mutations show no clear bias towards occurring in regions with any preference towards aggregation. In addition, the naturally occurring variants show no direct link between aggregation propensity and protein kinetic stability, while the consensus based mutations appear to reduce the aggregation propensity (see Figure 8.2.9-b). This result supports that the irreversible step (aggregation to a final state F) might be slowed down by consensus-based mutations.

## 8.3. Discussion

Our mechanistic and energetic studies on a set of human AGT variants have shown kinetic stabilities spanning over 11 orders of magnitude due to changes in activation free energies up to $\sim$15 $kcal \cdot mol^{-1}$ between disease-causing and consensus-based variants. The increase of the activation free energy barrier is due to large and highly compensating changes in the enthalpic-entropic contributions (see Appendix C).

Although the relevant denaturation transition state retains a native-like overall structure in terms of exposure to solvent, changes in kinetic stability seem to arise from those relatively small changes of the structure/energetic balance. These changes come mainly from solvation barriers between native and transition state, due to an increase of those networks of broken native internal contacts that have not been yet solvated in the transition state.

Some local changes in the structure or dynamics of the native state upon mutations could not be excluded. However, it is unlikely that they could explain the difference in ASA due to solvation barrier since these variants share commons traits such as similar hydrodynamic behaviour, similar activity, denaturation energetics consistent with similar degree of unfolding and, for some of them, almost superimposable crystal structures (see Appendix A and B). Therefore, it seems that the AGT stability depends on a very delicate balance between the structure and dynamics of the native and denaturation transition state.

It was shown, in the previous chapter, that consensus-based mutations enhance the kinetic stability at least partly by developing new favourable interactions in the AGT dimer. Differences in thermal denaturation and urea-induced global unfolding, suggest that the kinetic stability of AGT is modulated by changes in thermodynamic stability and the aggregation propensity of unfolded states. A complete thermodynamic characterization of AGT using chaotropic-induced unfolding cannot be performed due to irreversibility.[7,123]

Alternatively, the aggregation propensity of the unfolded states was evaluated by bioinformatics tools suggesting that the irreversible step might be slowed down by consensus-based mutations without a clear trend for naturally-occurring variants. This might imply that disease-causing mutations plausibly affect the thermodynamic stability of the native AGT, which could establish a

mechanistic link between a reduced kinetic stability and a reduced foldability upon expression in eukaryotic cells (see Figure 8.3.1).

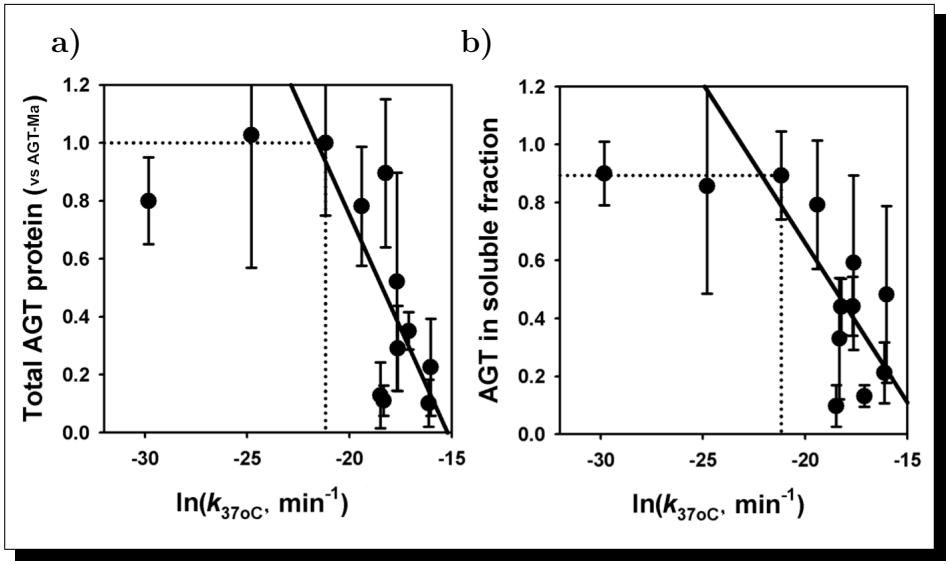


FIGURE 8.3.1. **Correlation between kinetic stability of holo AGT proteins and intracellular foldability. a)** total protein levels and **b)** fraction of soluble protein in eukaryotic cells as a function of kinetic stabiliy. Solid lines are linear fits of AGT variants with kinetic stabilities equal or lower than AGT-Ma while dotted lines show AGT-Ma as a threshold. Correlation coefficients ($r^2$) are a) 0.56 and b) 0.48. *Expression data were obtained by Prof. Eduardo Salido from Centre for Biomedical Research on Rare Diseases at University of La laguna through the collaboration on this project.*

Here we have also proposed that consensus-based mutations enhance the *in vitro* stability mostly by changes in solvation barriers that slightly affect the structure-dynamics of native and/or transition state. Nonetheless, consensus variants do not increase AGT foldability inside eukaryotic cells. Therefore it is plausible that human AGT-Ma represents a stability threshold to provide an optimum stability for intracellular folding to active dimers (90% of AGT-Ma is found in soluble fractions *in vivo*) and proper turnover inside peroxisomes (see Appendix A and Figure 8.3.1). As a result we hypothesize that the presence of consensus mutations in other vertebrate sequences may provide additional stability to compensate the fixation of destabilizing mutations that are beneficial for AGT function in the fitness of these organisms.

In terms of molecular evolution, a protein sequence under a selective pressure will explore a space of sequences where only those mutations that are compatible with the structure and the function of the protein will be fixed (fitness

threshold).[115] While consensus-based mutations have a high degree of conservation among vertebrates, naturally-occurring mutations on human AGT show a very low degree (see Figure 8.3.2-a and -b).

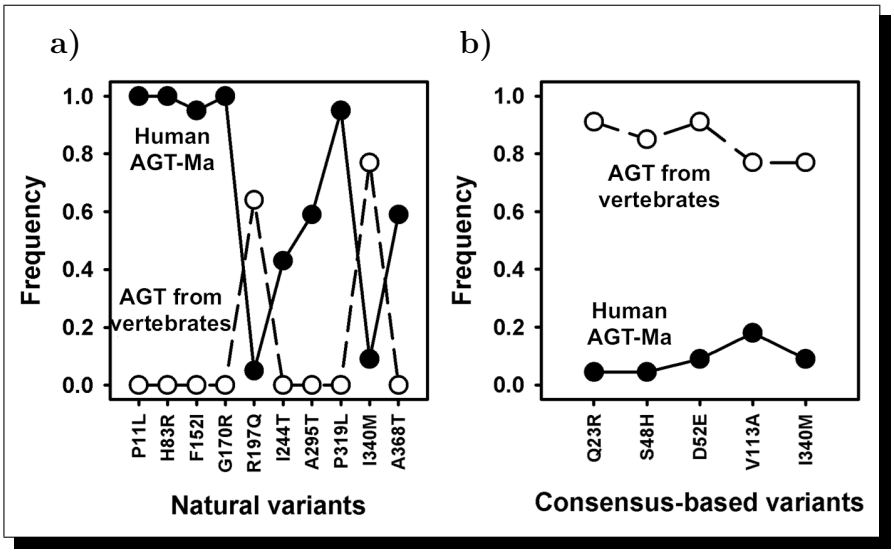


FIGURE 8.3.2. **Divergence in conservation of a) natural (disease-associated) and b) consensus-based variants in AGT sequences from vertebrates.** Closed symbols indicate the frequency of the residue found in human AGT-Ma while open symbols refer to the corresponding vertebrate sequences.

We have shown that the minor allele (AGT-Mi that has P11L and I340M substitution), which is found in about 20% of non-PH1 populations, shows a remarkable effect on protein stability and intracellular foldability. These effects are suggested to be mostly due to the presence of the P11L substitution. Thus the minor allele is known to predispose human AGT towards additional deleterious mutations and may represent a lower limit for AGT stability i.e., the threshold between health and disease. When a disease-causing mutation is found on the AGT-Mi background, the folding energetics of the protein may be located close to or beyond the proteostasis boundary. Therefore, mitochondrial mistargeting or aggregation events may be the consequence of modified interactions with elements of the proteostasis system such as chaperones.[7,59,170] Then, AGT activity in peroxisomes may decrease below a threshold incompatible with health and thus causing the onset of the disease.

In summary, human AGT-Ma displays an adequate conformational stability to fold inside cells and to be targeted to peroxisomes but it still interacts with molecular chaperones to reach its functional state.[7] Alternatively, AGT-Mi

may be close to the lower fitness requirements in a healthy cell, thus explaining the high frequency of disease-causing mutations on its background. Although molecular chaperones are able to buffer the effects of destabilizing mutations, the chronic expression of a misfolded protein together with changes in the proteostasis capacity through aging may influence the onset of the disease. In this situation, different elements of the proteostasis network may modulate the loss-of-function in PH1 by affecting the partition of AGT between aggregation, degradation or mistargeting.[83,93,171]

# Part 4

# Conclusions / Conclusiones

# Conclusions

Through the different chapters of this thesis, we have performed a detailed *in vitro* characterization of the human AGT protein from different perspectives with the purpose of finding new therapeutic approaches to PH1.

The *in vitro* characterization of disease-causing mutants has revealed a high destabilization of apo AGT, which partly correlates with kinetic trapping by molecular chaperones and intracellular protein foldability and mistargeting in cells.[7] In addition, we have observed a coenzyme induced kinetic overstabilization of some mutants that may explain the pyridoxine responsiveness found in patients carrying these mutations.[172] Therefore we suggest that native state kinetic stabilizers and protein homeostasis modulators may be suitable to correct folding and stability in PH1.[13]

We have considered the use of the nematode *C. elegans* as an animal model for PH1 and AGT proteostasis *in vivo*. The *in vitro* characterization of the orthologous AGT protein in this nematode and the key role of glyoxylate in the normal development of *C. elegans* indicate that it might not be suitable for developing a hyperoxaluric phenotype. Nonetheless *C. elegans* could still be a proper model to study interactions of AGT protein with different elements of the proteostasis network (ongoing project).[173]

The consensus approach has been proved to be a simple tool for improving kinetic stability and catalytic performance of human AGT. Consensus-based proteins could be used in gene and enzyme replacement therapies for PH1[15] (ongoing project), and in general for inborn errors of metabolism.[16] While expression of the enhanced AGT in cells is properly located in peroxisomes, the higher *in vitro* stability is not translated to higher resistance toward degradation. However, our study clearly shows that consensus approach optimizes interactions in AGT native structure that efficiently increase the kinetic free energy barrier for denaturation while the overall protein folding and enzyme function are kept and the crystallization process is sped up. We suggest that

AGT-Ma variant has been provided through evolution with a proper conformational stability[17] and turnover to perform its biological function in a harsh organelle such as peroxisome. Therefore, AGT-Ma may define an *upper* limit of kinetic stability beyond which an increase in stability is not translated in an increase of resistance to *in vivo* degradation. We also indicate that consensus approach could be a simple and generic strategy to obtain high resolution structural models of protein for which crystallization is a challenging task due to stability issues.

Finally, a deep insight into the energetics of the AGT protein has shown that disease-causing mutations mostly affect the thermodynamic stability of native AGT compromising the foldability in cells. At the same time, consensus-based mutations clearly enhance the kinetic stability mostly through changes in solvation barriers between native and transition state by developing new favourable interactions in the native state, and reduction of the aggregation propensity of partially unfolded states as proposed by bioinformatic algorithms. While AGT-Ma could represent an *upper* stability threshold for optimal folding in cells, AGT-Mi could represent a *lower* stability threshold between health and disease, i.e., AGT-Mi is a non pathogenic variant, but its folding energetic may be close to the proteostasis capacity so the presence of PH1 mutations induce the onset of the disease. Buffering effects of molecular chaperones on destabilizing mutations may allow disease-causing mutants to fold intracellularly but to a lower extent explaining the loss-of-function in these mutations. Our results also explain why pharmacological therapies based on native state ligands that boost AGT stability may lead to recovery of AGT function in cells.

# Conclusiones

A través de los diferentes capítulos de esta tesis hemos realizado una detallada caracterización *in vitro* de la enzima humana AGT desde diferentes perspectivas con el objetivo de encontrar nuevas aproximaciones terapéuticas.

La caracterización *in vitro* de los mutantes causantes de enfermedad ha revelado una alta desestabilización del estado apoenzima de la AGT, la cual correlaciona parcialmente con la existencia de estados atrapados cinéticamente por chaperonas moleculares y plegabilidad y localización mitocondrial errónea en células.[7] Además, hemos observado una sobre-estabilización de la estabilidad cinética por unión del coenzima, lo que puede explicar la respuesta a tratamientos con piridoxina en pacientes que presentan estas mutaciones.[172] En este sentido sugerimos que el empleo de estabilizadores del estado nativo y reguladores de la homeostasis proteica pueden permitir la corrección de los problemas de plegabilidad y estabilidad en HP1.[13]

Hemos considerado el empleo del nematodo *C. elegans* como un modelo animal para HP1 y la proteostasis *in vivo* de la enzima AGT. La caracterización *in vitro* de la proteína ortóloga a la enzima humana AGT en este nematodo, y el papel relevante que tiene el glioxilato en el desarrollo normal de *C. elegans* sugieren que quizás no sea un modelo adecuado para desarrollar un fenotipo de hiperoxaluria. No obstante, *C. elegans* podría aún ser un modelo adecuado para el estudio de las interacciones de la enzima AGT con los diferentes elementos de la red de proteostasis (projecto bajo desarrollo).[173]

La aproximación de consenso ha sido empleada como una herramienta simple para la mejora de la estabilidad cinética y la capacidad catalítica de la enzima humana AGT. Las proteínas basadas en la aproximación de consenso podrían ser usadas en terapia genética y de reemplazamiento enzimático para el tratamiento de HP1[15] (projecto en desarrollo) y, en general, para tratar errores congénitos del metabolismo.[16] Mientras la enzima AGT mejorada es adecuadamente importada a peroxisomas cuando se expresa en células, la alta estabilidad *in vitro* no se traslada directamente a una mayor resistencia a la degradación.

Sin embargo, nuestro estudio muestra claramente que a través de la aproximación de consenso hemos optimizado interacciones en la estructura nativa de la enzima AGT, y que esto se ha trasladado eficientemente en un incremento de la barrera de energía libre para la cinética de desnaturalización, mientras el plegamiento y la función de la enzima se han mantenido y se ha acelerado el proceso de cristalización. Sugerimos que la variante AGT-Ma ha sido provista, a través de la evolución, de una adecuada estabilidad conformacional[17] y reciclaje proteico para llevar a cabo su función biologica en un entorno adverso como es el peroxisoma. Por tanto, la variante AGT-Ma puede definir un límite *superior* de estabilidad cinética más allá del cual un incremento de la estabilidad no se traslada en un incremento de la resistencia a la degradación *in vivo*. También indicamos que la aproximación de consenso podría ser una estrategia simple y genérica para obtener modelos estructurales de alta resolución de aquellas proteínas que presentan problemas de estabilidad.

Finalmente, un estudio profundo sobre la energética de la proteína AGT ha mostrado que las mutaciones causantes de enfermedad afectan principalmente la estabilidad termodinámica del estado nativo comprometiendo la plegabilidad en células. Al mismo tiempo, las mutaciones de consenso claramente mejoran la estabilidad cinética principalmente a través de cambios en la barrera de solvatación entre el estado nativo y de transición debido al desarrollo de nuevas interacciones favorables y a reducción de la propensidad de agregación de los estados parcialmente plegados, como se desprende de los algoritmos informáticos aplicados. Mientras la variante AGT-Ma podría representar un límite *superior* para el óptimo plegamiento en células, la variante AGT-Mi podría suponer un límite *inferior* entre la salud y la enfermedad. Es decir, aunque AGT-Mi es una variante no patogénica, su energética de plegamiento podría estar cercana a la capacidad de proteostasis y por tanto la presencia de las mutaciones de HP1 sobre AGT-Mi podrían inducir la aparición de la enfermedad. Los efectos atenuadores de las chaperonas moleculares sobre las mutaciones desestabilizantes podrían permitir que los mutantes de enfermedad plieguen intracelularmente, aunque en pequeña medida, explicando la pérdida de función de estas mutaciones. Nuestros resultados también explican por qué las terapias farmacológicas basadas en ligandos, que mejoran la estabilidad del estado nativo, pueden conducir a la recuperación de la función AGT en células.

# Part 5

# Materials and Methods

CHAPTER 9

# Expression and purification of recombinant proteins

## 9.1. Reactives

Reagents were acquired from Sigma-Aldrich except the protease inhibitor cocktail EDTA-free (Roche), the imidazole (Merck) and the tryptone and yeast extract (BD).

## 9.2. Human AGT variants

**Plasmids construction.** The cDNA of AGT variants were generated by Prof. Eduardo Salido at Centre for Biomedical Research on Rare Diseases (University of La Laguna, Spain) using site-directed mutagenesis and standard subcloning procedures. All constructs were confirmed by sequencing. The cDNA was inserted in a cold shock expression system pCold$^{TM}$ plasmid (pColdII DNA - Takara Bio Inc.) and *E. coli* BL21 strains were transformed. This plasmid includes a His-tag sequence (at the N-terminal of the protein target) and ampicillin resistance.[174]

**Expression.** BL21 strains containing plasmids encoding AGT variants were grown in presence of ampicillin (0.1 mg/ml) until an absorbance of 0.6 at 600 nm was reached. Then, cultures were cooled down to a temperature below 15°C before induction with $isopropyl\,\beta - D - 1 - thiogalactopyranoside$ (IPTG) at 1 mM. The induction took place during 16 h at 4°C, and cells were collected by centrifugation at 4500 g for 10 minutes (Kokusan H-251 centrifuge). Cells were washed with binding buffer ($NaH_2PO_4$ 20 mM, $NaCl$ 200 mM, $Imidazole$ 50 mM, pH 7.4) and again harvested by centrifugation in a swing-out rotor (Hettich Universal 320R centrifuge) at 2800 g for 10 minutes. Finally cells were frozen for at least 10 h at -80°C prior the purification process.

**Purification.** The lysis of cells was performed by sonication (Sonics Vibracell VCX 130) with a 6 mm probe in cold lysis buffer ($NaH_2PO_4$ 20 mM, $NaCl$ 200 mM, $Imidazole$ 50 mM, $PLP$ 100 µM, *lysozyme* 0.2 µg/ml, protease

inhibitor cocktail EDTA-free, pH 7.4). After lysis, samples were ultracentrifugated at 70300 g for 20 minutes (Beckman-Coulter Optima LE-80K ultracentrifuge). The soluble fraction was loaded into an immobilized metal ion affinity chromatography (IMAC) step (His GravitTrap; GE-Healthcare) in order to bind the His-tagged AGT protein to the stationary phase (Sepharose$^{TM}$). The matrix was washed with binding buffer and the protein was eluted by the addition of the elution buffer ($NaH_2PO_4$ 20 mM, $NaCl$ 200 mM, $Imidazole$ 500 mM, pH 7.4). Samples were buffer exchanged by the use of a desalting step (PD10 Sephadex$^{TM}$ G-25; GE-Healthcare) in order to remove the imidazole. Next, samples were eluted with working buffer ($Na-Hepes$ 20 mM, $NaCl$ 200 mM, pH 7.4).

The dimeric form of AGT proteins was isolated by a size-exclusion chromatography step (HiLoad$^{TM}$ 16/60 Superdex$^{TM}$200 prep grade column; GE-Healthcare) in working buffer. The selected fractions (15 ml centered at retention volume of 81-84 ml) were collected and concentrated (Vivaspin$^{TM}$20 - 30000 MW cut-off). Samples were preserved in working buffer at concentrations ranging from 5 to 15 mg/ml using liquid nitrogen and stored at -80°C. The protein concentration was measured spectrophotometrically with a molar extinction coefficient of $\varepsilon_{280} = 1.069\, ml \cdot mg^{-1} \cdot cm^{-1}$, according to the AGT primary sequence.[175]

**Preparation of the AGT apo enzyme.** The coenzyme was removed from the holo AGT enzyme by using a modification of the protocol described by Cellini *et al.*,[62] which can be divided into two steps that can be monitored spectrophotometrically.

During the first step, the internal Schiff base between the coenzyme and the Lys209 is broken through the incubation of the holo enzyme (2 mg at 1 mg/ml) in the presence of 500 mM of L-alanine for 10 minutes at room temperature. After a concentration step, the incubation with L-alanine is repeated. In this stage the characteristic peak of the internal aldimine at 420 nm is replaced by a peak at 330 nm due to the formation of the external aldimine o Schiff base (PMP) (see Figure 9.2.1). Now that the coenzyme is non-covalently bound to the active site, it is released by successive steps of dilution and concentration (Vivaspin$^{TM}$20 - 30000 MW cut-off) in a buffer $KH_2PO_4$ 1 M pH 5.8. The low pH destabilizes the structure of the protein and the release of the coenzyme is made by displacement with the phosphate ion. At the end of this stage, the

absorbance spectra show that the peak at 330 nm is absent so the enzyme is in the apo state (see Figure 9.2.1).

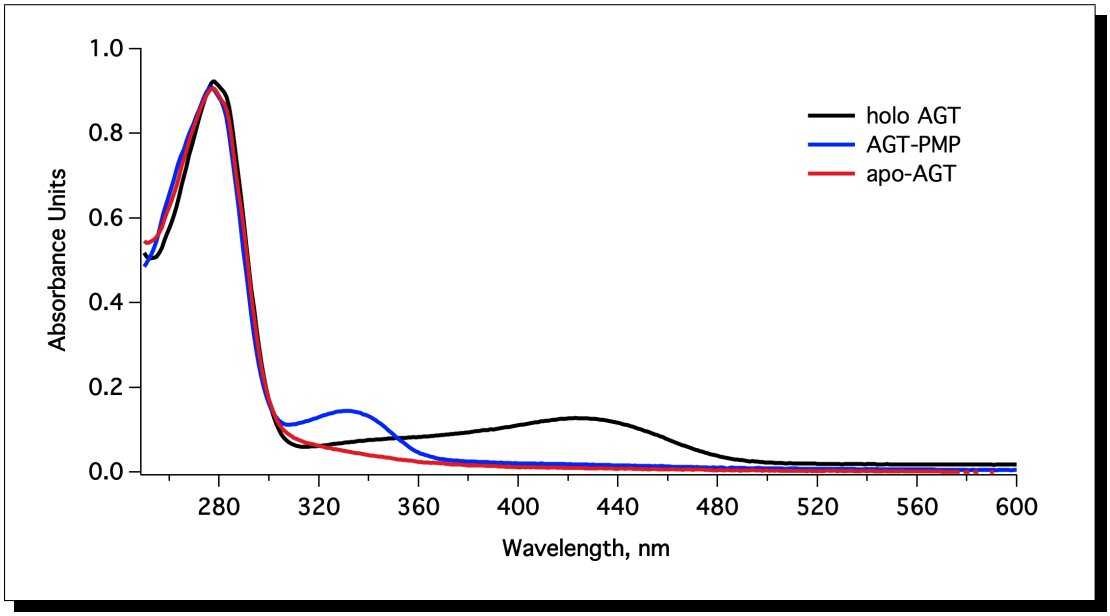


FIGURE 9.2.1. **Characteristic absorption spectra of an AGT variant through the formation of the apo enzyme.** The internal aldimine of the holo enzyme (black) has the characteristic absorption peak at 420 nm while the external aldimine (blue) presents the peak at 330 nm. Finally absorptium spectrum of the apo enzyme (red) lacks both peaks. Protein concentration 20 µM monomer.

Finally the pH is restored and the phosphate concentration is reduced by dilution-concentration steps with $KH_2PO_4$ 100 mM pH 7.4. Then, samples are buffer exchanged to buffer $Na - Hepes$ 20 mM, $NaCl$ 200 mM, pH 7.4 by a desalting step (PD10 Sephadex$^{TM}$ G-25; GE-Healthcare). In these conditions, more than 90% of the dimer is obtained as apo state. Nonetheless, the percentage of the protein lost in the process is increased in the PH1 variants due to the lower stability of these proteins. In each case the lack of aggregates in final samples were measured by evaluation of the absorbance at wavelength higher than 300 nm.

## 9.3. *C. elegans* T14D7.1 protein

**Plasmid construction.** The gene sequence of the predicted T14D7.1 protein was found indexed in the WormBase* as an orthologous to the human AGT

*International consortium of biologists and computer scientists dedicated to provide accurate, current, accessible information concerning the genetics, genomics and biology of the *C. elegans* and related nematodes (http://www.wormbase.org/).

enzyme.[103] The coding sequence was obtained from a *C. elegans* cDNA library and it was cloned into a pET-28 (Novagen) vector, by Dr Ana Calvo at the Department of Cell Biology (Yale University, CT, USA). This plasmid includes a His-tag sequence in the N-terminal with a cleavage site for thrombin and kanamycin resistance.

**Expression.** *E. coli* BL21 cells containing plasmid encoding T14D7.1 protein were grown in the presence of kanamycin (30 µg/ml) at 37°C until an absorbance of 0.6 at 600 nm. The induction was performed with IPTG (0.5 mM) during 16 h at room temperature. After the induction, cells were harvested and frozen following the same protocol and buffers as for human AGT.

**Purification.** T14D7.1 protein was purified following the same purification process that was used for human AGT (see previous section).

The protein concentration was measured spectrophotometrically with a molar extinction coefficient of $\varepsilon_{280} = 0.763\, ml \cdot mg^{-1} \cdot cm^{-1}$, according to the primary sequence of T14D7.1 protein.[175]

**Preparation of the T14D7.1 apo enzyme.** The procedure applied was based on the same protocol that was used for the human AGT variants (see previous section), including some different pH conditions (down to pH 5.5) and incubation times (from minutes to hours).

CHAPTER 10

# *In vitro* characterization of recombinant proteins

## 10.1. Buffers

All assays were performed in $Na-Hepes$ 20 mM, $NaCl$ 200 mM pH 7.4 at 25°C. Otherwise, it is explicitly indicated.

## 10.2. Spectroscopic analysis

**Absorption spectroscopy.** Absorption spectra were collected in a UV-visible spectrophotometer (Agilent 8453) using a 3-mm path length cuvette with a protein concentration of 20 µM in monomer.

**Circular dichroism spectroscopy.** The Near UV-visible circular dichroism spectra were measured in a Jasco J-710 spectropolarimeter using a protein concentration of 20 µM (in subunit) in a 5-mm path length cuvette. Spectra were collected at 50 nm/min and were averaged from 8 scans.

**Dynamic light scattering.** The evaluation of the hydrodynamic diameter of the holo variants was performed using a protein concentration of 5 µM (in subunit) in the presence of PLP 50 µM in a Zetasizer Nano ZS (Malvern Inc.) by using the Stokes-Einstein equation assuming a spherical shape. Measurements were made using a 3-mm path length cuvette and were the results of 3 independent measurements.

## 10.3. PLP affinity

The coenzyme affinity of human AGT variants was evaluated according to the equilibrium process in which each monomer of the protein is able to bind one molecule of the coenzyme. The equilibrium constant ($K_{D,PLP}$) of this process can be described as follows:

$$AGT_{monomer} - PLP \overset{K_{D,PLP}}{\leftrightarrow} AGT_{monomer} + PLP \tag{10.3.1}$$

(10.3.2) $$K_{D,PLP} = \frac{[AGT_{monomer}] \cdot [PLP]}{[AGT_{monomer} - PLP]}$$

The equilibrium constant was evaluated by using two different approaches:[62] the equilibrium approach is based in the establishment of the equilibrium between the coenzyme bound to the active site and the free coenzyme; alternatively the kinetic approach is based in the evaluation of the kinetic binding rates. Both approaches were applied to the human AGT-Ma, while for the other variants, the equilibrium or the kinetic approach was used according to the stability of the enzymes. Those variants that displayed aggregation events (detected by the decrease of fluorescence over time in the absence of ligand) during the incubation, were analyzed by the kinetic approach, while the equilibrium approach was used for the rest of variants.

For both approaches, the coenzyme binding was measured by quenching effect on the intrinsic Trp fluorescence due to the binding process. The intrinsic fluorescence was measured using excitation and emission wavelengths of 280 and 340 nm respectively with slits of 5 nm in a Cary Eclipse spectrofluorimeter (Agilent Technologies).

PLP stocks were freshly prepared for each assay in $Na{-}Hepes$ 20 mM, $NaCl$ 200 mM pH 7.4. PLP concentration was measured spectrophotometrically using a molar extinction coefficient of $\varepsilon_{388} = 4900\, M^{-1} \cdot cm^{-1}$.[176]

**Equilibrium approach.** The apo state of the protein (100 nM monomer) was incubated in the presence of PLP (0 - 10 µM) for 4 hours at 30°C. The stability of the fluorescence signal was checked to confirm that the equilibrium was reached. Fluorescence data were interpreted as follows:

(10.3.3) $$F = F_{apo} + (F_{holo} - F_{apo}) \cdot \alpha$$

where $F$ is the fluorescence intensity in the equilibrium for each total PLP concentration, $F_{holo}$ and $F_{apo}$ are the intrinsic fluorescence of the holo and apo state respectively, and $\alpha$ is the reaction progress and is defined as:

(10.3.4)
$$\alpha = \frac{[E]_t + [PLP]_t + K_{D,PLP} - \sqrt{([E]_t + [PLP]_t + K_{D,PLP})^2 - 4 \cdot [E]_t \cdot [PLP]_t}}{2 \cdot [E]_t}$$

where $[E]_t$ and $[PLP]_t$ are the total concentration of protein (monomer) and coenzyme respectively.

**Kinetic approach.** This approach is based on the establishment of a pseudo first-order process by keeping a low apo-protein:coenzyme ratio (i.e., $[apo - protein] \ll [PLP]$). Under these conditions, the binding kinetics of the coenzyme follows:

$$F(t) = F_0 + A \cdot exp(-k_{app} \cdot t) \tag{10.3.5}$$

where the fluorescence $F(t)$ is described as a function of time (t), $F_0$ is the fluorescence at $t = 0$ (apoenzyme), $A$ is the total quenching caused by the ligand and $k_{app}$ is the pseudo first-order constant of the process. The kinetic constants of the association ($k_{on}$, $s^{-1} \cdot M^{-1}$) and dissociation ($k_{off}$, $s^{-1}$) can be estimated from the linear regression of the $[PLP]$ vs $k_{app}$ according to the following expression:

$$k_{app} = k_{on} \cdot [PLP] + k_{off} \tag{10.3.6}$$

The equilibrium constant of the dissociation process can be evaluated as follows:

$$K_{D,PLP} = \frac{k_{off}}{k_{on}} \tag{10.3.7}$$

## 10.4. Enzyme activity measurements

The overall transamination activity of human AGT variants and the T14D7.1 protein, was analyzed according to a double displacement mechanism:

$$L - alanine + Glyoxylate \overset{AGT}{\leftrightarrow} Pyruvate + Glycine \tag{10.4.1}$$

This mechanism involves two substrates and two products, and it can also be described as the sum of two semi-reactions:

$$AGT_{PLP} + L - alanine \longleftrightarrow [AGT : L - ala] \rightarrow AGT_{PMP} + Pyruvate \tag{10.4.2}$$

(10.4.3) $$AGT_{PMP} + Glyoxylate \longleftrightarrow [AGT : Glyox] \rightarrow AGT_{PLP} + Glycine$$

The kinetic parameters of the protein activity were obtained from the measurement of the initial velocity ($\upsilon_0$) at different concentrations of substrates. A double displacement model was fit to the data according to the following expression:

(10.4.4) $$\upsilon_0 = \frac{V_{max} \cdot [S_{Ala}] \cdot [S_{Glyox}]}{K_{M,Ala} \cdot [S_{Glyox}] + K_{M,Glyox} \cdot [S_{Ala}] + [S_{Ala}] \cdot [S_{Glyox}]}$$

where $V_{max}$ is the maximum reaction rate, and $[S_{Ala}]$, $[S_{Ala}]$, $K_{M,Ala}$ and $K_{M,Glyox}$ are the substrate concentrations and Michaelis-Menten constants for L-alanine and glyoxylate respectively.

The assay was performed in two consecutive steps. In the **first step** the transamination reaction was carried out. The reaction was done at 37°C with a protein concentration of 5 µg/ml* in the presence of PLP (150 µM) in $Na-phosphate$ 100 mM pH 8 buffer. The glyoxylate concentrations were set from 0.25 mM to 2 mM, and the reactions were triggered by the addition of $L-alanine$ at concentrations from 0 to 100 mM. The time of reaction was set at 2 minutes. Under these conditions, the assay was linear to protein concentration and time and no more than 10% of substrates were consumed. After the time of reaction, the transamination was stopped by the addition of trichloroacetic acid (100 µl 50% w/v) and incubation on ice for 20 minutes. Then, supernatants were recovered after a centrifugation step and stored at -20°C.

In the **second step**, the formation of pyruvate was measured from the reaction mixtures by a coupled NADH:Lactate dehydrogenase assay (see Eq 10.4.5). These assays were performed in a 96 well plate (NanoQuant InfiniteM200 - Tecan) at 37°C by monitoring the reduction of the absorption at 340 nm, in the presence of NADH 0.7 mM and Tris-HCl 1 M pH 8.

(10.4.5) $$Pyruvate + NADH \overset{LDH}{\leftrightarrow} Lactate + NAD^+$$

*Due to the lower activity of the human AGT variant H83R-Mi, the total amount of protein used in the assay, was increased up to 15 µg/ml for a reaction time of 10 minutes. Conversely, the concentration was decreased (2.5 µg/ml) for the T14D7.1 protein while the time of reaction was kept in two minutes.

Transamination assays were done at least by duplicate, while the coupled assay was evaluated by triplicate. The pyruvate concentration was evaluated by performing calibration curves at known pyruvate concentrations for each set of experimental data.

Specific activities were measured at glyoxylate concentration of 10 mM and L-alanine concentration of 100 mM following the same procedure. The dependence of specific activity on pH was performed replacing Hepes with sodium phosphate (for pH from 6 to 8) and sodium acetate (for pH from 3.5 to 5.5) buffers during the assay. The time of the transamination reaction was set at 1.5 minutes (pH 7-8), 4 minutes (pH 6-6.5) and 10 minutes (pH 3.5-5.5). The dependence of specific activity on temperature was measured at 15, 20, 25, 30 and 37°C with times of reactions of 2 and 10 minutes.

## 10.5. Differential scanning calorimetry

The thermal unfolding of the enzymes was evaluated in a capillary VP-DSC differential scanning calorimeter (MicroCal, GE, Healthcare) with a cell volume of 135 µl at scan rates ranging from 1–3°C/$min$. Temperature ranges were usually 10-80°C (apo proteins) or 20-110°C (holo proteins), with a protein monomer concentration of 5 µM in the presence of $PLP$ 25 µM, $NaCl$ 200 mM, $Na-Hepes$ 20 mM pH 7.4 unless otherwise indicated. Samples were freshly prepared in filtered buffers and centrifugated prior to the experiments.

Thermal denaturation for all variants was irreversible, kinetically controlled and protein-concentration independent.[7] The analysis was done using a two state irreversible model:[121]

$$N \xrightarrow{k} F$$

where the native state $N$ is irreversibly unfolded into a final state $F$ and $k$ is a kinetic constant that changes with temperature according to the Arrhenius equation:

$$k = A \cdot exp\left(-\frac{E_a}{R \cdot T}\right) \tag{10.5.1}$$

where $R$ is the ideal gas constant, $A$ is the pre-exponential parameter and $E_a$ is the activation energy.

The output of the DSC experiment (apparent heat capacity – $C_{app}$; see Figure 10.5.1-a) can be analyzed as a function of the temperature as follows:

$$Cp_{app} = Cp_{pre} + (Cp_{post} - Cp_{pre}) \cdot (1 - x_N) - \Delta H \cdot \left(\frac{dx_N}{dT}\right) \tag{10.5.2}$$

where $x_N$ is the molar fraction of the native state, $\Delta H$ is the enthalpy of unfolding and finally $Cp_{pre}$ and $Cp_{post}$ are the baseline of the pre- and post-transition as a linear function of the temperature respectively.

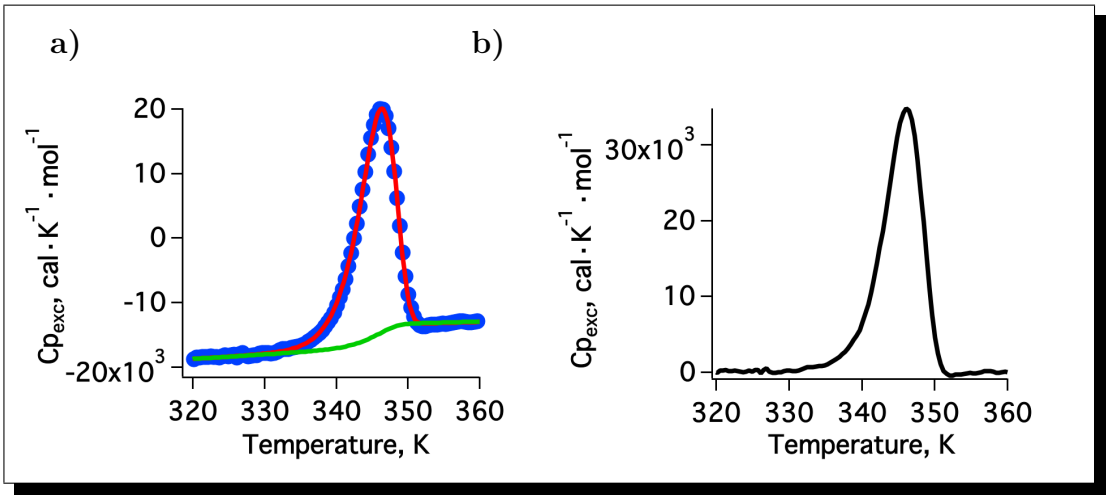


FIGURE 10.5.1. **Output of a DSC experiment for F152I-Mi analyzed by a two-state kinetic model. a)** experimental temperature-dependence of the apparent heat capacity (blue markers). The red line is the fit of the two-state irreversible model to data and the green line corresponds to the chemical baseline; **b)** the excess heat capacity is calculated taking the chemical baseline as a reference.

The first two summands of the the Eq 10.5.2 represent the chemical baseline (green line in Figure 10.5.1-a), a smooth curve that connects the apparent heat capacities of the native ($Cp_{pre}$) and denatured ($Cp_{post}$) states of the protein. The third summand represents the peak or transition which is attributed to the heat absorption associated with the denaturation of the protein. The subtraction of the chemical baseline from the apparent heat capacity ($Cp_{app}$) allows us the evaluation of the excess heat capacity ($Cp_{exc}$; see Figure 10.5.1-b) as the apparent heat capacity of the protein measured from an adequate reference level (the chemical baseline):

$$Cp_{exc} = -\Delta H \cdot \left(\frac{dx_N}{dT}\right) \tag{10.5.3}$$

**Two-state irreversible model with first order kinetics.** Therefore, from Eq 10.5.3 we can particularize an expression for a first order kinetic process, obtaining the following expression:[121]

(10.5.4)
$$Cp_{exc} = -\Delta H \cdot \left(\frac{-E_a}{R \cdot T_m^2}\right) \cdot exp\left(\frac{-E_a \cdot (T - T_m)}{R \cdot T_m^2}\right) \cdot exp\left(-exp\left(\frac{-E_a \cdot (T - T_m)}{R \cdot T_m^2}\right)\right)$$

where $Cp_{exc}$ is the excess heat capacity in reference with the chemical baseline, $T_m$ is the temperature corresponding to the maximum of the excess heat capacity, $E_a$ is the activation energy and $\Delta H$ is the denaturation enthalpy per mol of protein of the thermal unfolding.

The proper description of the thermal unfolding by the two state irreversible model, was tested by the evaluation of the activation energy with the application of the consistency tests:[121]

- Test A: the kinetic constant values evaluated with Eq 10.5.5 and the linearization of the Arrhenius equation allow the plot of the $ln(k)$ vs $1/T_m$, from where the activation energy can be obtained from the slope.
- Test B: the dependence of the $T_m$ parameter with the scan rate ($\upsilon$) can be evaluated with the linearization of the Eq 10.5.6, where the activation energy can be again evaluated from the slope of the plot $ln(\upsilon/T_m^2)$ vs $1/T_m$.
- Test C: this test is based in the dependence of the heat involved at any temperature according to the Eq 10.5.7. The plot of the $ln(ln(Q_T/(Q_T - Q)))$ vs $1/T$ allows the evaluation of the activation energy from the slope.

(10.5.5)
$$k = \frac{\upsilon \cdot Cp_{exc}}{(Q_T - Q)}$$

(10.5.6)
$$\frac{\upsilon}{T_m^2} = \frac{A \cdot R}{E_a} \cdot exp\left(\frac{-E_a}{R \cdot T_m}\right)$$

(10.5.7)
$$ln\left(ln\left(\frac{Q_T}{Q_T - Q}\right)\right) = \frac{E_a}{R} \cdot \left(\frac{1}{T_m} - \frac{1}{T}\right)$$

where $Q_T$ corresponds to the total heat involved upon the unfolding process and $Q$ corresponds to the heat involved at each temperature.

**Two-state irreversible model with non first-order kinetics.** The oligomerization change between the native protein and the transition state for the rate-limiting step upon the thermal unfolding, determines the reaction order $(1/\mu)$ of the process. The possible dissociation of the dimer into monomers prior to the transition state can be evaluated in the Eq 10.5.3 obtaining the following expression:[122]

(10.5.8)

$$Cp_{exc} = \frac{\Delta H \cdot E_a}{R \cdot T_m^2} \cdot exp\left(\frac{(T - T_m) \cdot E_a}{R \cdot T_m^2}\right) \cdot \left(1 + \frac{1-\mu}{\mu} \cdot exp\left(\frac{(T - T_m) \cdot E_a}{R \cdot T_m^2}\right)\right)^{\frac{1}{\mu - 1}}$$

where $\mu$ is the dissociation grade of the native state to the transition state (where $\mu = 1$ represents no dissociation and $\mu = 2$ represents a total dissociation).

This model is based on the condition that the concentration of the unfolded state $(U)$ must be much lower than the native state $(N)$, therefore for fitting purposes, the range of temperatures must be constrained to $T < T$' to avoid inconsistencies during the fitting. $T$' is provided by:

(10.5.9)

$$T' = T_m + \frac{R \cdot T_m^2}{E_a} \cdot ln\left(\frac{\mu}{\mu - 1}\right)$$

## 10.6. Mutational effects on the activation energetics

Mutational effects on the activation free energy $(\Delta\Delta G^{\neq})$, enthalpy $(\Delta\Delta H^{\neq})$ and entropy $(\Delta\Delta S^{\neq})$ were determined on the basis of the transition state theory.[108] The effects of mutations on activation energetic parameters were determined using the following expressions:[166]

(10.6.1)

$$\Delta\Delta G^{\neq}(T) = -R \cdot T \cdot ln\left(\frac{k_T\,(mutant)}{k_T\,(AGT - Ma)}\right)$$

(10.6.2)

$$\Delta\Delta H^{\neq}(T) = E_a\,(mutant) - E_a\,(AGT - Ma)$$

(10.6.3)

$$-T \cdot \Delta\Delta S^{\neq}(T) = \Delta\Delta\, G^{\neq}(T) - \Delta\Delta H^{\neq}(T)$$

These parameters were evaluated at the physiological temperature of 37°C using the kinetic constants of the thermal unfolding extrapolated to this temperature.

## 10.7. Dependence of the activation free energy of unfolding on urea concentration ($m^{\neq}$)

The denaturant concentration dependence of the activiation free-energy changes ($m^{\neq}$) is defined as following:

$$m^{\neq} = -\left(\frac{\delta \Delta G^{\neq}}{\delta\,[urea]}\right) = R \cdot T \cdot \left(\frac{\delta ln(k)}{\delta\,[urea]}\right) \tag{10.7.1}$$

where $\Delta G^{\neq}$ is the activation free energy and $k$ is the first-order rate constant.

**Kinetic $m$ values for irreversible thermal denaturation.** Instead of evaluating the kinetic $m$ values ($m^{\neq}$) from rate constant values ($k$), it is more convenient to derive them from the urea dependence of the transition temperature ($T_m$) and activation energy ($E_a$) values[166] because they are accurately determined by DSC. Therefore, the kinetic $m$ was evaluated according to the following expression:

$$m^{\neq} = -\frac{E_a}{T_m}\left(\frac{dT_m}{d\,[urea]}\right) - R \cdot T_m \cdot \left(\frac{d\,ln\,(E_a/R \cdot T_m^2)}{d\,[urea]}\right) \tag{10.7.2}$$

The derivatives in Eq 10.7.2 are available from experimental urea concentration dependencies of $T_m$ and $E_a$, for a given scan rate. Plots of $T_m$ *vs* $[urea]$ were highly linear while $ln\,(E_a/R \cdot T_m^2)$ *vs* $[urea]$ show some curvature. Nonetheless the second-term in the right-hand-side of Eq 10.7.2 is almost negligible compared to the first-one.

## 10.8. *Unfolding* and *solvation barrier* contribution to activation free enthalpy and entropy

The decomposition of the activation free enthalpy ($\Delta H^{\neq}$) into the unfolding ($\Delta H_{unf}$) and solvation barrier ($\Delta H^{*}$) contributions was determined by the following expressions:[18,165]

$$\Delta H^{\neq} = \Delta H^{*} + \Delta H_{unf} \tag{10.8.1}$$

$$\Delta H_{unf} = \Delta H \cdot \left( \frac{m^{\neq}}{m^{eq}} \right) \tag{10.8.2}$$

$$\Delta H^{*} = E_a - \Delta H_{unf} \tag{10.8.3}$$

where $\Delta H$ is the calorimetric enthalpy derived from the analysis of the DSC transition, and the ratio $m^{\neq}/m^{eq}$ reflect the degree of unfolding in transition state as estimated from the kinetic ($m^{\neq}$) and equilibrium[167] ($m^{eq}$) $m$ values.

The changes in total accessible area ($\Delta ASA^{\neq}$) between transition and native state were calculated from the unfolding and solvation barrier contribution to the activation enthalpy:[133,165,177]

$$\Delta ASA^{\neq} = \Delta ASA_{unf} + \Delta ASA^{*} \tag{10.8.4}$$

$$\Delta ASA_{unf} = \left( \frac{m^{\neq}}{m^{eq}} \right) \cdot \Delta ASA \tag{10.8.5}$$

$$\Delta H^{*} = \beta \cdot ASA^{*} \tag{10.8.6}$$

where $\beta = 24\, J \cdot mol^{-1} \cdot \text{Å}^2$ according to the parametrization of enthalpy changes for the breakup of internal interactions.[177]

The unfolding and solvation barrier components to the activation entropy ($\Delta S^{\neq} = \Delta S^{*} + \Delta S_{unf}$) were determined by invoking the existence of a enthalpy-entropy compensation for the solvation barrier contributions.[165] The rate constant for unfolding can be written as:

$$k = k_0 \cdot exp\left( \frac{-\Delta G^{\neq}}{R \cdot T} \right) = k_0 \cdot exp\left( \frac{-\Delta H^{\neq}}{R \cdot T} \right) \cdot exp\left( \frac{\Delta S^{\neq}}{R} \right) \tag{10.8.7}$$

and rearranged as:

$$R \cdot lnk + \frac{\Delta H^{\neq}}{T} - \Delta S_{unf} = R \cdot lnk_0 + \Delta S^{*} \tag{10.8.8}$$

where the unfolding component can be determined as:[133]

(10.8.9) $$\Delta S_{unf} = \left(0.0088 + 0.058 \cdot ln\left(\frac{T}{333.15}\right)\right) \cdot \left(\frac{m^{\neq}}{m^{eq}}\right) \cdot N_r$$

The left-hand side of the Eq 10.8.8 can be evaluated for a given temperature providing an estimated value for $R \cdot lnk_0 + \Delta S^*$. In order to separate these two terms the proposed enthalpy-entropy compensation for the solvation barrier is invoked:

(10.8.10) $$\Delta H^* = T_0 \cdot \Delta S^*; \qquad \Delta S^* = \Delta H^*/T_0$$

where $T_0$ is the compensation temperature that can be evaluated (434.78 K for this set of proteins). A plot of the left-hand-side of Eq 10.8.8 *vs* $\Delta H^*$ should be linear with a slope equal to $1/T_0$ and with an intercept equal to $R{\cdot}lnk_0$. Using the Eq 10.8.10 and the value of $T_0$, the solvation-barrier entropic contributions can be evaluated, and therefore the activation entropy can be decomposed.

## 10.9. Electrostatic and Accessible surface area calculations

**Accessible surface area.** The accessible surface area (ASA) calculations were performed using a home-built software, kindly provided by Prof. Jose Manuel Sanchez-Ruiz. By the use of atomic coordinates of the dimeric human AGT-Ma,[50] I340M-Ma, HEM-Ma and RHEAM-Ma[†], the ASA was evaluated with a radius of 1.4 Å and the Chothia set for the protein atoms.[178]

**Electrostatic interactions.** The evaluation of the energy of charge-charge interactions ($E_{q-q}$) were performed using the solvent-accessibility-corrected Tanford-Kirkwood model described previously by Ibarra-Molero *et al.*[151] The calculations were performed using a home-built software, kindly provided by Prof. Jose Manuel Sanchez-Ruiz, at the simulated conditions of pH 7.4 and 200 mM of ionic strength in order to mimic the conditions used in the DSC experiments. The output of these analyses provide the energy of charge-charge interactions of a given ionizable residue with all the ionizable residues in the protein dimer ($E_{q-q}$).

[†]Through the collaboration on this project, the crystallographic structures of I34M-Ma, HEM-Ma and RHEAM-Ma were determined by Dr Armando Albert at Department of Crystallography and Structural Biology, IQF-Rocasolano, CSIC.

### 10.10. Urea-induced global unfolding

Equilibrium unfolding curves were measured at 25°C after an overnight incubation of the holo protein (5 µM) at different urea concentrations in presence of 1 mM of $Tris(2-carboxyethyl)phosphine$ (TCEP) in a *Hepes* 20 mM, $NaCl$ 200mM and pH 7.4 buffer. The degree of unfolding was monitored by measuring ellipticity at 222 nm in a Jasco J-710 spectropolarimeter using 1-mm path-length cuvettes thermostatized with a Peltier element. The data were analyzed by a two-state model ($N \Leftrightarrow U$) according to the following expression:

$$Y = \frac{Y_N + \alpha_N \cdot [urea] + (Y_U + \alpha_U \cdot [urea]) \cdot exp\left(\frac{-m(C_m - [urea])}{R \cdot T}\right)}{1 + exp\left(\frac{-m(C_m - [urea])}{R \cdot T}\right)} \quad (10.10.1)$$

where $Y$ is the measured signal, $Y_N$ and $Y_U$ the signal values of the native and fully unfolded states, $\alpha_N$ and $\alpha_U$ represent the dependence of the signal of native and unfolded states on urea concentration respectively and $C_m$ is the midpoint of urea required for unfolding and $m$ is the slope of the curve at $C_m$.

Kinetics of protein unfolding was also monitored by measuring ellipticity at 222 nm using a thermostatized 1-mm path-length cuvettes. Urea solutions were prepared daily in *Hepes* 20 mM, $NaCl$ 200mM and pH 7.4 and urea concentration was measured by refractive index.[179] Proteins were incubated at the target temperature (25-70°C) for 10 min in concentrated protein solutions (routinely AGT concentration of 50 µM monomer with 10-fold excess of PLP). Then, solutions were diluted 1:10, manually mixed with urea solutions (7-9 M) containing a final concentration of TCEP 1 mM, and the time dependence of ellipticity at 222 nm was registered. Data were corrected by the dead time of experiment (15-20 sec). All the kinetic traces were described by a single exponential function that yields an apparent rate constant ($k_{app}$). The dependence of this constant on urea concentration was found to be linear and was analyzed by the following expression:

$$ln(k_{app}) = ln(k_{unf-0M}) + \frac{m^{\neq}}{R \cdot T} \cdot [urea] \quad (10.10.2)$$

where $k_{unf-0M}$ is the rate constant extrapolated to the absence of urea and $m^{\neq}$ is the unfolding kinetic $m$ value. Finally, the temperature dependence of the rate constant was analyzed using the Arrhenius equation (see Eq 10.5.1).

# Part 6

# Bibliography

# Bibliography


[1] Danpure, C. J. Peroxisomal alanine:glyoxylate aminotransferase and prenatal diagnosis of primary hyperoxaluria type 1. *Lancet* **2**, 1168 (1986).

[2] Purdue, P. E., Takada, Y. & Danpure, C. J. Identification of mutations associated with peroxisome-to-mitochondrion mistargeting of alanine/glyoxylate aminotransferase in primary hyperoxaluria type 1. *J Cell Biol* **111**, 2341–51 (1990).

[3] Danpure, C. J. *et al.* Enzymological and mutational analysis of a complex primary hyperoxaluria type 1 phenotype involving alanine:glyoxylate aminotransferase peroxisome-to-mitochondrion mistargeting and intraperoxisomal aggregation. *Am J Hum Genet* **53**, 417–32 (1993).

[4] Salido, E., Pey, A. L., Rodriguez, R. & Lorenzo, V. Primary hyperoxalurias: disorders of glyoxylate detoxification. *Biochim Biophys Acta* **1822**, 1453–64 (2012).

[5] Cochat, P. *et al.* Primary hyperoxaluria type 1: indications for screening and guidance for diagnosis and treatment. *Nephrol Dial Transplant* **27**, 1729–36 (2012).

[6] Williams, E. L. *et al.* Primary hyperoxaluria type 1: update and additional mutation analysis of the agxt gene. *Hum Mutat* **30**, 910–7 (2009).

[7] Pey, A. L., Salido, E. & Sanchez-Ruiz, J. M. Role of low native state kinetic stability and interaction of partially unfolded states with molecular chaperones in the mitochondrial protein mistargeting associated with primary hyperoxaluria. *Amino Acids* **41**, 1233–45 (2011).

[8] Wickner, S., Maurizi, M. R. & Gottesman, S. Posttranslational quality control: folding, refolding, and degrading proteins. *Science* **286**, 1888–93 (1999).

[9] Frydman, J. Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* **70**, 603–47 (2001).

[10] Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–32 (2011).

[11] Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* **82**, 323–55 (2013).

[12] Markaki, M. & Tavernarakis, N. Modeling human diseases in caenorhabditis elegans. *Biotechnol J* **5**, 1261–76 (2010).

[13] Calamini, B. *et al.* Small-molecule proteostasis regulators for protein conformational diseases. *Nat Chem Biol* **8**, 185–96 (2012).

[14] Lehmann, M. & Wyss, M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr Opin Biotechnol* **12**, 371–5 (2001).

[15] Salido, E. C. *et al.* Alanine-glyoxylate aminotransferase-deficient mice, a model for primary hyperoxaluria that responds to adenoviral gene transfer. *Proc Natl Acad Sci U S A* **103**, 18249–54 (2006).

[16] Margaritis, P. *et al.* Catalytic domain modification and viral gene delivery of activated factor vii confers hemostasis at reduced expression levels and vector doses in vivo. *Blood* **117**, 3974–82 (2011).

[17] Godoy-Ruiz, R. *et al.* Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencies of amino acid occurrences in sequence alignments. *J Mol Biol* **362**, 966–78 (2006).

[18] Rodriguez-Larrea, D., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. Energetic and structural consequences of desolvation/solvation barriers to protein folding/unfolding assessed from experimental unfolding rates. *Biophys J* **91**, L48–50 (2006).

[19] Sanchez-Ruiz, J. M. Protein kinetic stability. *Biophys Chem* **148**, 1–15 (2010).

[20] Danpure, C. J. *The Online Metabolic and Molecular Bases of Inherited Disease*, chap. 133 - Primary hyperoxaluria, 1–116 (McGraw-Hill, 2006).

[21] Harambat, J., Fargue, S., Bacchetta, J., Acquaviva, C. & Cochat, P. Primary hyperoxaluria. *Int J Nephrol* **2011**, 864580 (2011).

[22] Hoppe, B., Beck, B. B. & Milliner, D. S. The primary hyperoxalurias. *Kidney Int* **75**, 1264–71 (2009).

[23] Wanders, R. J. A. & Waterham, H. R. Biochemistry of mammalian peroxisomes revisited. *Annu Rev Biochem* **75**, 295–332 (2006).

[24] Vamecq, J., Cherkaoui-Malki, M., Andreoletti, P. & Latruffe, N. The human peroxisome in health and disease: The story of an oddity becoming a vital organelle. *Biochimie* **98C**, 4–15 (2014).

[25] Santana, A., Torres, A. & Salido, E. [molecular pathology of primary hyperoxaluria]. *Nefrologia* **23 Suppl 1**, 90–7 (2003).

[26] Beck, B. B. & Hoppe, B. Is there a genotype-phenotype correlation in primary hyperoxaluria type 1? *Kidney Int* **70**, 984–6 (2006).

[27] Hoppe, B. Evidence of true genotype-phenotype correlation in primary hyperoxaluria type 1. *Kidney Int* **77**, 383–5 (2010).

[28] Cochat, P. *et al.* Epidemiology of primary hyperoxaluria type 1. société de néphrologie and the société de néphrologie pédiatrique. *Nephrol Dial Transplant* **10 Suppl 8**, 3–7 (1995).

[29] Danpure, C. J. Molecular etiology of primary hyperoxaluria type 1: new directions for treatment. *Am J Nephrol* **25**, 303–10 (2005).

[30] Gibbs, D. A. & Watts, R. W. Biochemical studies on the treatment of primary hyperoxaluria. *Arch Dis Child* **42**, 505–8 (1967).

[31] Monico, C. G., Rossetti, S., Olson, J. B. & Milliner, D. S. Pyridoxine effect in type i primary hyperoxaluria is associated with the most common mutant allele. *Kidney Int* **67**, 1704–9 (2005).

[32] Hoyer-Kuhn, H. *et al.* Vitamin b6 in primary hyperoxaluria i: First prospective trial after 40 years of practice. *Clin J Am Soc Nephrol* (2014).

[33] Coulter-Mackie, M., White, C., Hurley, R., Chew, B. & Lange, D. *Gene Reviews*, chap. Primary Hyperoxaluria Type 1 (2002), wa): university of washington, seattle; 1993-2013 edn.

[34] Harambat, J. *et al.* Genotype-phenotype correlation in primary hyperoxaluria type 1: the p.gly170arg agxt mutation is associated with a better outcome. *Kidney Int* **77**, 443–9 (2010).

[35] Hatch, M., Gjymishka, A., Salido, E. C., Allison, M. J. & Freel, R. W. Enteric oxalate elimination is induced and oxalate is normalized in a mouse model of primary hyperoxaluria following intestinal colonization with oxalobacter. *Am J Physiol Gastrointest Liver Physiol* **300**, G461–9 (2011).

[36] Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W. & Balch, W. E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu Rev Biochem* **78**, 959–91 (2009).

[37] Strittmatter, L. *et al.* Clybl is a polymorphic human enzyme with malate synthase and beta-methylmalate synthase activity. *Hum Mol Genet* (2014).

[38] Holmes, R. P. & Assimos, D. G. Glyoxylate synthesis, and its modulation and influence on oxalate synthesis. *J Urol* **160**, 1617–24 (1998).

[39] Ichiyama, A. Studies on a unique organelle localization of a liver enzyme, serine:pyruvate (or alanine:glyoxylate) aminotransferase. *Proc Jpn Acad Ser B Phys Biol Sci* **87**, 274–86 (2011).

[40] Takayama, T. *et al.* Control of oxalate formation from l-hydroxyproline in liver mitochondria. *J Am Soc Nephrol* **14**, 939–46 (2003).

[41] Coulter-Mackie, M. B. 4-hydroxyproline metabolism and glyoxylate production: A target for substrate depletion in primary hyperoxaluria? *Kidney Int* **70**, 1891–3 (2006).

[42] Thompson, G. N., Purkiss, P. & Danpure, C. J. The subcellular metabolism of glyoxylate in primary hyperoxaluria type 1: the relationship between glycine production and oxalate overproduction. *J Inherit Metab Dis* **11 Suppl 2**, 212–4 (1988).

[43] Poore, R. E., Hurst, C. H., Assimos, D. G. & Holmes, R. P. Pathways of hepatic oxalate synthesis and their regulation. *Am J Physiol* **272**, C289–94 (1997).

[44] Baker, P. R. S., Cramer, S. D., Kennedy, M., Assimos, D. G. & Holmes, R. P. Glycolate and glyoxylate metabolism in hepg2 cells. *Am J Physiol Cell Physiol* **287**, C1359–65 (2004).

[45] Rokka, A. *et al.* Pxmp2 is a channel-forming protein in mammalian peroxisomal membrane. *PLoS One* **4**, e5090 (2009).

[46] Caplin, B. *et al.* Alanine-glyoxylate aminotransferase-2 metabolizes endogenous methylarginines, regulates no, and controls blood pressure. *Arterioscler Thromb Vasc Biol* **32**, 2892–900 (2012).

[47] Behnam, J. T., Williams, E. L., Brink, S., Rumsby, G. & Danpure, C. J. Reconstruction of human hepatocyte glyoxylate metabolic pathways in stably transformed chinese-hamster ovary cells. *Biochem J* **394**, 409–16 (2006).

[48] Thompson, J. S. & Richardson, K. E. Isolation and characterization of an l-alanine: glyoxylate aminotransferase from human liver. *J Biol Chem* **242**, 3614–9 (1967).

[49] Eliot, A. C. & Kirsch, J. F. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem* **73**, 383–415 (2004).

[50] Zhang, X. *et al.* Crystal structure of alanine:glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1. *J Mol Biol* **331**, 643–52 (2003).

[51] Danpure, C. J. Primary hyperoxaluria type 1: Agt mistargeting highlights the fundamental differences between the peroxisomal and mitochondrial protein import pathways. *Biochim Biophys Acta* **1763**, 1776–84 (2006).

[52] Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim Biophys Acta* **1814**, 1407–18 (2011).

[53] Yoshimura, T., Jhee, K. H. & Soda, K. Stereospecificity for the hydrogen transfer and molecular evolution of pyridoxal enzymes. *Biosci Biotechnol Biochem* **60**, 181–7 (1996).

[54] Motley, A. *et al.* Mammalian alanine/glyoxylate aminotransferase 1 is imported into peroxisomes via the pts1 translocation pathway. increased degeneracy and context specificity of the mammalian pts1 motif and implications for the peroxisome-to-mitochondrion mistargeting of agt in primary hyperoxaluria type 1. *J Cell Biol* **131**, 95–109 (1995).

[55] Knott, T. G. *et al.* The peroxisomal targeting sequence type 1 receptor, pex5p, and the peroxisomal import efficiency of alanine:glyoxylate aminotransferase. *Biochem J* **352 Pt 2**, 409–18 (2000).

[56] Huber, P. A. J. *et al.* Peroxisomal import of human alanine:glyoxylate aminotransferase requires ancillary targeting information remote from its c terminus. *J Biol Chem* **280**, 27111–20 (2005).

[57] Fodor, K., Wolf, J., Erdmann, R., Schliebs, W. & Wilmanns, M. Molecular requirements for peroxisomal targeting of alanine-glyoxylate aminotransferase as an essential determinant in primary hyperoxaluria type 1. *PLoS Biol* **10**, e1001309 (2012).

[58] Ghosh, D. & Berg, J. M. A proteome-wide perspective on peroxisome targeting signal 1(pts1)-pex5p affinities. *J Am Chem Soc* **132**, 3973–9 (2010).

[59] Santana, A., Salido, E., Torres, A. & Shapiro, L. J. Primary hyperoxaluria type 1 in the canary islands: a conformational disease due to i244t mutation in the p11l-containing alanine:glyoxylate aminotransferase. *Proc Natl Acad Sci U S A* **100**, 7277–82 (2003).

[60] Coulter-Mackie, M. B. & Lian, Q. Consequences of missense mutations for dimerization and turnover of alanine:glyoxylate aminotransferase: study of a spectrum of mutations. *Mol Genet Metab* **89**, 349–59 (2006).

[61] Cellini, B., Montioli, R. & Voltattorni, C. B. Human liver peroxisomal alanine:glyoxylate aminotransferase: characterization of the two allelic forms and their pathogenic variants. *Biochim Biophys Acta* **1814**, 1577–84 (2011).

[62] Cellini, B., Bertoldi, M., Montioli, R., Paiardini, A. & Borri Voltattorni, C. Human wild-type alanine:glyoxylate aminotransferase and its naturally occurring g82e variant: functional properties and physiological implications. *Biochem J* **408**, 39–50 (2007).

[63] Lumb, M. J. & Danpure, C. J. Functional synergism between the most common polymorphism in human alanine:glyoxylate aminotransferase and four of the most common disease-causing mutations. *J Biol Chem* **275**, 36415–22 (2000).

[64] Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–30 (1973).

[65] Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–8 (2002).

[66] Levinthal, C. Are there pathways for protein folding. *J. Chim. phys* **65**, 44–45 (1968).

[67] Cortazzo, P. *et al.* Silent mutations affect in vivo protein folding in escherichia coli. *Biochem Biophys Res Commun* **293**, 537–41 (2002).

[68] Hartl, F. U. & Hayer-Hartl, M. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16**, 574–81 (2009).

[69] Gershenson, A. & Gierasch, L. M. Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol* **21**, 32–41 (2011).

[70] Evans, M. S., Clarke, T. F., 4th & Clark, P. L. Conformations of co-translational folding intermediates. *Protein Pept Lett* **12**, 189–95 (2005).

[71] Feldman, D. E. & Frydman, J. Protein folding in vivo: the importance of molecular chaperones. *Curr Opin Struct Biol* **10**, 26–33 (2000).

[72] Prakash, S. & Matouschek, A. Protein unfolding in the cell. *Trends Biochem Sci* **29**, 593–600 (2004).

[73] Varshavsky, A. The ubiquitin system, an immense realm. *Annu Rev Biochem* **81**, 167–76 (2012).

[74] Toyama, B. H. & Hetzer, M. W. Protein homeostasis: live long, won't prosper. *Nat Rev Mol Cell Biol* **14**, 55–61 (2013).

[75] Hutt, D. & Balch, W. E. Cell biology. the proteome in balance. *Science* **329**, 766–7 (2010).

[76] McClellan, A. J., Tam, S., Kaganovich, D. & Frydman, J. Protein quality control: chaperones culling corrupt conformations. *Nat Cell Biol* **7**, 736–41 (2005).

[77] Bagola, K. & Sommer, T. Protein quality control: on ipods and other junq. *Curr Biol* **18**, R1019–21 (2008).

[78] Albanèse, V., Yam, A. Y.-W., Baughman, J., Parnot, C. & Frydman, J. Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. *Cell* **124**, 75–88 (2006).

[79] Wendler, P. *et al.* Atypical aaa+ subunit packing creates an expanded cavity for disaggregation by the protein-remodeling factor hsp104. *Cell* **131**, 1366–77 (2007).

[80] Olzmann, J. A., Li, L. & Chin, L. S. Aggresome formation and neurodegenerative diseases: therapeutic implications. *Curr Med Chem* **15**, 47–60 (2008).

[81] Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–9 (2003).

[82] Bachmair, A., Finley, D. & Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179–86 (1986).

[83] Morley, J. F., Brignull, H. R., Weyers, J. J. & Morimoto, R. I. The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in caenorhabditis elegans. *Proc Natl Acad Sci U S A* **99**, 10417–22 (2002).

[84] Morley, J. F. & Morimoto, R. I. Regulation of longevity in caenorhabditis elegans by heat shock factor and molecular chaperones. *Mol Biol Cell* **15**, 657–64 (2004).

[85] Carrell, R. W. & Lomas, D. A. Conformational disease. *Lancet* **350**, 134–8 (1997).

[86] Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–9 (2008).

[87] Qu, B. H., Strickland, E. H. & Thomas, P. J. Localization and suppression of a kinetic defect in cystic fibrosis transmembrane conductance regulator folding. *J Biol Chem* **272**, 15739–44 (1997).

[88] Fan, J. Q., Ishii, S., Asano, N. & Suzuki, Y. Accelerated transport and maturation of lysosomal alpha-galactosidase a in fabry lymphoblasts by an enzyme inhibitor. *Nat Med* **5**, 112–5 (1999).

[89] Gidalevitz, T., Krupinski, T., Garcia, S. & Morimoto, R. I. Destabilizing protein polymorphisms in the genetic background direct phenotypic expression of mutant sod1 toxicity. *PLoS Genet* **5**, e1000399 (2009).

[90] Lindquist, S. L. & Kelly, J. W. Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation diseases: progress and prognosis. *Cold Spring Harb Perspect Biol* **3** (2011).

[91] Teixeira-Castro, A. *et al.* Neuron-specific proteotoxicity of mutant ataxin-3 in c. elegans: rescue by the daf-16 and hsf-1 pathways. *Hum Mol Genet* **20**, 2996–3009 (2011).

[92] Diomede, L. *et al.* Expression of a2v-mutated a beta in caenorhabditis elegans results in oligomer formation and toxicity. *Neurobiol Dis* **62**, 521–32 (2014).

[93] Gidalevitz, T., Wang, N., Deravaj, T., Alexander-Floyd, J. & Morimoto, R. I. Natural genetic variation determines susceptibility to aggregation or toxicity in a c. elegans model for polyglutamine disease. *BMC Biol* **11**, 100 (2013).

[94] Petriv, O. I., Pilgrim, D. B., Rachubinski, R. A. & Titorenko, V. I. Rna interference of peroxisome-related genes in c. elegans: a new model for human peroxisomal disorders. *Physiol Genomics* **10**, 79–91 (2002).

[95] Thieringer, H., Moellers, B., Dodt, G., Kunau, W.-H. & Driscoll, M. Modeling human peroxisome biogenesis disorders in the nematode caenorhabditis elegans. *J Cell Sci* **116**, 1797–804 (2003).

[96] *WormBook: the online review of Caenorhabditis elegans biology* (2012).

[97] Kenyon, C. The nematode caenorhabditis elegans. *Science* **240**, 1448–53 (1988).

[98] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802–5 (1994).

[99] Hunt-Newbury, R. *et al.* High-throughput in vivo analysis of gene expression in caenorhabditis elegans. *PLoS Biol* **5**, e237 (2007).

[100] C. elegans Sequencing Consortium. Genome sequence of the nematode c. elegans: a platform for investigating biology. *Science* **282**, 2012–8 (1998).

[101] Boulin, T. & Hobert, O. From genes to function: the c. elegans genetic toolbox. *Wiley Interdiscip Rev Dev Biol* **1**, 114–37 (2012).

[102] Fire, A. *et al.* Potent and specific genetic interference by double-stranded rna in caenorhabditis elegans. *Nature* **391**, 806–11 (1998).

[103] Kuwabara, P. E. & O'Neil, N. The use of functional genomics in c. elegans for studying human development and disease. *J Inherit Metab Dis* **24**, 127–38 (2001).

[104] Cohen, F. E. & Kelly, J. W. Therapeutic approaches to protein-misfolding diseases. *Nature* **426**, 905–9 (2003).

[105] Silva, M. C. *et al.* A genetic screening strategy identifies novel regulators of the proteostasis network. *PLoS Genet* **7**, e1002438 (2011).

[106] Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. Forces contributing to the conformational stability of proteins. *FASEB J* **10**, 75–83 (1996).

[107] Becktel, W. J. & Schellman, J. A. Protein stability curves. *Biopolymers* **26**, 1859–77 (1987).

[108] Laidler, K. J. & King, M. C. Development of transition-state theory. *The Journal of physical chemistry* **87**, 2657–2664 (1983).

[109] Taverna, D. M. & Goldstein, R. A. Why are proteins marginally stable? *Proteins* **46**, 105–9 (2002).

[110] Plaza del Pino, I. M., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases. *Proteins* **40**, 58–70 (2000).

[111] Johnson, S. M. *et al.* Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidoses. *Acc Chem Res* **38**, 911–21 (2005).

[112] Brannigan, J. A. & Wilkinson, A. J. Protein engineering 20 years on. *Nat Rev Mol Cell Biol* **3**, 964–70 (2002).

[113] Penning, T. M. & Jez, J. M. Enzyme redesign. *Chem Rev* **101**, 3027–46 (2001).

[114] Steipe, B., Schiller, B., Plückthun, A. & Steinbacher, S. Sequence statistics reliably predict stabilizing mutations in a protein domain. *J Mol Biol* **240**, 188–92 (1994).

[115] Kimura, M. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet* **66**, 367–86 (1991).

[116] Lehmann, M. *et al.* The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng* **15**, 403–11 (2002).

[117] Lehmann, M. *et al.* The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein engineering* **15**, 403–411 (2002).

[118] Nikolova, P. V., Henckel, J., Lane, D. P. & Fersht, A. R. Semirational design of active tumor suppressor p53 dna binding domain with enhanced stability. *Proc Natl Acad Sci U S A* **95**, 14675–80 (1998).

[119] Rodriguez-Larrea, D. *et al.* Role of conservative mutations in protein multi-property adaptation. *Biochem J* **429**, 243–9 (2010).

[120] Cellini, B. *et al.* Molecular defects of the glycine 41 variants of alanine glyoxylate aminotransferase associated with primary hyperoxaluria type i. *Proc Natl Acad Sci U S A* **107**, 2896–901 (2010).

[121] Sánchez-Ruiz, J. M., López-Lacomba, J. L., Cortijo, M. & Mateo, P. L. Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. *Biochemistry* **27**, 1648–52 (1988).

[122] Sanchez-Ruiz, J. M. Theoretical analysis of lumry-eyring models in differential scanning calorimetry. *Biophys J* **61**, 921–35 (1992).

[123] Cellini, B., Lorenzetto, A., Montioli, R., Oppici, E. & Voltattorni, C. B. Human liver peroxisomal alanine:glyoxylate aminotransferase: Different stability under chemical stress of the major allele, the minor allele, and its pathogenic g170r variant. *Biochimie* **92**, 1801–11 (2010).

[124] Albert, A. *et al.* Structure of groel in complex with an early folding intermediate of alanine glyoxylate aminotransferase. *J Biol Chem* **285**, 6371–6 (2010).

[125] Cellini, B., Montioli, R., Paiardini, A., Lorenzetto, A. & Voltattorni, C. B. Molecular insight into the synergism between the minor allele of human liver peroxisomal alanine:glyoxylate aminotransferase and the f152i mutation. *J Biol Chem* **284**, 8349–58 (2009).

[126] Hayashi, H. Pyridoxal enzymes: mechanistic diversity and uniformity. *J Biochem* **118**, 463–73 (1995).

[127] di Salvo, M. L., Contestabile, R. & Safo, M. K. Vitamin b(6) salvage enzymes: mechanism, structure and regulation. *Biochim Biophys Acta* **1814**, 1597–608 (2011).

[128] Casanueva, M. O., Burga, A. & Lehner, B. Fitness trade-offs and environmentally induced mutation buffering in isogenic c. elegans. *Science* **335**, 82–5 (2012).

[129] Nollen, E. A. A. *et al.* Genome-wide rna interference screen identifies previously undescribed regulators of polyglutamine aggregation. *Proc Natl Acad Sci U S A* **101**, 6403–8 (2004).

[130] Nussbaum-Krammer, C. I. & Morimoto, R. I. Caenorhabditis elegans as a model system for studying non-cell-autonomous mechanisms in protein-misfolding diseases. *Dis Model Mech* **7**, 31–9 (2014).

[131] Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol* **7**, 539 (2011).

[132] Chan, P. & Warwicker, J. Evidence for the adaptation of protein ph-dependence to subcellular ph. *BMC Biol* **7**, 69 (2009).

[133] Robertson, A. D. & Murphy, K. P. Protein structure and the energetics of protein stability. *Chem Rev* **97**, 1251–1268 (1997).

[134] Somero, G. N. Proteins and temperature. *Annu Rev Physiol* **57**, 43–68 (1995).

[135] Somero, G. N. Adaptation of enzymes to temperature: searching for basic "strategies". *Comp Biochem Physiol B Biochem Mol Biol* **139**, 321–33 (2004).

[136] Somero, G. N. Temperature adaptation of enzymes: biological optimization through structure-function comprimises. *Annual Review of Ecology and Systematics* **9**, 1–29 (1978).

[137] Gurvitz, A. *et al.* Predicting the function and subcellular location of caenorhabditis elegans proteins similar to saccharomyces cerevisiae beta-oxidation enzymes. *Yeast* **17**, 188–200 (2000).

[138] Motley, A. M., Hettema, E. H., Ketting, R., Plasterk, R. & Tabak, H. F. Caenorhabditis elegans has a single pathway to target matrix proteins to peroxisomes. *EMBO Rep* **1**, 40–6 (2000).

[139] Vanni, P., Giachetti, E., Pinzauti, G. & McFadden, B. A. Comparative structure, function and regulation of isocitrate lyase, an important assimilatory enzyme. *Comp Biochem Physiol B* **95**, 431–58 (1990).

[140] Khan, F. R. & McFadden, B. A. Caenorhabditis elegans: decay of isocitrate lyase during larval development. *Exp Parasitol* **54**, 47–54 (1982).

[141] Wadsworth, W. G. & Riddle, D. L. Developmental regulation of energy metabolism in caenorhabditis elegans. *Dev Biol* **132**, 167–73 (1989).

[142] Liu, F., Thatcher, J. D., Barral, J. M. & Epstein, H. F. Bifunctional glyoxylate cycle protein of caenorhabditis elegans: a developmentally regulated protein of intestine and muscle. *Dev Biol* **169**, 399–414 (1995).

[143] Braeckman, B. P., Houthoofd, K. & Vanfleteren, J. R. Intermediary metabolism. *WormBook* 1–24 (2009).

[144] O'Riordan, V. B. & Burnell, A. M. Intermediary metabolism in the dauer larva of the nematode caenorhabditis elegans - ii. the glyoxylate cycle and fatty-acid oxidation. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **95**, 125–130 (1990).

[145] Tsuboi, D., Qadota, H., Kasuya, K., Amano, M. & Kaibuchi, K. Isolation of the interacting molecules with gex-3 by a novel functional screening. *Biochem Biophys Res Commun* **292**, 697–701 (2002).

[146] Salido, E. *et al.* Phenotypic correction of a mouse model for primary hyperoxaluria with adeno-associated virus gene transfer. *Mol Ther* **19**, 870–5 (2011).

[147] Suda, T. *et al.* Progress toward liver-based gene therapy. *Hepatol Res* **39**, 325–40 (2009).

[148] Kamimura, K., Suda, T., Zhang, G. & Liu, D. Advances in gene delivery systems. *Pharmaceut Med* **25**, 293–306 (2011).

[149] Danpure, C. J. Advances in the enzymology and molecular genetics of primary hyperoxaluria type 1. prospects for gene therapy. *Nephrol Dial Transplant* **10 Suppl 8**, 24–9 (1995).

[150] Steipe, B. Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes. *Methods Enzymol* **388**, 176–86 (2004).

[151] Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I. & Sanchez-Ruiz, J. M. Thermal versus guanidine-induced unfolding of ubiquitin. an analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **38**, 8138–49 (1999).

[152] Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* **38**, 16419–23 (1999).

[153] Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–90 (2003).

[154] Ghaemmaghami, S. & Oas, T. G. Quantitative protein stability measurement in vivo. *Nat Struct Biol* **8**, 879–82 (2001).

[155] Risso, V. A., Gavira, J. A., Gaucher, E. A. & Sanchez-Ruiz, J. M. Phenotypic comparisons of consensus variants versus laboratory resurrections of precambrian proteins. *Proteins* **82**, 887–96 (2014).

[156] Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. & Sanchez-Ruiz, J. M. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian beta-lactamases. *J Am Chem Soc* **135**, 2899–902 (2013).

[157] Tomschy, A. *et al.* Engineering of phytase for improved activity at low ph. *Appl Environ Microbiol* **68**, 1907–13 (2002).

[158] Bianchi, E., Venturini, S., Pessi, A., Tramontano, A. & Sollazzo, M. High level expression and rational mutagenesis of a designed protein, the minibody. from an insoluble to a soluble molecule. *J Mol Biol* **236**, 649–59 (1994).

[159] Das, D. & Georgiadis, M. M. A directed approach to improving the solubility of moloney murine leukemia virus reverse transcriptase. *Protein Sci* **10**, 1936–41 (2001).

[160] Lawson, D. M. *et al.* Solving the structure of human h ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**, 541–4 (1991).

[161] Cooper, D. R. *et al.* Protein crystallization by surface entropy reduction: optimization of the ser strategy. *Acta Crystallogr D Biol Crystallogr* **63**, 636–45 (2007).

[162] Price, W. N., 2nd *et al.* Understanding the physical properties that control protein crystallization by analysis of large-scale experimental data. *Nat Biotechnol* **27**, 51–7 (2009).

[163] Bartoszewska, M. *et al.* Peroxisomal proteostasis involves a lon family protein that functions as protease and chaperone. *J Biol Chem* **287**, 27380–95 (2012).

[164] Privalov, P. & Khechinashvili, N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *Journal of molecular biology* **86**, 665–684 (1974).

[165] Costas, M. *et al.* Between-species variation in the kinetic stability of tim proteins linked to solvation-barrier free energies. *J Mol Biol* **385**, 924–37 (2009).

[166] Rodriguez-Larrea, D., Minning, S., Borchert, T. V. & Sanchez-Ruiz, J. M. Role of solvation barriers in protein kinetic stability. *J Mol Biol* **360**, 715–24 (2006).

[167] Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138–48 (1995).

[168] Vendruscolo, M. & Tartaglia, G. G. Towards quantitative predictions in cell biology using chemical properties of proteins. *Mol Biosyst* **4**, 1170–5 (2008).

[169] Tartaglia, G. G. & Vendruscolo, M. The zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* **37**, 1395–401 (2008).

[170] Fargue, S., Lewin, J., Rumsby, G. & Danpure, C. J. Four of the most common mutations in primary hyperoxaluria type 1 unmask the cryptic mitochondrial targeting sequence of alanine:glyoxylate aminotransferase encoded by the polymorphic minor allele. *J Biol Chem* **288**, 2475–84 (2013).

[171] Voisine, C., Pedersen, J. S. & Morimoto, R. I. Chaperone networks: tipping the balance in protein folding diseases. *Neurobiol Dis* **40**, 12–20 (2010).

[172] Pey, A. L., Albert, A. & Salido, E. Protein homeostasis defects of alanine-glyoxylate aminotransferase: new therapeutic strategies in primary hyperoxaluria type i. *Biomed Res Int* **2013**, 687658 (2013).

[173] Westerheide, S. D. & Morimoto, R. I. Heat shock response modulators as therapeutic tools for diseases of protein conformation. *J Biol Chem* **280**, 33097–100 (2005).

[174] Qing, G. *et al.* Cold-shock induced high-yield protein production in escherichia coli. *Nat Biotechnol* **22**, 877–82 (2004).

[175] Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**, 2411–23 (1995).

[176] Peterson, E. A. & Sober, H. A. Preparation of crystalline phosphorylated derivatives of vitamin b6. *Journal of the American Chemical Society* **76**, 169–175 (1954).

[177] Hilser, V. J., Gómez, J. & Freire, E. The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculations. *Proteins* **26**, 123–33 (1996).

[178] Chothia, C. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* **105**, 1–12 (1976).

[179] Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **131**, 266–80 (1986).

# Part 7

# Appendices

# APPENDIX A

OPEN ACCESS Freely available online

PLOS ONE

# The Role of Protein Denaturation Energetics and Molecular Chaperones in the Aggregation and Mistargeting of Mutants Causing Primary Hyperoxaluria Type I

**Noel Mesa-Torres[1☯], Israel Fabelo-Rosa[2☯], Debora Riverol[2], Cristina Yunta[3], Armando Albert[3], Eduardo Salido[2]*, Angel L. Pey[1]***

**1** Department of Physical Chemistry, Faculty of Sciences, University of Granada, Granada, Spain, **2** Centre for Biomedical Research on Rare Diseases, Instituto Tecnologías Biomédicas, University of La Laguna, Tenerife, Spain, **3** Department of Crystallography and Structural Biology, Instituto de Química Física "Rocasolano", Consejo Superior de Investigaciones Científicas, Madrid, Spain

## Abstract

Primary hyperoxaluria type I (PH1) is a conformational disease which result in the loss of alanine:glyoxylate aminotransferase (AGT) function. The study of AGT has important implications for protein folding and trafficking because PH1 mutants may cause protein aggregation and mitochondrial mistargeting. We herein describe a multidisciplinary study aimed to understand the molecular basis of protein aggregation and mistargeting in PH1 by studying twelve AGT variants. Expression studies in cell cultures reveal strong protein folding defects in PH1 causing mutants leading to enhanced aggregation, and in two cases, mitochondrial mistargeting. Immunoprecipitation studies in a cell-free system reveal that most mutants enhance the interactions with Hsc70 chaperones along their folding process, while in vitro binding experiments show no changes in the interaction of folded AGT dimers with the peroxisomal receptor Pex5p. Thermal denaturation studies by calorimetry support that PH1 causing mutants often kinetically destabilize the folded apo-protein through significant changes in the denaturation free energy barrier, whereas coenzyme binding overcomes this destabilization. Modeling of the mutations on a 1.9 Å crystal structure suggests that PH1 causing mutants perturb locally the native structure. Our work support that a misbalance between denaturation energetics and interactions with chaperones underlie aggregation and mistargeting in PH1, suggesting that native state stabilizers and protein homeostasis modulators are potential drugs to restore the complex and delicate balance of AGT protein homeostasis in PH1.

**Citation:** Mesa-Torres N, Fabelo-Rosa I, Riverol D, Yunta C, Albert A, et al. (2013) The Role of Protein Denaturation Energetics and Molecular Chaperones in the Aggregation and Mistargeting of Mutants Causing Primary Hyperoxaluria Type I. PLoS ONE 8(8): e71963. doi:10.1371/journal.pone.0071963

**Editor:** Eugene A. Permyakov, Russian Academy of Sciences, Institute for Biological Instrumentation, Russian Federation

**Received** June 13, 2013; **Accepted** July 5, 2013; **Published** August 27, 2013



**Funding:** This work was supported by the Spanish ministry of Science and Innovation (RYC2009-04147 and CSD2009-00088 to ALP, SAF2011-23933 to ES, and CSD2006-00015, S2010/BMD-2457 and BFU2011-25384 to AA) and Junta de Andalucia (P11CTS-7187 ALP); FPI predoctoral fellowships from the Spanish ministry of Science and Innovation to IF-R and NM-T. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: esalido@ull.es (ES); angelpey@ugr.es (ALP)

## Introduction

Primary hyperoxaluria type I (PH1) is an autosomal recessive inborn error of metabolism caused by mutations in the *AGXT* gene, coding for the enzyme alanine-glyoxylate aminotransferase (AGT). AGT catalyzes the transamination of L-alanine to pyruvate and glyoxylate to glycine in the presence of pyridoxal 5′-phosphate (PLP) as cofactor [1]. AGT deficiency causes glyoxylate accumulation that is subsequently oxidized to oxalate, leading to the production of calcium oxalate crystals that result in progressive renal failure, and eventually, a life-threatening systemic build-up of oxalate known as oxalosis [2]. Liver and kidney transplantation is the only curative option to date, but this aggressive treatment poses significant morbidity and mortality [3]. A limited number of patients with specific genotypes have been reported to respond to pharmacological doses of pyridoxine, even though the molecular mechanisms involved in the response are unclear [4,5,6,7]. *AGXT* has two polymorphic variants, the most frequent ("wild-type"; AGT WT) called the *major* allele (haplotype) and a less common polymorphic variant called the *minor* allele (referred to as AGT LM) which appears in 20% of control subjects and 46% of PH1 patients [1]. The minor allele shows two single amino acid substitutions (p.P11L and p.I340M) among other genomic changes. Even though the minor allele does not cause PH1 by itself, it is known to exacerbate the deleterious effects of additional mutations [1,4,7]. About 150 mutations in the *AGXT* gene have been described in PH1 patients, 50% of them being missense mutations [1], and a few of them, such as p.G170R and p.I244T, are relatively common. Several molecular mechanisms seem to contribute to AGT loss-of-function in PH1 at the protein level: i) Mitochondrial mistargeting, where the AGT enzyme is imported to mitochondria [2]; ii) Protein aggregation [4,8,9]; iii) Accelerated proteasomal degradation [8]; iv) Catalytic defects [4]. However, the molecular details underlying protein mistargeting, aggregation and degradation in PH1 remain unclear. Beyond their

interest in PH1, some of these mutations have important implications in cell biology and genetics. AGT mistargeting mutations are unique models to try to understand some of the principles behind enzyme compartmentalization in the cell, which also has relevant evolutionary connotations [10]. In addition, the necessary synergy between common polymorphisms and disease-causing mutations of the *AGXT* minor haplotype is one of the best characterized examples of such interaction in human genetics [2].

Human genetic diseases are often caused by alterations in protein homeostasis [11]. The ability of a protein to fold into its native and functional conformation relies on intrinsic physico-chemical properties (thermodynamic stability, folding, unfolding, misfolding and aggregation rates) in the crowded intracellular milieu, as well as in the interaction of different conformations populated along the folding/unfolding process with elements of the protein homeostasis network, an array of pathways involved in the control of protein synthesis, folding, post-translational modification, trafficking, disaggregation and degradation [11,12,13,14]. In the context of protein homeostasis, we have recently suggested a role of the low kinetic stability of the apo-AGT (which shows no coenzyme bound [6]) and enhanced interactions with Hsc70, Hsp90 and Hsp60 chaperones [6,9,15] in the aggregation (for the I244T variant -p.P11L, p.I244T and p.I340M in *cis* -) and mistargeting (for the G170R variant -p.P11L, p.G170R and p.I340M in *cis* -) of PH1 mutations. Whether these aberrant protein features are specific to these mutations and/or intrinsic to a certain pathogenic mechanism (aggregation vs. mistargeting) remains unclear. From the perspective of the protein homeostasis network, understanding the role of specific protein features in the disease-causing mechanisms is required to develop new therapeutic strategies aimed to restore protein function [11,13,14]. However, owing to the large complexity and interactivity of the protein homeostasis pathways, involving at least 800 different proteins [13], a detailed characterization of protein homeostasis defects in PH1 represents a remarkable challenge.

In this work, we have performed a multidisciplinary characterization of four polymorphic variants (WT, p.P11L, p.I340M and p.P11L/I340M or *minor allele*, LM) and six known PH1 mutations present in the *minor allele* (named p.H83R, p.F152I, p.G170R, p.I244T, p.P319L and p.A368T through this paper). We also analyzed two rare variants of uncertain pathogenicity: p.R197Q and p.A295T, both in the *minor allele*. Our main aim is to shed light on the complex mutational effects of AGT folding and stability in PH1 from biochemical, biophysical, cell and structural biology perspectives required to deeply understand PH1 as a conformational disease. Our results show that PH1 causing mutations associated with aggregation and mistargeting display common alterations in protein folding, stability and interaction with molecular chaperones. This suggests that the protein homeostasis pathways involved in both mechanisms are shared and, consequently, that the final fate of the mutant proteins is likely determined by their specific regulatory elements. Nevertheless, our results indicate that native state stability and molecular chaperones are key points to understand PH1 pathogenesis, that might be targeted pharmacologically to restore protein homeostasis in PH1 patients.

## Materials and Methods

### Construction, expression and purification of Pex5p-pbd and AGT proteins in *E.coli*

Cloning of the PTS1-binding domain (amino acids 235–602 in reference sequence NM_000319) of human Pex5 (Pex5p-pbd) was performed after reverse transcription of normal human liver mRNA and PCR amplification with primers BgNPEX5-F: AGATCTCATATGGAGTTTGAACGAGCCAAG and SlRPEX5-R: TGCGACGAATTCACTGGGGCAGGCCAAAC. The NdeI and EcoRI sites designed at the 5′end of the primers were used to clone the amplification product into pCOLDII expression vector. The AGT expression constructs were generated as previously described [6], using site-directed mutagenesis, standard subcloning procedures and confirmed by sequencing. *E. coli* BL21strain containing pCOLDII plasmids encoding AGT and Pex5p-pbd proteins were grown in the presence of ampicillin 0.1 mg/ml and induced with 0.4 mM IPTG for 6 h at 4°C. His-tagged AGT and Pex5-pbd proteins were purified from soluble extracts using IMAC-columns (GE Healthcare) as recommended by the manufacturer. Proteins were further purified by size-exclusion chromatography (SEC) as previously described [6]. Holo- and apo-AGT were prepared and stored as previously described [6]. Protein concentration was measured spectrophotometrically using $\varepsilon_{280(1\ mg/ml)} = 1.069$ (AGT) and 1.243 (Pex5p-pbd), calculated from their sequences [16].

### Spectroscopic analyses

All spectroscopic analyses were performed in 20 mM Na-Hepes, 200 mM NaCl pH 7.4 at 25°C. The hydrodynamic behavior of dimeric holo-AGT proteins was evaluated by dynamic light scattering (DLS) using 5 μM AGT (in subunit) and 50 μM PLP in a Zetasizer Nano ZS (Malvern Inc.). UV-visible absorption spectra were acquired in an Agilent 8453 diode-array spectrophotometer using 3 mm path length cuvettes and 20 μM AGT. Near-UV/Visible circular dichroism measurements were performed as described [6]. Fluorescence measurements were performed as previously described [6,17] with some minor modifications (see SI text).

### Enzyme kinetic analysis

The AGT overall transaminase activity was customarily measured as described [18]. Briefly, 5 μg/ml of AGT incubated in Na-Phosphate 0.1 M pH 8 buffer at 25°C in the presence of 150 μM PLP, 0–5 mM glyoxylate and the reaction was triggered by adding 0–100 mM L-Alanine. Pyruvate formed in the reactions was measured in a Tecan Infinite M200 Pro microplate reader by a coupled NADH:lactate dehydrogenase assay after 2 min reaction at 25°C. Global fittings of activity measurements were performed using a double-displacement mechanism [17].

### Differential scanning calorimetry (DSC)

DSC measurements were performed and analyzed using a two-state irreversible kinetic model as described [6] (a detailed description of DSC fittings can be found in the SI text). Denaturation rate constants $k$ are determined from the profiles of excess heat capacity vs. temperature profiles using equation 1:

$$k = \frac{v \cdot C_{p(exc)}}{\Delta H - < H >} \quad (1)$$

Where $C_{p(exc)}$ and $<H>$ are the excess heat capacity and the excess enthalpy at each temperature, $v$ and $\Delta H$ stand for the scan rate and the calorimetric enthalpy, respectively. The temperature dependence of the rate constants follows the Arrhenius equation:

$$k = A \cdot \exp\left[-\frac{E_a}{R \cdot T}\right] \quad (2)$$

Where $E_a$ is the activation energy and $T$ is the absolute temperature in K.

Mutational effects on the activation free energy ($\Delta\Delta G^{\neq}$), enthalpy ($\Delta\Delta H^{\neq}$) and entropy ($\Delta\Delta S^{\neq}$) where determined on the basis of the transition state theory as described in [19]. The values of $\Delta\Delta H^{\neq}$ and $\Delta\Delta S^{\neq}$ where considered to be constant within the temperature range involved in extrapolations based on the highly linear Arrhenius plots found in all cases as well as the nice agreement between the values obtained upon determination of these parameters at 37°C (physiological temperature) and at 60°C (approximately the average and median $T_m$ value for all AGT enzymes; data not shown). Mutational effects on activation energetic parameters were determined using equations (3)–(5):

$$\Delta\Delta G^{\neq} = -R\cdot T\cdot \ln\left(\frac{k_{(37^oC)}(mut)}{k_{(37^oC)}(WT)}\right) \quad (3)$$

$$\Delta\Delta H^{\neq} = E_a(mut) - E_a(WT) \quad (4)$$

$$-T\Delta\Delta S^{\neq} = \Delta\Delta G^{\neq} - \Delta\Delta H^{\neq} \quad (5)$$

## Expression and characterization of AGT variants in chinese hamster ovary (CHO) cells

CHO cells (ATTC, USA) were grown in α-MEM (Lonza, Germany) supplemented with glutamine, penicillin/streptomycin and 5% fetal bovine serum. Cell transfections were performed using AGT cDNA variants subcloned in pCIneo plasmids (Promega, USA) in 6-well plates with Transfast reagent (Promega, USA), following manufacturer's guidelines. After 24 h, cells were passed to 100 mm Petri dishes containing 4 glass coverslips for immunofluorescence studies. Cultures were harvested 48 h after transfection.

Cells were sonicated in lysis buffer (100 mM potassium phosphate pH 8.0, 250 mM sucrose, 0.05% triton X-100, 100 µM PLP) and centrifuged at 4°C, 5000 *g* for 10 min (Beckman, USA) to obtain soluble fractions (supernatants). The pellets were washed with phosphate buffer saline (PBS) and resuspended in RIPA buffer. The protein concentration was measured by using bicinchoninic acid (BCA), and equal amounts of total protein were run in 10% acrylamide-SDS gels. The upper portion of the gel (above 70 kDa) was stained with Coomassie blue and scanned to control for equal protein loading, while the remaining of the gel was transferred to nitrocellulose membranes for western analysis. Purified AGT protein were included to calibrate the amount of AGT present in the cell lysates. Membranes were probed with anti-AGT rabbit serum followed by HRP-conjugated anti-rabbit IgG (Jackson ImmunoResearch, USA) and enhanced chemiluminescence substrate (Roche, Germany). Chemiluminescent signals were measured in VersaDoc 4000 MP and ChemiDoc MP devices and analysed using ImageLab Software (BioRad, Hercules, CA).

AGT activities were measured using cell lysates containing 100 µg total protein incubated in 100 mM K-Phosphate buffer pH 8.0 at 37°C for 30 min in the presence of 40 µM PLP, 10 mM glyoxylate and 40 mM L-Alanine. Pyruvate formed in the reactions was measured by a coupled NADH:lactate dehydrogenase assay [18].

For immunofluorescence confocal microscopy, cells grown on two coverslips were pulse-labeled *in vivo* with 100 nM MitoTracker Red (Invitrogen USA) during 15 min, followed by additional 15 min chase in label-free medium. PBS, pH 7.4, was used as a buffer in all subsequent washes and incubations. Cells were fixed in 3% paraformaldehyde at room temperature for 10 min. and permeabilized with 0.1% Triton X100-PBS. For AGT labeling on mitotracker stained cells, rabbit anti-human AGT (a gift from Dr. Danpure, University College London, UK) and Alexa Fluor 488 goat anti-rabbit IgG (Invitrogen, USA) were used. To label AGT and peroxisomes on the same cells, guinea pig anti-human AGT (also provided by Dr. Danpure) and rabbit anti-human PMP70 (Abcam, UK) were used, followed by incubations with Alexa Fluor 488 goat anti-guinea pig IgG (Invitrogen, USA) and Alexa Fluor 555 goat anti-rabbit IgG (invitrogen, USA). The coverslips were mounted with PBS-glycerol. The images were taken with a 60× objective in a confocal laser-scanning microscope (Olympus Inverted IX81, Japan).

## *In vitro* expression of AGT variants in a cell-free system and interaction with Hsc70 chaperones

Cell-free expression of AGT variants was performed in rabbit reticulocyte lysates (TnT system, Promega, USA) at 30°C for 2 h using $^{35}$S-Met and AGT cDNA variants subcloned in pCIneo plasmids. Protein synthesis was stopped with 100 µg/ml cycloheximide, and $1/25^{th}$ of the reaction product was set aside for analysis. Hsc70 immunoprecipitation was carried out with the remaining TnT product, using rat-Hsc70 antibodies (Abcam, Cambridge, UK) as previously described [9]. The immunoprecipitated proteins and the initial TnT products were denatured in Laemmli's buffer and analyzed by SDS-PAGE and fluorography.

## Molecular modeling

p.I340M crystallization assays were carried out on a 60-well microbath under oil (Terasaki plates) at 291 K. Crystals were obtained using a precipitant solution containing 15% PEG 3350; 0.1M Bis-Tris pH 5.2, 150 mM $Li_2SO_4$ and 5% w/v octyl-b-D-glucoside as additive. The best crystal forms were obtained by mixing drops of 1 µl protein solution, 2 µl precipitant solution and 0.45 µl additive. In general, the crystals appear and grow in the following 24 hours. Crystals were mounted in a fiber loop, transferred to the cryoprotectant (20% glycerol on the crystal mother liquor solution) and flash-frozen at 100 K in a nitrogen gas steam. p.I340M crystals diffraction data set was collected using a ADSC Q4 CCD detector at ID14.4 beamline of the European Synchrotron Radiation Facility (Grenoble). Diffraction data were processed with XDS [20] and scaled with SCALA from the CCP4 package (Collaborative Computational Project, Number 4) [21]. A summary of the diffraction protocol, data-collection and refinement statistics are given in Table S2. p.I340M diffracted to 1.90 Å resolution and belonged to space group $P2_12_12_1$, with unit-cell parameters a = 54.5, b = 103.5, c = 153. The p.I340M structure was solved by molecular replacement using Phaser [22] with the coordinates from the native protein (PDB 1H0C [23]). Several cycles of restrained refinement with PHENIX [24] and iterative model building with COOT [24,25] yielded to the final model with an $R/R_{free}$ 0.17/0.20. The water structure was also modeled. The stereochemistry of the models was verified with MolProbity [26]. Ribbon figures were produced using PyMol [27]. The coordinates and structure factor amplitudes have been deposited in the Protein Data Bank (PDB code: 2YOB).

## Statistical analysis

For statistical comparison, a two-sample Student's t-test was performed.

## Results

### PH1 causing mutants decrease protein yields and enhance protein aggregation in mammalian cells

We have explored the impact of mutations and polymorphisms in AGT protein folding efficiency and intracellular trafficking upon transient transfection of CHO cells (Figure 1 and 2). We define AGT intracellular *foldability* as the ability to fold into native dimers inside the cell, which is determined by the partition of the newly synthesized protein into folding, aggregation and degradation pathways. Most of the mutations studied are shown to reduce the total immunoreactive AGT levels (soluble *plus* insoluble in Figure 1A) compared to WT, from ~8–10-fold (p.F152I, p.P319L and p.I244T) to ~3–4-fold (LM, p.P11L, p.H83R, p.G170R and p.A368T), while only three mutants show WT-like levels (p.I340M, p.A295T and p.R197Q). Since total AGT protein levels represent the balance between protein synthesis, folding and degradation, these results indicate that most of the PH1 mutants display folding defects, possibly leading to enhanced AGT turnover (remarkably for p.F152I, p.P319L and p.I244T) which would be consistent with previous studies showing increased proteasomal turnover of PH1 mutants in cell-free systems [8,9].

AGT levels and specific activities (per mg of total protein) in soluble extracts closely correlate (Figure 1). The specific activity (per mg of AGT) measured in these soluble extracts (0.9–1.5 mmol Pyruvate·$h^{-1}$·$mg^{-1}$, with the exception of p.H83R) and in purified AGT dimers (Table 1) are also found to be similar. Therefore, the AGT levels detected in CHO soluble extracts mostly reflect folded AGT dimers. AGT WT exists mainly in the soluble fraction (~90% of the total AGT protein), while the mutations decrease the presence of AGT in soluble extracts by ~2-fold (p.P11L and LM, $p<0.01$; p.A295T, $p = 0.054$; p.H83R and p.A368T, $p = 0.10$), ~3–4-fold (p.F152I and p.I244T, $p \leq 0.01$) and ~8-fold (p.P319L and p.G170R, $p<0.01$) (Figure 1A). Noteworthy, p.F152I, p.G170R, p.I244T and pP319L showed significantly decreased fraction of AGT protein ($p<0.05$ in all cases) and specific activity ($p<0.01$ except for p.P319L which was not detectable) in soluble extracts compared to AGT LM. Overall, these results indicate that p.F152I, p.G170R, p.I244T and p.P319L further reduce AGT intracellular foldability compared to AGT LM also by enhancing protein aggregation.

The intracellular targeting of the AGT variants has been also analyzed in CHO cells by immunofluorescence confocal microscopy (Figure 2). Only two mutants were found to mistarget to mitochondria (p.G170R and p.F152I), where AGT is thought to be unable to detoxify glyoxylate [2], while the remaining variants were found at the correct peroxisomal location. Since peroxisomal import of AGT is mediated by the interaction of its peroxisomal targeting sequence (PTS1) with Pex5p [28], a possible explanation for mitochondrial mistargeting would be alterations in the binding of mutant AGT folded dimers with Pex5p. However, calorimetric titrations of native AGT proteins with the PTS1 binding domain of Pex5p (Figure S1) show little or no effects of these PH1 mutations in the molecular recognition by Pex5p (Table S1).

### PH1 causing mutants strongly interact with Hsc70 chaperones

An alternative explanation for protein misfolding and mistargeting observed for some mutants (see section above) may be that AGT variants abnormally interact with molecular chaperones along their folding process, as previously described by us for the p.G170R and p.I244T variants and chaperones of the Hsp60, Hsp70 and Hsp90 families [6,9,15]. Thus, we have tested the interaction of all AGT variants with Hsc70 chaperones upon expression in a cell-free system by immunoprecipitation (Figure 3). All the AGT variants tested showed stronger interactions with

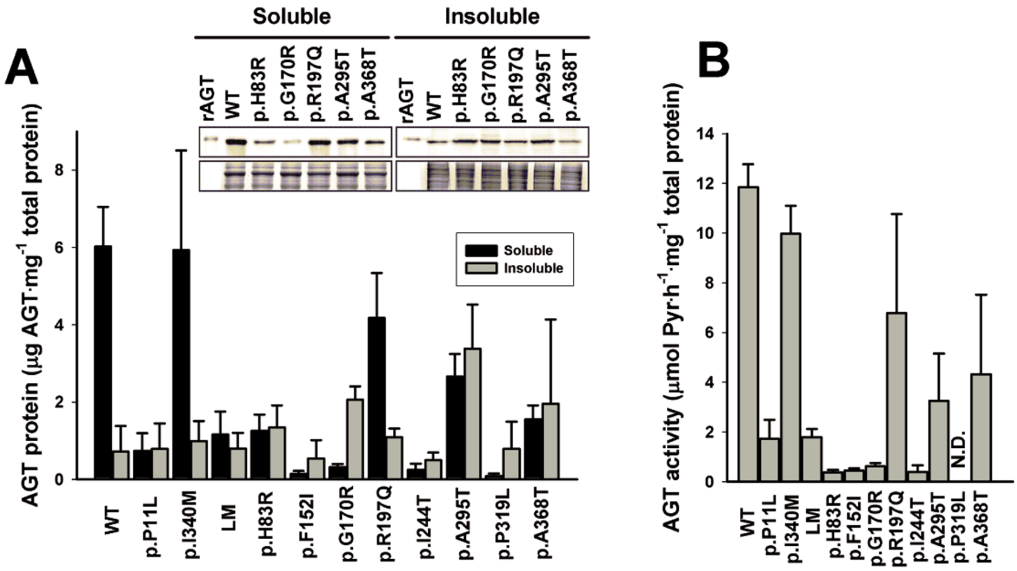


**Figure 1. Expression studies of AGT variants in CHO cells.** A) Levels of AGT immunoreactive protein as determined in soluble (black bars) and insoluble (grey bars) extracts. The inset shows representative immunodetection experiments of AGT variants (upper image) and the corresponding loading controls (lower image); rAGT: recombinant His-tagged AGT. B) AGT activities in soluble extracts. Data are means±s.d. of 2–5 independent experiments. N.D. not detectable.
doi:10.1371/journal.pone.0071963.g001

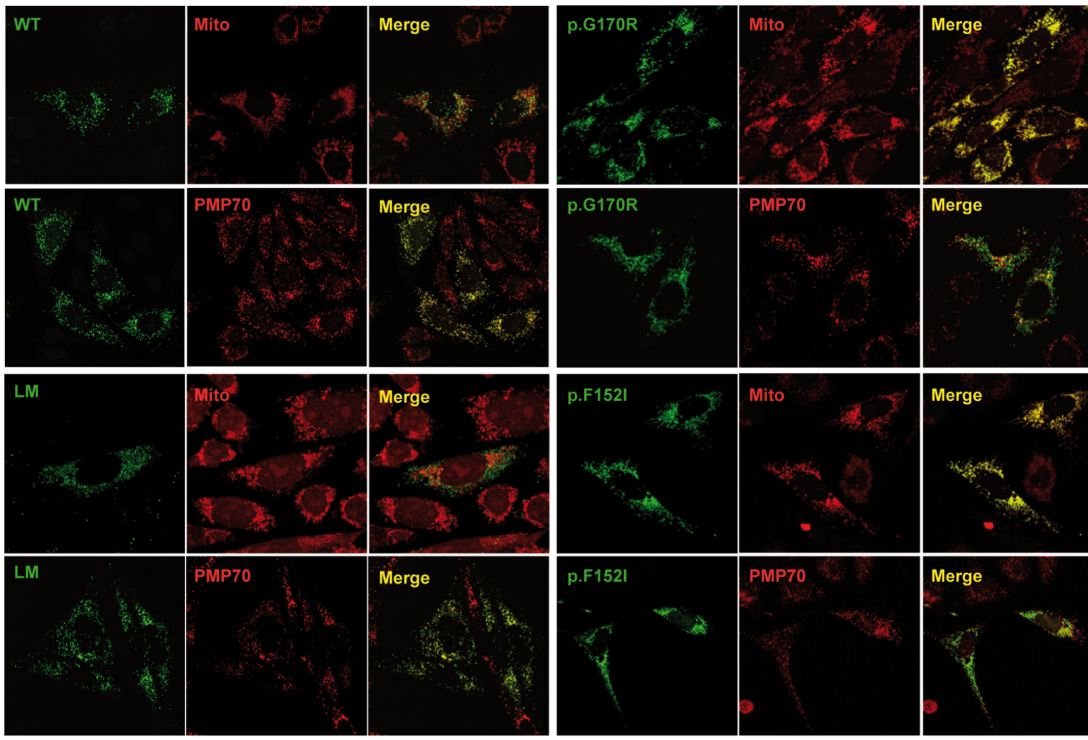


**Figure 2. Immunolocalization studies of AGT variants in CHO cells.** WT (upper left panel), LM (lower left panel), p.G170R (upper right panel) and p.F152I (lower right panel). In each panel, the upper row shows mitochondrial immunolocalization (AGT variant, Mitotracker probe and their merge), while lower row shows peroxisomal immunolocalization (AGT variant, PMP70 and their merge).
doi:10.1371/journal.pone.0071963.g002

**Table 1.** Functional properties and hydrodynamic diameter of AGT variants.

| AGT Variant | $V_{max}$ (mmol·h$^{-1}$·mg$^{-1}$)[a] | $K_{M(Alanine)}$ (mM)[a] | $K_{M(Glyoxilate)}$ (μM)[a] | $K_{d(PLP)}$ (nM)[b] | $k_{on(PLP)}$ ($M^{-1}·s^{-1}$) | $k_{off(PLP)}$ ($10^5$) ($s^{-1}$) | Diameter (mean±s.d.; in nm)[c] |
|---|---|---|---|---|---|---|---|
| WT | 2.22±0.09 | 19.5±1.4 | 245±29 | 100±30 (1400±800) | 197±33 | 28±18 | 8.1±2.3 |
| p.P11L | 2.43±0.08 | 18.3±1.2 | 185±22 | (1070±90) | 150±1 | 16±1 | 9.8±2.8 |
| p.I340M | 3.03±0.10 | 22.3±1.3 | 277±27 | 172±42 | N.det. | N.det. | 8.2±0.1 |
| LM | 2.26±0.09 | 19.8±1.5 | 197±27 | 157±15 | N.det. | N.det. | 7.8±0.1 |
| p.H83R | 0.141±0.005 | 42.4±2.4 | 156±19 | (1180±380) | 1470±160 | 174±53 | 8.9±2.3 |
| p.F152I | 1.90±0.19 | 15.7±3.1 | 201±68 | (490±300) | 2990±70 | 146±90 | 8.9±0.2 |
| p.G170R | 2.54±0.13 | 21.9±2.0 | 293±42 | (1380±170) | 125±2 | 17±2 | 6.6±0.9 |
| p.R197Q | 1.84±0.06 | 18.1±1.2 | 168±21 | 156±34 | N.det. | N.det. | 9.9±0.7 |
| p.I244T | 2.82±0.20 | 16.8±2.4 | 193±49 | (3060±240) | 51±1 | 16±1 | 7.1±0.7 |
| p.A295T | 2.10±0.10 | 18.5±1.3 | 292±30 | 71±16 | N.det. | N.det. | 9.5±0.4 |
| p.P319L | 1.54±0.13 | 16.2±2.7 | 212±60 | (240±360) | 268±7 | 6±10 | 7.3±1.0 |
| p.A368T | 2.12±0.13 | 23.0±2.5 | 297±50 | 134±26 | N.det. | N.det. | 8.2±0.2 |

[a]Enzyme kinetic parameters were obtained from global fittings of 2–4 independent experimental series using a double-displacement mechanism;
[b]$K_d$ values were estimated from single titrations except for AGT WT protein where $K_d$ value is the mean±s.d. from three independent titrations. Data in parenthesis are $K_d$ estimates obtained from the kinetic binding experiments ($K_d = k_{off}/k_{on}$).
[c]Hydrodynamic diameter of holo-AGT variants determined by dynamic light scattering (DLS). Data are mean±s.d. of 3–9 independent measurements.
doi:10.1371/journal.pone.0071963.t001

Hsc70 chaperones ($p<0.01$ in all cases), except p.I340M. Polymorphic p.P11L and LM showed a 4-fold increase in immunoprecipitated AGT compared to the WT protein, while disease-causing mutants showed somewhat higher levels, from 4.8-fold (p.H83R, $p = 0.46$ vs. LM) to 7-fold (p.P319L, $p = 0.078$ vs. LM). Owing to the significant variability in these experiments, the changes detected are not robust enough to claim mutation-specific differences compared to AGT LM, but the trend observed suggests that disease-causing mutations may increase the interaction with Hsc70 chaperones along their folding process vs. LM protein.

## PH1 causing mutants do not generally perturb AGT oligomerization and function

To probe whether PH1 mutants may affect AGT oligomerization and catalytic properties (as recently found for some PH1 mutants [4]), we have expressed recombinant AGT variants in *E.coli* and purified AGT dimers to homogeneity. Size-exclusion chromatography analysis showed a single peak for all variants with an elution volume consistent with a ≈90 kDa dimeric form (data not shown). The molecular dimensions of purified dimers were further studied by dynamic light scattering (DLS), showing a hydrodynamic diameter of $8.1 \pm 2.3$ nm for holo-WT and $8.4 \pm 1.1$ nm for the rest of holo-AGT variants (Table 1). Within the experimental uncertainty, these results are consistent with the size of dimeric AGT WT previously reported by DLS [29] and also imply no noticeable perturbation of dimer size and/or dimer-monomer equilibrium by these mutations.

Thus, we evaluated the effect of these variants on the functionality of the AGT protein. Enzyme kinetic analyses based on a double-displacement mechanism ([17]; see Figure 4A–B for representative examples) were performed for all the AGT variants and the enzyme kinetic parameters obtained are compiled in Table 1. All the AGT variants tested showed similar specific activity ($V_{max}$) to that found for AGT WT with the exception of p.H83R mutant, which displayed a ~15-fold reduction in specific activity. No large changes in $K_m$ values for L-Ala and glyoxylate were found among the AGT variants studied. The environment of the bound coenzymes to WT and p.H83R was characterized by Near-UV and visible absorption and circular dichroism spectroscopies, revealing a ~10–15 nm blue shift in the absorbance and dichroic bands of bound PLP and PMP (Figure 4C–D), which supports a distorted coenzyme binding mode to the p.H83R mutant. We have further tested this hypothesis by incubating holo-WT and p.H83R (at a final monomer concentration of 36 μM) in the presence of L-Ala for 1 h at 25°C and analyzed the the amount of PMP released by UV-absorption spectroscopy upon filtration using microfilters of 10 kDa cut-off. Under these conditions, ~3% of the PMP was released in the WT enzyme (consistent with tight binding of PMP along the overall transamination reaction previously reported; [17]), while ~44% of PMP was released in the p.H83R, indicating a large decrease in PMP binding affinity for this mutant. Thus, we conclude that the large decrease in catalytic performance of p.H83R is caused by a distortion in the binding mode of PLP and PMP coenzymes.

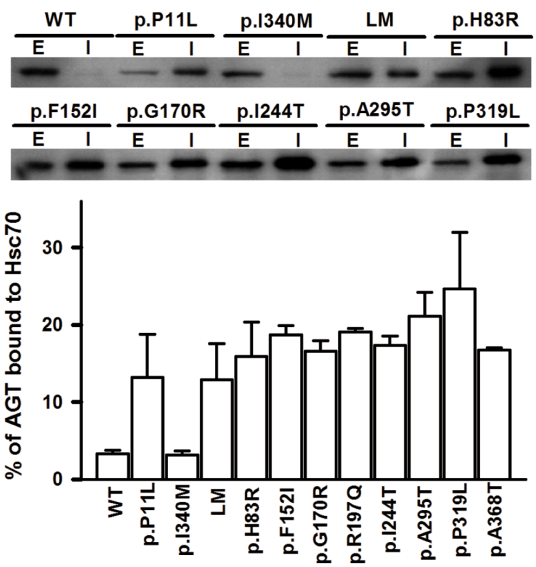


**Figure 3. Interaction of AGT variants with Hsc70 chaperones in a cell-free system.** Representative autoradiograms of AGT proteins labeled with $^{35}$S-Met are shown for several AGT variants (E: total AGT synthesized in extracts; I: AGT immunoprecipitated using anti-Hsc70 antibodies; note that 1 μl of the TnT reaction was loaded in E lanes, while the protein immunoprecipitated from 6 μl TnT lysate was loaded in I lanes). Data in the lower panel are expressed as percentage of immunoprecipitated AGT compared to the total AGT synthesized, and are mean±s.d. from three independent experiments.
doi:10.1371/journal.pone.0071963.g003

We have also measured the binding affinity of the apo-variants for PLP using fluorescence spectroscopy (Figure S2). Direct equilibrium measurements allowed to determine the $K_d$ values for the WT and some stable mutants at 30°C (Table 1). All the AGT variants studied under equilibrium conditions showed affinity for PLP similar to the WT protein ($K_d$ about 100 nM). Under these conditions, aggregation in the absence of PLP was found for p.P11L, p.H83R, p.F152I, p.G170R, p.I244T and p.P319L mutants. Lowering the temperature or adding 10% glycerol did not prevent aggregation in the time scale required for AGT:PLP equilibration. Alternatively, we estimated the affinity for PLP by performing kinetic binding experiments under pseudo-first order conditions (Figure S2B) providing only an upper limit for the $K_d$ values mostly due to the large uncertainties associated to the determination of the $k_{off}$ (Table 1). No clear differences were observed for the estimated $K_d$ values for PLP between WT and the mutants studied using this kinetic approach (Table 1).

## PH1 mutations kinetically destabilize the apo-AGT form

Thermal denaturation of AGT shows a single transition for all AGT variants studied by differential scanning calorimetry (DSC; Figure 5A). Denaturation of AGT variants is a purely kinetic process, which is well described by the irreversible two-state conversion of the native dimer to a final aggregated state (N→F). This kinetic process is characterized by a denaturation rate constant $k$ [6,30], which is inversely proportional to its half-life for denaturation. In this scenario, protein kinetic stability (as a given denaturation rate or half-life under certain experimental conditions) is determined by the height of the free energy barrier that AGT must cross from the native state to reach the transition state of the denaturation rate-limiting step [7]. Thus, mutational and PLP effects on AGT thermal transition ($T_m$ and $E_a$ values) are translated into changes in kinetic stability (Table 2 and Figure S3). Remarkably, removal of bound PLP has a dramatic effect on the stability of most of AGT variants, reducing by ~25°C and 4–5 orders of magnitude the $T_m$ values and the corresponding kinetic stabilities (Table 2 and Figure S3).

DSC is particularly suitable to study the stability of AGT enzymes because it provides not only a $T_m$ value, which can be

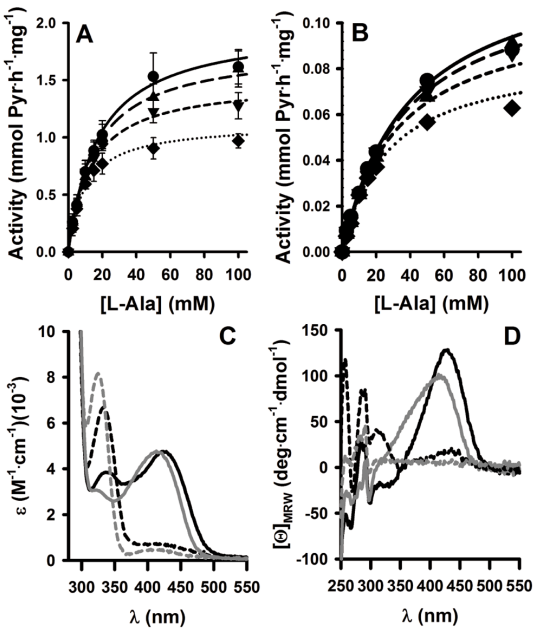


**Figure 4. Functional characterization of WT and p.H83R.** A and B) Enzyme activity measurements for WT (A) and p.H83R (B) at different L-Ala concentrations in the presence of 0.25 mM (diamonds), 0.5 mM (down triangles), 1 mM (up triangles) and 2 mM (circles) glyoxylate. Data in panel A are means±s.d from four independent measurements while data in panel B are means from two independent measurements. Lines are best fits for the different glyoxylate concentrations obtained from global fittings using a double-displacement mechanism. C and D) Absorption (C) and circular dichroism (D) spectra for holo-WT (black) and holo-p.H83R (grey) acquired upon incubation for at least 10 min in the absence (continuous lines; PLP bound) or presence (dashed lines; PMP bound) of 500 mM L-alanine.
doi:10.1371/journal.pone.0071963.g004

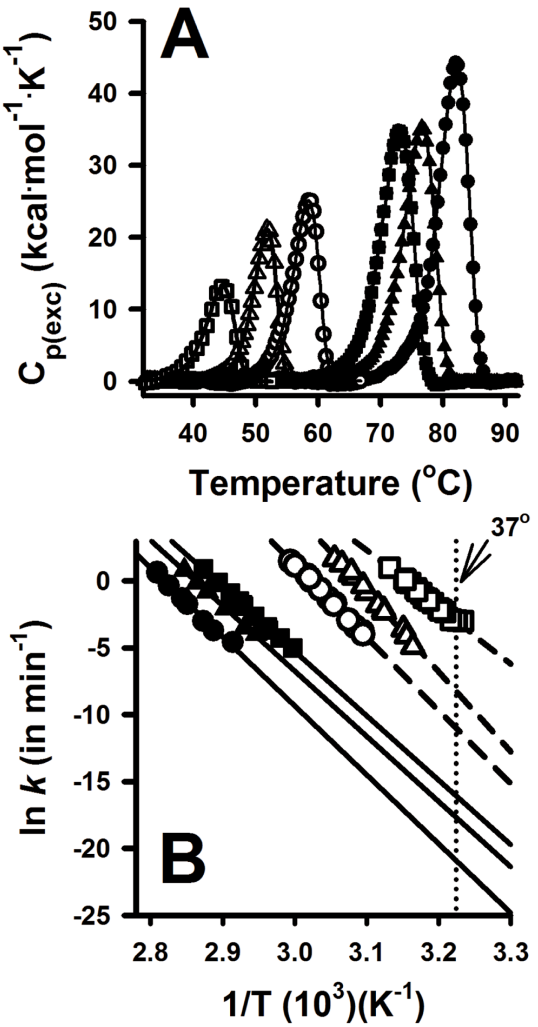


**Figure 5. Thermal denaturation of AGT variants studied by differential scanning calorimetry (DSC).** A) Representative DSC traces obtained at 3°C/min; Lines are best-fits from a two-state irreversible denaturation model with first-order kinetics [6]; B) Arrhenius plots for the irreversible denaturation of AGT variants, indicating also the extrapolated rate constants at physiological temperature (intercept with the vertical dotted line). Symbols are: WT (circle), LM (triangle) and p.F152I (square). Data for holo proteins are shown as closed symbols and for apo proteins as open symbols.
doi:10.1371/journal.pone.0071963.g005

alternatively obtained by other techniques, but also allow determine accurately the rate constants as a function of temperature, to compare kinetic stabilities over widely different time scales (in this work half-lives for denaturation range from several minutes to thousands of years, extrapolated to physiological temperature; Table 2) and provide insightful parameters of the denaturation process such as denaturation enthalpies and activation energies [6,19,30]. For instance, the temperature dependence of denaturation enthalpies shows a common behavior for holo- and apo-AGT enzymes (Figure 6A), yielding a global slope (i.e. a denaturation heat capacity) of 10.2±0.6 kcal·mol$^{-1}$·K$^{-1}$ which is slightly lower than the theoretical value expected for global unfolding of the AGT dimer (14.8 kcal·mol$^{-1}$·K$^{-1}$; based on [31]). This analysis supports that AGT thermal denaturation involves a large loss of tertiary structure, possibly reflecting denaturation of both domains in AGT, and also that all the AGT enzymes studied here in their holo- and apo-forms denature to a similar extent.

DSC scans are used to determine denaturation rates ($k$) or half-lives at different temperatures by building Arrhenius plots (Figure 5B) which allow study kinetic stabilities at physiological temperature (Table 2). Most of the holo variants are kinetically stable at 37°C with extrapolated half-lives ranging from ~12 to ~80000 years (Table 2), with the exception of p.H83R, with a very short half-life of ~4 h likely due to a distortion in the binding mode of PLP (Figure 4). The apo-proteins are much less stable, remarkably the p.G170R, p.A295T, p.H83R, p.F152I, p.I244T and p.P319L enzymes, which reduce ~20–150-fold the kinetic stability vs. the LM variant, and they denature on a time scale of a few minutes to hours (Table 2). Indeed, the half-lives for enzyme inactivation upon incubation of apo-proteins at 37°C and standard activity measurements agree very well with those obtained from our DSC analyses (Table 2). This significant kinetic destabilization of the apo vs. holo-forms in most of PH1 mutants is not due to

**Table 2.** Thermal denaturation and kinetic stability parameters for AGT enzymes.

| | $T_m$ (°C)[a] | $K_{37°C}$ ($min^{-1}$)[b,c] | $\Delta h_{cal}$ (kcal·$mol^{-1}$)[b,d] | $E_a$ (kcal·$mol^{-1}$)[e] | μ[b] |
|---|---|---|---|---|---|
| *Holo-AGT* | | | | | |
| WT | 82.1 | $6.4\pm2.2\cdot10^{-10}$ | 548±5 | 109±5 | 0.95±0.02 |
| p.P11L | 73.8 | $1.1\pm0.2\cdot10^{-7}$ | 400±4 | 95±3 | 0.98±0.01 |
| p.I340M | 84.1 | $1.7\pm1.1\cdot10^{-11}$ | 530±24 | 115±4 | 0.90±0.03 |
| LM | 76.6 | $9.3\pm0.9\cdot10^{-9}$ | 375±13 | 101±5 | 1.00±0.01 |
| p.H83R | 58.2 | $2.9\pm0.4\cdot10^{-3}$ | 236±15 | 55±3 | 0.90±0.03 |
| p.F152I | 73.1 | $9.9\pm3.3\cdot10^{-8}$ | 454±18 | 97±4 | 0.92±0.04 |
| p.G170R | 75.5 | $3.7\pm0.6\cdot10^{-8}$ | 458±17 | 100±2 | 0.99±0.01 |
| p.R197Q | 77.9 | $3.7\pm0.7\cdot10^{-9}$ | 401±6 | 110±6 | 0.96±0.01 |
| p.I244T | 75.8 | $1.1\pm0.1\cdot10^{-8}$ | 405±10 | 103±2 | 0.95±0.02 |
| p.A295T | 77.5 | $1.2\pm0.2\cdot10^{-8}$ | 500±9 | 105±8 | 0.91±0.01 |
| p.P319L | 76.5 | $8.6\pm1.5\cdot10^{-9}$ | 395±5 | 101±3 | 0.98±0.04 |
| p.A368T | 76.5 | $6.8\pm2.1\cdot10^{-9}$ | 440±6 | 102±7 | 0.98±0.04 |
| *Apo-AGT* | | | | | |
| WT | 58.4 | $1.6\pm0.7\cdot10^{-5}$ | 255±17 | 111±10 | 1.30±0.02 |
| p.P11L | 49.9 | $2.1\pm0.8\cdot10^{-3}$ | 201±14 | 105±3 | 1.14±0.01 |
| p.I340M | 61.1 | $2.0\pm1.0\cdot10^{-7}$ | 370±21 | 143±11 | 1.04±0.03 |
| LM | 51.8 | $2.6\pm1.5\cdot10^{-4}$ ($3.7\pm2.0\cdot10^{-4}$) | 204±10 | 121±6 | 1.15±0.06 |
| p.H83R | 46.9 | $3.8\pm0.8\cdot10^{-2}$ | 156±7 | 62±6 | 0.91±0.04 |
| p.F152I | 44.7 | $4.9\pm0.4\cdot10^{-2}$ | 137±17 | 87±7 | 1.13±0.04 |
| p.G170R | 48.5 | $6.8\pm0.5\cdot10^{-3}$ ($9.2\pm3.0\cdot10^{-3}$) | 201±13 | 92±9 | 1.00±0.04 |
| p.R197Q | 52.3 | $5.5\pm0.5\cdot10^{-4}$ | 200±4 | 110±10 | 0.96±0.06 |
| p.I244T | 47.3 | $3.6\pm0.4\cdot10^{-2}$ ($2.6\pm0.5\cdot10^{-2}$) | 162±10 | 63±1 | 0.88±0.02 |
| p.A295T | 49.4 | $5.3\pm1.2\cdot10^{-3}$ ($2.4\pm0.7\cdot10^{-3}$) | 244±14 | 92±9 | 0.89±0.01 |
| p.P319L | 47.1 | $3.9\pm0.2\cdot10^{-2}$ | 194±37 | 62±2 | 0.88±0.01 |
| p.A368T | 51.8 | $5.9\pm0.5\cdot10^{-4}$ | 201±3 | 112±6 | 1.15±0.06 |

[a]determined at 3°C/min scan rate.
[b]mean±s.d. from independent experiments performed at three different scan rates.
[c]Values in parentheses are the inactivation rate constants measured at 37°C.
[d]expressed per mol of AGT dimer (i.e. the unfolding cooperative unit).
[e]mean±s.d. from the four consistency tests proposed by [47].
doi:10.1371/journal.pone.0071963.t002

large changes in the affinity for PLP (Table 1), but rather, it suggests that cofactor binding overcomes some destabilizing interactions present in the native state of the apo-forms [7]. Regarding the polymorphic variants, p.P11L and LM are kinetically destabilizing, while p.I340M enhances AGT kinetic stability compared to the WT protein, either in the apo- or the holo-form (Table 2).

The large kinetic stabilization induced by PLP binding, as well as the destabilization induced by some mutants in their apo-form must originate from changes in the height of the denaturation free energy barrier (i.e. the free energy difference between the native and denaturation transition state [6]). Our DSC analyses also provide information on the reaction order of AGT denaturation by determining the μ value (i.e. reaction order is $1/\mu$), and hence, on the oligomerization state of the denaturation transition state: a μ value close to one indicates first-order kinetics, involving a dimeric transition state for AGT, while a value close to two supports that the transition state is monomeric. As we show in Table 2, most of the AGT variants display first-order denaturation kinetics ($\mu\approx1$), with only small deviations for a few apo-forms. Thus, the AGT denaturation transition state is essentially a partially unfolded dimer for all AGT variants, indicating that other unfolding steps (dimer dissociation and monomer unfolding) must occur after the denaturation rate-limiting step, and hence, these steps do not contribute to the effect of mutations, polymorphisms or PLP binding on AGT kinetic stability. PLP mediated kinetic stabilization arise from its preferential binding and stabilization of the native AGT dimeric structure, raising the denaturation free energy barrier by ~7 kcal·$mol^{-1}$ (see [6] and Table 1).

## Structural and energetic basis of AGT kinetic destabilization by PH1 mutations

A plot of the activation energies for denaturation of all AGT enzymes as a function of their $T_m$ value (which exponentially correlate well with their kinetic stabilities; see Figure S3) shows that those AGT enzymes with lower kinetic stability display lower values of $E_a$ (Figure 6B). The high linearity of Arrhenius plots (Figure 5B) indicates a negligible activation heat capacity, and thus, $E_a$ values can be considered essentially as temperature-independent. Hence, the results shown in Figure 6B can be rationalized as mutational effects on the *energetic balance* (entropic and enthalpic contributions) of the free energy barrier for denaturation (previously shown for denaturation of other protein

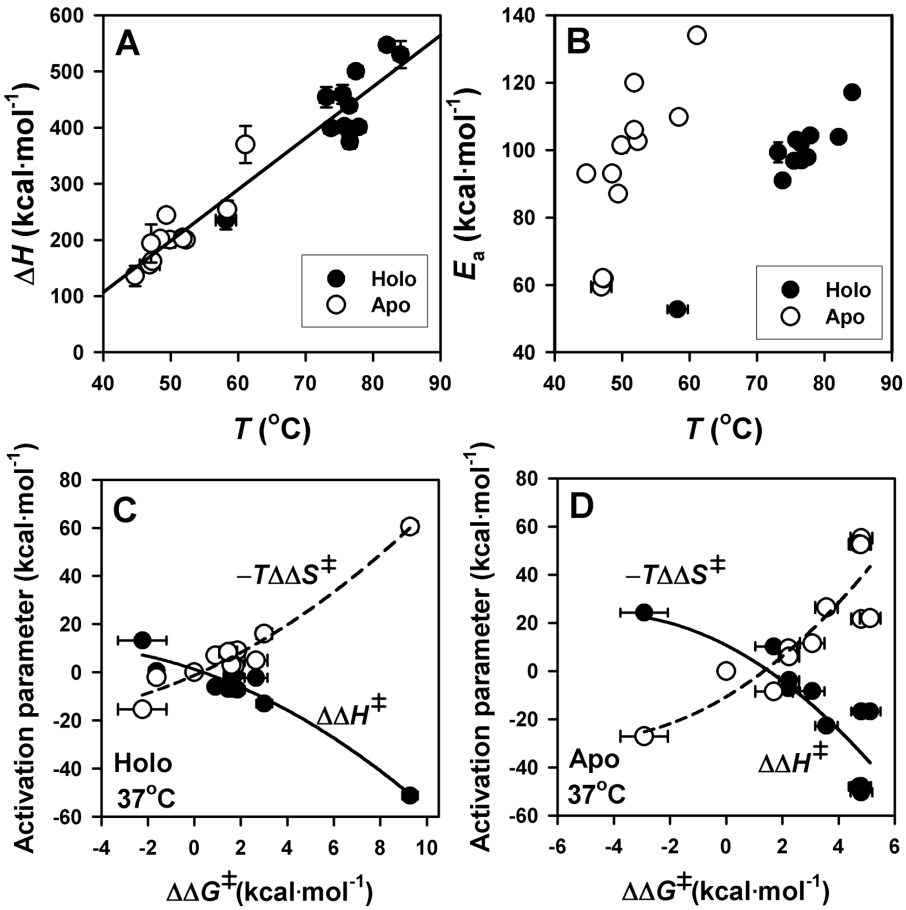


**Figure 6. Structure-energetics relationships for thermal denaturion of holo- and apo-AGT enzymes.** A) Temperature dependence of denaturation enthalpies ($\Delta H$) for holo- (closed symbols) and apo-(open symbols) proteins. The linear fit provides the value of $\Delta C_p$ ($=10.2\pm0.6$ kcal.mol$^{-1}$). B) Activation energy ($E_a$) plotted vs. the $T_m$ for holo- (closed symbols) and apo-(open symbols) AGT enzymes. C and D) changes in activation enthalpic and entropic contributions to AGT kinetic stability as a function of changes in activation free energies for holo-(C) and apo-(D) AGT enzymes. Lines in C and D are meant to guide the eye and have no theoretical meaning.
doi:10.1371/journal.pone.0071963.g006

systems; [19,32,33]). We have thus calculated the changes in the activation enthalpies ($\Delta\Delta H^{\ddagger}$) and entropies ($-T\Delta\Delta S^{\ddagger}$) and plotted them as a function of the mutational effects on the kinetic stability ($\Delta\Delta G^{\ddagger}$) for the holo-(Figure 6C) and apo-(Figure 6D) enzymes. We found that changes in kinetic stability ($\Delta\Delta G^{\ddagger}$) arise from large changes of opposite sign in $\Delta\Delta H^{\ddagger}$ and $-\Delta\Delta TS^{\ddagger}$ that largely cancel out for holo- and apo-AGT enzymes. Since $\Delta\Delta H^{\ddagger}$ reflect structural differences (~solvent exposure) between the native and the transition states, the analyses shown in Figure 6 support that the most destabilizing mutants decrease the magnitude of the structural change occurring between the native and denaturation transition states. We must note that there is no experimental evidence of large structural changes in the native dimer upon mutation: they show similar hydrodynamic radius and, with the exception of p.H83R, similar enzyme activities (Table 1) as well as denaturation enthalpies consistent with similarly folded native states (Table 2), and thus, these changes may primarily affect the denaturation transition state. Using structure-energetics relationships, a value of $-60$ kcal·mol$^{-1}$ for $\Delta\Delta H^{\ddagger}$ (found for several AGT mutants; see Figure 6) is translated into a difference of ~90 folded residues or ~8500 Å$^2$ of solvent exposed surface [31] between the native and transition state (per dimer), supporting the existence of significant mutational effects on the structure and energetics of the denaturation transition state (see [33] for a similar situation with mutants of human phosphoglycerate kinase 1).

## Structure of p.I340M and modeling of the AGT variants

The crystal structure of p.I340M has been determined at an unprecedented 1.9 Å resolution to improve our modeling efforts on PH1 causing mutants. The structures of p.I340M and the WT (PDB code 1H0C [23]) are nearly identical. AGT homodimer has each protomer folded into a large N-terminal domain, a smaller C-terminal domain and a 22 amino acid long unstructured N-terminal tail that grabs the other subunit within the dimer (Figure 7A). None of the substitutions observed in the same

haplotype are clustered, and thus, the effects of these substitutions are likely to be additive rather than synergic. The P11L and I340M polymorphisms are found in the dimer interface. P11L is expected to produce steric clashes with residues from the other subunit, which would explain the reduced stability and foldability of p.P11L, while I340M establishes several new favorable interactions hence explaining the stabilization observed for p.I340M, and the partial compensation of P11L destabilizing effect when they both occur in *cis* (i.e. the *minor* allele).

The mutations analyzed on the minor allele occur in locations far from the dimerization interface (Figure 7A), and with the exception of H83R (Figure 7B), none of them are found near the active site. Modeling of Arg83 predicts a loss of interactions with PLP and the appearance of repulsions with surrounding residues, likely affecting the orientation of PLP (and/or PMP), which could explain the low catalytic activity, altered PLP/PMP binding mode and low stability as holo-protein of H83R. A295T, P319L and A368T substitutions are found in the vicinity of the domain:domain interface of AGT monomer (Figure 7B), and cause an increase in the size of the side chain. Interestingly, these three mutations show more pronounced apo-AGT destabilization (Table 2) and deleterious effects in CHO cells (as total and soluble AGT protein levels; Figure 1) as closer the substitution is to the domain interface (P319L>A295T≥A368T), suggesting that a fine tuning of interactions in AGT domain:domain interface may be important for apo AGT kinetic stability and intracellular stability and foldability. G170R, F152I and R197Q mutations are not expected to cause important alterations on AGT structure based on the simple modeling of these substitutions or the crystal structures available [34], although molecular dynamic simulations have suggested a significant structural impact of the F152I mutation [35].

## Discussion

A majority of the most common alleles associated to PH1 affect protein folding and stability, suggesting that most PH1 patients

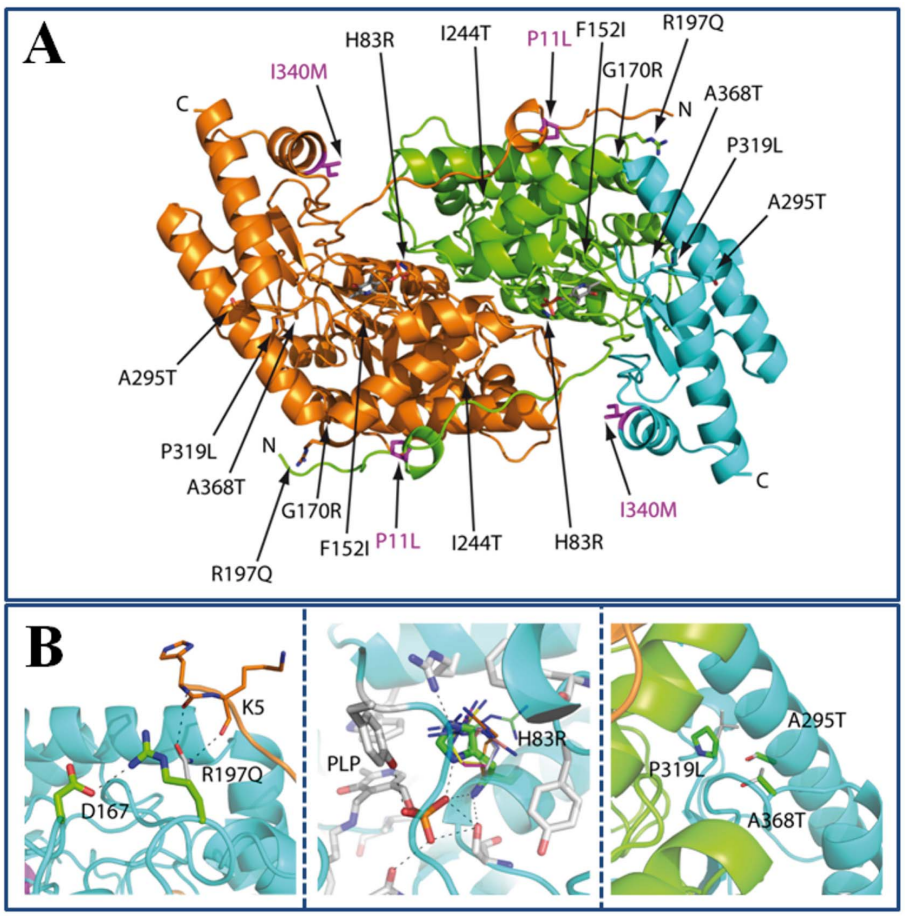


**Figure 7. Structural modeling of AGT mutations and polymorphisms.** A) Representation of AGT dimer structure. The domain structure of one of the subunits is emphasized using green and blue colors. The AGT mutations found in PH1 patients are labeled and pointed by the arrows while the polymorphisms constituting the *minor allele* are highlighted in magenta. B) Representation of the structural environment of the Arg197 (left), His83 (middle) and Ala295, Pro319 and Ala368 (right). The mutations have been modeled as thinner sticks.
doi:10.1371/journal.pone.0071963.g007

may show defective AGT function due to alterations in its protein homeostasis (this work; [4,6,9,36]). To understand the different mechanisms underlying PH1, we have integrated in this work concepts from protein structure, biophysics, biochemistry and cell biology to provide a comprehensive view on the mutational effects on protein folding, assembly, transport, mistargeting and degradation of the AGT protein (summarized in Table 3). Our work delineates important checkpoints in AGT protein homeostasis, such as the stability of the apo-proteins and the recognition of folding intermediates by molecular chaperones (Figure 8) that might be specifically targeted to restore AGT function in PH1 patients. We also provide insight to current potential therapies for PH1 such as pyridoxine supplementation [4,5,6,37].

PLP binding to AGT enhances native state kinetic stability by 4–5 orders of magnitude ([6]; this work), making holo-proteins highly kinetically stable at physiological temperature, with the only exception of the catalytic mutant p.H83R. Interestingly, we found that p.F152I, p.G170R, p.I244T, p.A295T and p.P319L are markedly destabilized as apo-proteins compared to the non-pathogenic LM variant (Table 2), and actually, they denature at a relatively fast rate at physiological temperature. As we have previously discussed for p.G170R [6], the low kinetic stability of these mutants in the apo-form may have important implications for PH1 pathogenesis, since it is likely that a significant fraction of AGT may transiently exist as apo-protein *in vivo*, and thus, it might be susceptible to intracellular irreversible alterations such as mitochondrial import, protein aggregation and degradation. Consequently, four of these variants (p.F152I, p.G170R, p.I244T and p.P319L) show evident signs of misfolding and, possibly, accelerated turnover in CHO cells, while two of them (p.G170R and p.F152I) also cause protein mitochondrial mistargeting (Table 3). This indicates some degree of correlation between the apo- stability and protein mitochondrial import and/or intracellular aggregation. According to this interpretation, we propose a beneficial effect of pyridoxine supplementation in patients carrying these four mutations (F152I, G170R, I244T and P319L). In fact, p.F152I and p.G170R have been described as pyridoxine-responsive genotypes in PH1 patients [38,39].

Our DSC analyses provide molecular insights on the effect of PH1 causing mutants and PLP binding on the AGT kinetic stability, denaturation mechanism and structural/energetic features of its denaturation free energy barrier previously unexplored. Within the set of PH1 mutants studied here, only the thermal stability of p.F152I and p.G170R have been reported earlier (by circular dichroism and inactivation experiments) showing thermal destabilization (lower $T_m$) mainly for their apo-forms [35,40,41]. Our detailed DSC kinetic analyses further show that the kinetic overstabilization exerted by PLP bound to p.F152I, p.G170R, p.I244T, p.A295T and p.P319L as holo-proteins arises from subtle changes in the enthalpic and entropic contributions to the denaturation free energy barrier, since a similar pattern of enthalpy/entropy compensation is found for holo- and apo-AGT proteins (Figure 6C and 6D). Moreover, these effects must concern mainly to the structure and energetics of the dimeric transition state for denaturation, and thus, mutational effects on dimer dissociation and full monomer unfolding do not contribute to the relevant kinetic stability of AGT enzymes since they occur after the denaturation rate-limiting step.

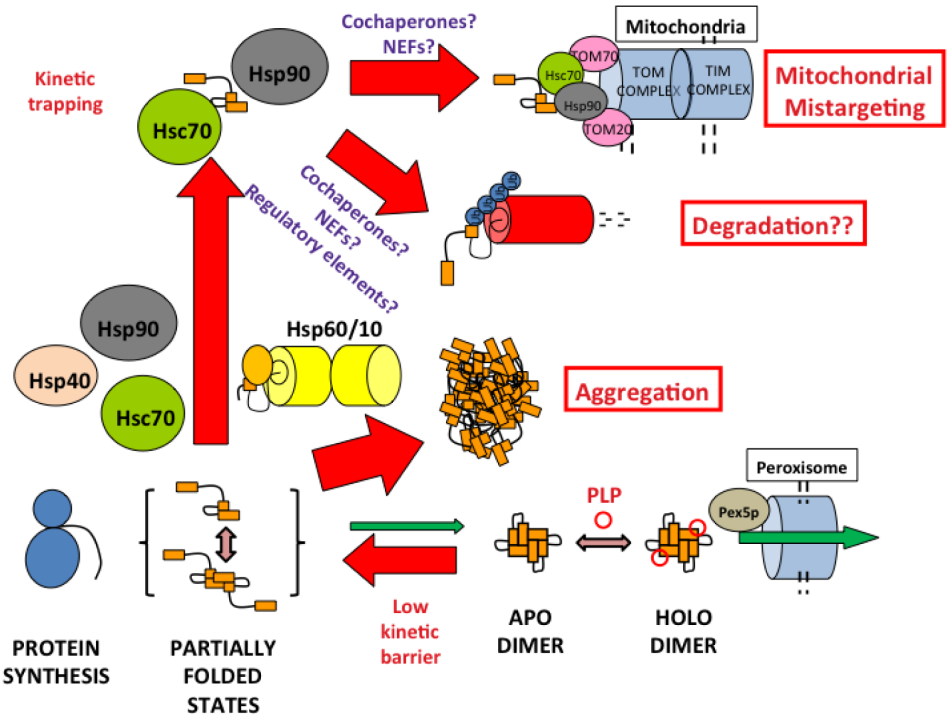


**Figure 8. Folding and misfolding checkpoints of PH1 causing mutants.**
doi:10.1371/journal.pone.0071963.g008

**Table 3.** Summary of mutational effects on molecular properties of AGT protein.

| AGT enzyme | Kinetic stability | | Eukaryotic cell expression | | | | Interaction with Hsc70 | Structural effect |
|---|---|---|---|---|---|---|---|---|
| | Holo | Apo | Total AGT protein | Soluble AGT protein | AGT activity | Targeting | | |
| WT | ↑ | ↑ | ↑↑ | ↑ | ↑ | Peroxisomal | ↓↓ | |
| p.P11L | ↓ | ↓ | = | = | = | Peroxisomal | = | Destabilizing clashes between protomers (this work) |
| p.I340M | ↑↑ | ↑↑ | ↑↑ | ↑ | ↑ | Peroxisomal | ↓↓ | New stabilizing interactions with Q23 and D52 (this work) |
| LM | = | = | = | = | = | Peroxisomal | = | Combination of P11L and I340M (this work) |
| p.H83R | ↓↓↓ | ↓↓ | = | = | ↓↓ | Peroxisomal | =↑ | Structural rearrangement at the active site and destabilization of bound cofactors (Figure 7B; this work) |
| p.F152I | ↓ | ↓↓ | ↓ | ↓ | ↓ | Mitochondrial | =↑ | Conservative change (this work); Creates a cavity in the active site displacing W108 which stacks PMP [35] |
| p.G170R | = | ↓↓ | = | ↓↓ | ↓↓ | Mitochondrial | =↑ | Local conformational rearrangement [34] |
| p.R197Q | ↑ | = | ↑↑ | ↑ | ↑ | Peroxisomal | =↑ | Disruption of a salt bridge (R197-D167) and formation of a new hydrogen bond (Q197-K5; Figure 7B; this work) |
| p.I244T | = | ↓↓ | ↓ | ↓ | ↓ | Peroxisomal | =↑ | α-helix destabilization by a new hydrogen bond [23] |
| p.A295T | = | ↓↓ | ↑ | = | = | Peroxisomal | ↑ | Destabilization of domain:domain interface by a bulkier side-chain (Figure 7B). |
| p.P319L | = | ↓↓ | ↓ | ↓↓ | ↓↓ | Peroxisomal | ↑ | Destabilization of domain:domain interface by a bulkier side-chain (Figure 7B). P319 is at the interface. |
| p.A368T | = | = | = | = | = | Peroxisomal | =↑ | Destabilization of domain:domain interface by a bulkier side-chain (Figure 7B). |

Molecular properties (kinetic stability, expression analyses in CHO cells and interaction with Hsc70 chaperones) are semiquantitatively compared taking AGT LM as reference.
The symbols indicate: ↑/↓, increased/decreased; =↑, slightly increased; =, unchanged.
doi:10.1371/journal.pone.0071963.t003

The intracellular homeostasis of AGT protein seems to rely on a delicate balance between protein folding, misfolding, degradation and intracellular trafficking. Importantly, the AGT LM protein, a variant which is not disease-causing but it is known to sensitize AGT towards deleterious mutations, is shown to notably reduce protein kinetic stability and to enhance protein misfolding and degradation, while most of disease-causing mutations further exacerbate at least some of these defects (Table 3). The partial correlation between these molecular defects suggests that multiple elements in the protein homeostasis networks play a role in determining the fate of PH1 mutants (including chaperones, cochaperones and regulatory proteins [13] and vitamin B6 salvage enzymes implicated in the recycling and targeting of PLP to apo-enzymes [42]). This complexity in the homeostasis of AGT proteins, as well as individual differences in the protein homeostasis network (which may even occur among isogenic individuals [43]), may explain inter-individual variability in clinical presentations and residual activities for patients sharing a given genotype [7,44] and the different fate of mutant proteins (aggregation vs. mitochondrial mistargeting) when expressed under different experimental conditions (this work and [36]). Despite our findings, those specific events and interactions responsible for the partition between protein mitochondrial mistargeting, aggregation and degradation remain elusive.

In the present work, we show that all PH1 mutations of the *minor* haplotype strongly interact with Hsc70 chaperones, adding to our previous work on p.I244T and p.G170R that also showed enhanced interactions with Hsc90 [6,9] and bacterial Hsp60 [15]. We have recently reported that p.G170R interacts with Hsc70 and Hsp90 chaperones through a *molten globule* folding intermediate [6], while p.I244T interacts with Hsp60 chaperones in partially folded monomeric state with the folded N-terminal and C-terminal domains in an *open* conformation [15]. We thus propose that the last steps involving docking of tertiary structure elements and acquisition of the dimeric quaternary structure are crucial for proper folding of AGT. Moreover, enhanced interaction of PH1 mutants with these molecular chaperones suggest a rougher folding landscape for these mutants (with a higher population of kinetic/equilibrium intermediates [13,45,46]). Thus, molecular chaperones emerge as an important checkpoint in the folding of PH1 mutants, likely by partitioning AGT folding intermediates into productive formation of native dimers and peroxisomal import, presentation of partially folded states to the mitochondrial import systems, aggregation and proteasomal degradation [7]. Hsp70, Hsp60 and Hsp90 chaperone systems are known to cooperate in assisting protein folding, and the regulation of chaperone activity by cochaperones and regulatory proteins may lead to different fates (i.e. folding vs. degradation) for the client proteins [13]. Overall, all these results suggest that at least these three chaperone systems (Hsp60, Hsp70 and Hsp90) are potential targets for correction of the folding defects displayed by PH1 mutants. Consequently, the detailed characterization of the chaperone requirements for efficient folding of PH1 mutants will open new approaches for therapeutic intervention in PH1. We have already initiated such studies in cell and animal models of PH1 (ongoing work).

## Conclusions

In this work, we present a multidisciplinary approach that provides clues on the protein homeostasis defects displayed by PH1 causing mutations leading to protein aggregation and mistargeting. We observe a significant correlation between mutation-induced kinetic destabilization of the apo-AGT dimer, kinetic trapping by molecular chaperones and intracellular protein foldability and mistargeting. Detailed kinetic and structure-energetics analyses also show that cofactor induced overstabilization of some mutants is caused by subtle changes in the enthalpic/entropic contributions to denaturation free energy barriers, which may also explain the pyridoxine responsiveness found in patients carrying these mutations. We propose that native state kinetic stabilizers and protein homeostasis modulators may be suitable pharmacological therapies to correct folding and stability defects in PH1.

## Supporting Information

**Figure S1 Interaction of Pex5p-pbd and AGT-WT by isothermal titration calorimetry (ITC).** A) Raw calorimetric data; B) Binding isotherm (the line shows the best-fit to one-independent-type-of sites).
(TIF)

**Figure S2 Equilibrium (A) and kinetic (B) PLP binding experiments to apo-AGT.** Line in panel A shows the best fit to a 1:1 equilibrium binding model; Line in panel B are linear fits of the experimental data, the slope providing the value of $k_{on}$ and the y-intercept the value of $k_{off}$. Data are from means±s.d. from three independent experiments.
(TIF)

**Figure S3 Exponential relationship between the kinetic stability at physiological temperature and $T_m$ values for holo- (open symbols) and apo-(closed symbols) AGT enzymes.**
(TIF)

**Table S1 Thermodynamic binding parameters for the interaction between holo-AGT variants with Pex5p-pbd.**
(DOC)

**Table S2 Data collection and refinement statistics.**
(DOC)

**Materials S1 Materials and Methods.**
(DOC)

## Acknowledgments

We thank Dr. J.A. Gavira and Ms. R. Fernández for their assistance in DLS experiments and Ms. B. Rodriguez for her help with *E. coli* expression.

## Author Contributions

Conceived and designed the experiments: ALP ES AA. Performed the experiments: NM-T IF-R CY AA ES ALP. Analyzed the data: NM-T IF-R CY AA ES ALP. Contributed reagents/materials/analysis tools: ALP ES AA DR. Wrote the paper: ALP ES AA.

## References

1. Williams EL, Acquaviva C, Amoroso A, Chevalier F, Coulter-Mackie M, et al. (2009) Primary hyperoxaluria type 1: update and additional mutation analysis of the AGXT gene. Hum Mutat 30: 910–917.
2. Danpure CJ (2006) Primary hyperoxaluria type 1: AGT mistargeting highlights the fundamental differences between the peroxisomal and mitochondrial protein import pathways. Biochim Biophys Acta 1763: 1776–1784.
3. Cochat P, Hulton SA, Acquaviva C, Danpure CJ, Daudon M, et al. (2012) Primary hyperoxaluria Type 1: indications for screening and guidance for diagnosis and treatment. Nephrol Dial Transplant 27: 1729–1736.
4. Cellini B, Montioli R, Voltattorni CB (2011) Human liver peroxisomal alanine:glyoxylate aminotransferase: characterization of the two allelic forms and their pathogenic variants. Biochim Biophys Acta 1814: 1577–1584.

5. Lumb MJ, Birdsey GM, Danpure CJ (2003) Correction of an enzyme trafficking defect in hereditary kidney stone disease in vitro. Biochem J 374: 79–87.
6. Pey AL, Salido E, Sanchez-Ruiz JM (2011) Role of low native state kinetic stability and interaction of partially unfolded states with molecular chaperones in the mitochondrial protein mistargeting associated with primary hyperoxaluria. Amino Acids 41: 1233–1245.
7. Salido E, Pey AL, Rodriguez R, Lorenzo V (2012) Primary hyperoxalurias: Disorders of glyoxylate detoxification. Biochim Biophys Acta 1822: 1453–1464.
8. Coulter-Mackie MB, Lian Q (2006) Consequences of missense mutations for dimerization and turnover of alanine:glyoxylate aminotransferase: study of a spectrum of mutations. Mol Genet Metab 89: 349–359.
9. Santana A, Salido E, Torres A, Shapiro LJ (2003) Primary hyperoxaluria type 1 in the Canary Islands: a conformational disease due to I244T mutation in the P11L-containing alanine:glyoxylate aminotransferase. Proc Natl Acad Sci U S A 100: 7277–7282.
10. Martin W (2010) Evolutionary origins of metabolic compartmentalization in eukaryotes. Philos Trans R Soc Lond B Biol Sci 365: 847–855.
11. Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE (2009) Biological and chemical approaches to diseases of proteostasis deficiency. Annu Rev Biochem 78: 959–991.
12. Balch WE, Morimoto RI, Dillin A, Kelly JW (2008) Adapting proteostasis for disease intervention. Science 319: 916–919.
13. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475: 324–332.
14. Gomes CM (2012) Protein misfolding in disease and small molecule therapies. Curr Top Med Chem 12: 2460–2469.
15. Albert A, Yunta C, Arranz R, Pena A, Salido E, et al. (2010) Structure of GroEL in complex with an early folding intermediate of alanine glyoxylate aminotransferase. J Biol Chem 285: 6371–6376.
16. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4: 2411–2423.
17. Cellini B, Bertoldi M, Montioli R, Paiardini A, Borri Voltattorni C (2007) Human wild-type alanine:glyoxylate aminotransferase and its naturally occurring G82E variant: functional properties and physiological implications. Biochem J 408: 39–50.
18. Rumsby G, Weir T, Samuell CT (1997) A semiautomated alanine:glyoxylate aminotransferase assay for the tissue diagnosis of primary hyperoxaluria type 1. Ann Clin Biochem 34 (Pt 4): 400–404.
19. Rodriguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM (2006) Role of solvation barriers in protein kinetic stability. J Mol Biol 360: 715–724.
20. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66: 125–132.
21. (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 50: 760–763.
22. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674.
23. Zhang X, Roe SM, Hou Y, Bartlam M, Rao Z, et al. (2003) Crystal structure of alanine:glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1. J Mol Biol 331: 643–652.
24. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213–221.
25. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
26. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66: 12–21.
27. DeLano WL (2002) Pymol. DeLano Scientific LLC.
28. Fodor K, Wolf J, Erdmann R, Schliebs W, Wilmanns M (2012) Molecular requirements for peroxisomal targeting of alanine-glyoxylate aminotransferase as an essential determinant in primary hyperoxaluria type 1. PLoS Biol 10: e1001309.
29. Cellini B, Montioli R, Paiardini A, Lorenzetto A, Maset F, et al. (2010) Molecular defects of the glycine 41 variants of alanine glyoxylate aminotransferase associated with primary hyperoxaluria type I. Proc Natl Acad Sci U S A 107: 2896–2901.
30. Sanchez-Ruiz JM (1992) Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. Biophys J 61: 921–935.
31. Robertson AD, Murphy KP (1997) Protein Structure and the Energetics of Protein Stability. Chem Rev 97: 1251–1268.
32. Costas M, Rodriguez-Larrea D, De Maria L, Borchert TV, Gomez-Puyou A, et al. (2009) Between-species variation in the kinetic stability of TIM proteins linked to solvation-barrier free energies. J Mol Biol 385: 924–937.
33. Pey AL, Mesa-Torres N, Chiarelli LR, Valentini G (2013) Structural and Energetic Basis of Protein Kinetic Destabilization in Human Phosphoglycerate Kinase 1 Deficiency. Biochemistry 52: 1160–1170.
34. Djordjevic S, Zhang X, Bartlam M, Ye S, Rao Z, et al. (2010) Structural implications of a G170R mutation of alanine:glyoxylate aminotransferase that is associated with peroxisome-to-mitochondrion mistargeting. Acta Crystallogr Sect F Struct Biol Cryst Commun 66: 233–236.
35. Cellini B, Montioli R, Paiardini A, Lorenzetto A, Voltattorni CB (2009) Molecular Insight into the Synergism between the Minor Allele of Human Liver Peroxisomal Alanine:Glyoxylate Aminotransferase and the F152I Mutation. J Biol Chem 284: 8349–8358.
36. Fargue S, Lewin J, Rumsby G, Danpure CJ (2013) Four of the most common mutations in primary hyperoxaluria type 1 unmask the cryptic mitochondrial targeting sequence of alanine:glyoxylate aminotransferase encoded by the polymorphic minor allele. J Biol Chem 288: 2475–2484.
37. Fargue S, Rumsby G, Danpure CJ (2013) Multiple mechanisms of action of pyridoxine in primary hyperoxaluria type 1. Biochim Biophys Acta.
38. Monico CG, Rossetti S, Olson JB, Milliner DS (2005) Pyridoxine effect in type I primary hyperoxaluria is associated with the most common mutant allele. Kidney Int 67: 1704–1709.
39. van Woerden CS, Groothoff JW, Wijburg FA, Annink C, Wanders RJ, et al. (2004) Clinical implications of mutation analysis in primary hyperoxaluria type 1. Kidney Int 66: 746–752.
40. Cellini B, Lorenzetto A, Montioli R, Oppici E, Voltattorni CB (2010) Human liver peroxisomal alanine:glyoxylate aminotransferase: Different stability under chemical stress of the major allele, the minor allele, and its pathogenic G170R variant. Biochimie 92: 1801–1811.
41. Pittman AM, Lage MD, Poltoratsky V, Vrana JD, Paiardini A, et al. (2012) Rapid profiling of disease alleles using a tunable reporter of protein misfolding. Genetics 192: 831–842.
42. di Salvo ML, Contestabile R, Safo MK (2011) Vitamin B(6) salvage enzymes: mechanism, structure and regulation. Biochim Biophys Acta 1814: 1597–1608.
43. Casanueva MO, Burga A, Lehner B (2012) Fitness trade-offs and environmentally induced mutation buffering in isogenic C. elegans. Science 335: 82–85.
44. Danpure CJ, editor(2001) Primary hyperoxalurias.
45. Chakraborty K, Chatila M, Sinha J, Shi Q, Poschner BC, et al. (2010) Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. Cell 142: 112–122.
46. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122: 209–220.
47. Sanchez-Ruiz JM, Lopez-Lacomba JL, Cortijo M, Mateo PL (1988) Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. Biochemistry 27: 1648–1652.

# APPENDIX B

Biochem. J. (2014) **462**, 453–463 (Printed in Great Britain) doi:10.1042/BJ20140250





# The consensus-based approach for gene/enzyme replacement therapies and crystallization strategies: the case of human alanine–glyoxylate aminotransferase

Noel MESA-TORRES*, Cristina YUNTA†, Israel FABELO-ROSA‡, Juana María GONZALEZ-RUBIO†, José M. SÁNCHEZ-RUIZ*, Eduardo SALIDO‡[1], Armando ALBERT†[1,2] and Angel L. PEY*[1,2]

*Department of Physical Chemistry, Faculty of Sciences, University of Granada, Av/Fuentenueva s/n, 18071 Granada, Spain
†Departmento de Cristalografía y Biología Estructural, Instituto de Química-Física "Rocasolano", Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain
‡Center for Biomedical Research on Rare Diseases (CIBERER), University Hospital of the Canary Islands and CIBICAN, University of La Laguna, 38320 Tenerife, Spain

Protein stability is a fundamental issue in biomedical and biotechnological applications of proteins. Among these applications, gene- and enzyme-replacement strategies are promising approaches to treat inherited diseases that may benefit from protein engineering techniques, even though these beneficial effects have been largely unexplored. In the present study we apply a sequence-alignment statistics procedure (consensus-based approach) to improve the activity and stability of the human AGT (alanine–glyoxylate aminotransferase) protein, an enzyme which causes PH1 (primary hyperoxaluria type I) upon mutation. By combining only five consensus mutations, we obtain a variant (AGT-RHEAM) with largely enhanced *in vitro* thermal and kinetic stability, increased activity, and with no side effects on foldability and peroxisomal targeting in mammalian cells. The structure of AGT-RHEAM reveals changes at the dimer interface and improved electrostatic interactions responsible for increased kinetic stability. Consensus-based variants maintained the overall protein fold, crystallized more easily and improved the expression as soluble proteins in two different systems [AGT and CIPK24 (CBL-interacting serine/threonine-protein kinase) SOS2 (salt-overly-sensitive 2)]. Thus the consensus-based approach also emerges as a simple and generic strategy to increase the crystallization success for hard-to-get protein targets as well as to enhance protein stability and function for biomedical applications.

Key words: inherited disease, protein crystallization, protein function, protein stability, sequence-alignment statistics, therapy.

## INTRODUCTION

Inborn errors of liver metabolism are rare inherited disorders with significant morbidity and mortality. For many of them, liver transplantation and dietary restriction are the only therapeutic option currently available to increase the lifespan of patients and, thus, novel treatments such as GT (gene therapy) and ERT (enzyme replacement therapy) are being investigated. For GT, an important issue is the percentage of hepatocytes to be corrected with a therapeutic gain and low toxicity (e.g. as little vector dose as possible). Modifications of the coding sequence included in the vector for GT is an attractive strategy to improve its efficiency. Similarly, ERT efficiency can be enhanced by protein engineering of the coding sequence. Thus GT and ERT strategies may benefit from an improvement of the protein target by protein engineering techniques to prolong therapeutic effects and/or reduce the dose required for therapeutic action [1,2].

PH1 (primary hyperoxaluria type I) is a severe inborn error of liver metabolism due to the inherited deficit of AGT (alanine–glyoxylate aminotransferase; E.C. 2.6.1.44), which conveys high oxalate production by the liver and subsequent renal damage [3]. AGT is a peroxisomal PLP (pyridoxal-5′-phosphate)-dependent enzyme that catalyses the amino transfer from L-alanine to glyoxylate to yield pyruvate and glycine. Current treatments for PH1 often rely on a combined liver and kidney transplantation [4,5]. Most disease-associated PH1 mutations cause protein misfolding (peroxisomal aggregation and/or accelerated protein turnover) or mitochondrial mistargeting (instead of normal peroxisomal targeting) (see [5] and references therein). GT and ERT are new therapeutic strategies for PH1 [6,7]. Working with an animal model for PH1, we have previously shown that about 40 % hepatocytes have to be transduced with the WT (wild-type) human AGT in order to normalize oxalate production [8].

The 'consensus approach' to improve protein stability is a simple strategy that uses statistical information contained in sequence alignments of proteins with moderate to high homology [9]. Consensus mutations are often viewed as reversions to the ancestral amino acid residues [10] and, consequently, protein stability enhancement could be reasonably attributed to the fact that consensus-sequence proteins may capture to some extent the extreme properties (such as thermophilicity) of ancestral proteins [11–13]. The consensus approach has been successfully applied to improve the thermodynamic and thermal stability of several proteins [14–20]. Interestingly, some of these studies have also

Abbreviations: AGT, alanine–glyoxylate aminotransferase; CCD, charge-coupled device; CHO, Chinese-hamster ovary; CIPK, CBL-interacting serine/threonine-protein kinase; DSC, differential scanning calorimetry; ERT, enzyme replacement therapy; GT, gene therapy; ITC, isothermal titration calorimetry; Pex5p-pbd, Pex5p-PTS1 binding domain; PH1, primary hyperoxaluria type I; PLP, pyridoxal-5′-phosphate; SOS2, salt-overly-sensitive 2; WT, wild-type.

[1] These authors share senior authorship.

[2] Correspondence may be addressed to either of these authors (email xalbert@iqfr.csic.es or angelpey@ugr.es).

The co-ordinates and structure factor amplitudes have been deposited in the PDB under codes 4CBR and 4CBS for AGT-HEM and AGT-RHEAM respectively.

shown that the consensus-based approach also improves protein kinetic stability [16–18]. Therefore the consensus approach appears to be a simple and excellent choice to develop robust proteins for GT and ERT, provided that these changes do not trigger significant immune response. Moreover, a recent study has also shown that the consensus approach can simultaneously increase protein activity and stability (analysed in [18]).

In the present study, we have used the consensus-based approach to improve the stability and function of human AGT. A combination of five mutations (Q23R/S48H/D52E/V113A/I340M, referred to as AGT-RHEAM) improves protein kinetic stability (as a half-life for denaturation at physiological temperature) by 2–3 orders of magnitude and also, remarkably, increases enzyme activity 2.2-fold. The X-ray crystal structure of the AGT-RHEAM protein reveals that several mutations target the AGT monomer–monomer interface, contributing to improve the protein kinetic stability of the unfolding co-operative unit (i.e. the dimer) [24]. AGT-RHEAM is well expressed and imported into peroxisomes and its activity is enhanced in eukaryotic cells. In addition, we report that consensus-based AGT proteins crystallized more easily than the WT protein, which was confirmed using the plant SOS2 (salt-overly-sensitive 2)/CIPK24 (CBL-interacting serine/threonine-protein kinase 24) protein [21], a protein that is very hard to express and crystallize. This suggests that the protein stabilization and increased solubility provided by the consensus approach could be a simple and generic method to enhance the ability of a protein sample to crystallize as an alternative to the use of chemical ligands and optimal solvent conditions [22,23].

## EXPERIMENTAL

### Sequence alignment and consensus-based identification of potentially stabilizing mutations

Stabilizing mutations for human AGT were searched on the basis of the consensus approach to protein stability [9]. Briefly, the human AGT sequence (NCBI GI: 4557289 and Reference Sequence: NM_000030) was aligned with 21 AGT sequences from vertebrates and 88 AGT sequences from eukaryotes (obtained using BlastP) using the MUSCLE program (http://www.drive5.com/muscle/). These sequence alignments were used to determine those mutations displaying highest frequencies in the alignments that were not present in the human AGT sequence. Eight mutations with ratios (highest frequency in the alignment/frequency of the amino acid present in human AGT) $>8$ in the alignment with sequences from vertebrates were selected for recombinant expression and purification in *Escherichia coli* cells and further analysis. An identical procedure was followed for the kinase domain of CIPK24/SOS2. For CIPK24/SOS2, the sequence from *Arabidopsis thaliana* (At5g35410) was aligned with 500 CIPK sequences from viridiplantae. The N- and C-terminals of the proteins were excluded from the analysis. Five mutations out of 12 possible candidates on the kinase domain of *A. thaliana* CIPK24/SOS2 were chosen and two multiple mutants generated in *E. coli* cells for further analyses: 3MxCIPK24/SOS2 (E107K/S109D/C127S) and 5MxCIPK24/SOS2 (P81K/E107K/S109D/C127S/L266K).

### Protein expression and purification

*E. coli* BL21(DE3) bacterial cells containing pCold II plasmids (TaKaRa) encoding AGT protein variants or Pex5p-pbd (Pex5p-PTS1 binding domain) were grown and proteins expressed and purified as described in [24]. Holo- and apo-AGT were prepared as described in [25] and stored as described in [26]. Full details of CIPK24/SOS2 cloning, expression and purification along with determination of the crystal structure solution have also been performed (J.M. Gonzalez-Rubio and A. Albert, unpublished work). Briefly, the kinase domain of CIPK24/SOS2 (residues 7–308) and the mutated constructs were cloned into the pGEX4T2 expression vector (GE Healthcare) and recombinant proteins were expressed in Rosetta 2 (DE3) pLysS cells (Stratagene) at 16 °C. Recombinant proteins were purified using glutathione Sepharose 4B (GE Healthcare) followed by an overnight preScission cleavage. Samples were further purified by gel-filtration chromatography on a Hi Load 16/60 Superdex 200 prep-grade column previously equilibrated in 20 mM Tris/HCl (pH 9.0), 100 mM NaCl and 5 mM DTT (GE Healthcare). Protein concentration was measured spectrophotometrically using $\varepsilon_{280(1\text{mg/ml})} = 1.069$ (AGT), 1.243 (Pex5p-pbd) and 0.964 (CIPK24/SOS2) derived from their primary sequence [27].

### Differential scanning calorimetry

Measurements were performed on a capillary VP-DSC (differential scanning calorimetry; MicroCal; GE Healthcare) calorimeter with a cell volume of 0.135 ml at the indicated scan rates. DSC scans were customarily performed in 20 mM sodium Hepes and 200 mM NaCl (pH 7.4) (AGT proteins) or 100 mM sodium phosphate and 20 mM NaCl (pH 8) (SOS2 proteins), in a temperature range of 20–110 °C, and using a protein concentration of 5 $\mu$M (in protein subunit) as described in [26]. DSC profiles were analysed on the basis of a two-state irreversible model depicted by the scheme $N \rightarrow F$, where $N$ is the native state and $F$ is the final state which cannot fold back to $N$ under the given experimental conditions [28,29]. Additional details of the DSC analyses have been described previously [26].

### Spectroscopic analyses

All spectroscopic analyses were performed in 20 mM sodium Hepes and 200 mM NaCl (pH 7.4) at 25 °C as described previously [24].

### Binding of AGT proteins to Pex5p-pbd by isothermal titration calorimetry

ITC (isothermal titration calorimetry) measurements were carried out as described in [24].

### Activity measurements

Steady-state measurements were typically performed using 5 $\mu$g/ml of AGT incubated in 0.1 M sodium phosphate buffer (pH 8) at 25 °C in the presence of 40 $\mu$M PLP and 0–2 mM glyoxylate and the reaction was triggered by the addition of 0–100 mM L-alanine. Each set of experiments included activity measurements at 32 pairs of alanine/glyoxylate concentrations. Pyruvate formed in the reactions was measured in a Tecan Infinite M200 Pro microplate reader by a coupled NADH/lactate dehydrogenase assay after 2 min of reaction at 25 °C [30]. Global fittings of activity measurements were performed according to a double-displacement mechanism as described in [24].

### Protein crystallization

Before crystallization AGT-HEM and AGT-RHEAM were dialysed in 50 mM Tris/HCl (pH 7.5) and 250 mM NaCl, and concentrated to 13.5 mg/ml. CIPK24/SOS2 and

5MxCIPK24/SOS2 were concentrated to 5.0 mg/ml. We employed a 10 kDa cutoff Amicon protein concentrator (YM-10; Millipore Corporation) and kept the proteins at 4 °C. Crystallization assays were carried out on a 60-well microbath under oil (Terasaki plates) at 18 °C. The best AGT-HEM and AGT-RHEAM crystals were obtained using a precipitant solution containing 15 % PEG 10K, 0.1 M Hepes (pH 7.5), 2.5 % (v/v) isopropanol and 5 % glycerol as additives. For 5MxCIPK24/SOS2 the best crystallization conditions were obtained using a precipitant solution containing 24 % PEG 4K and 20 mM Tris/HCl (pH 8.5). Crystals were obtained by mixing drops of 2 $\mu$l of protein solution and 2 $\mu$l of precipitant solution that grew over a few hours. Several strategies were used to optimize these crystallization conditions, which included adjusting the protein sample composition, the precipitant concentration and pH values.

To estimate the crystallization propensity of the consensus optimized proteins compared with the WT forms we performed a number of crystallization experiments in which we varied the pH and precipitant concentration around the optimal crystallization condition. For AGT-RHEAM compared with AGT-WT, we screened at pH 7.0, 7.5 and 7.8 and PEG 10K at 8 %, 10 %, 12 % and 15 %, and for 5MxCIPK24/SOS2 compared with CIPK24/SOS2 we screened pH 8.2, 8.5, 8.8 and 9 and PEG 4K at 16 %, 18 %, 20 % and 22 %, monitoring the number of crystallization hits at different times (0–36 h). We considered as a crystallization hit any kind of protein crystal observable under a $\times$50 magnification microscope regardless of the size or shape of the crystal. AGT hits were identified as yellow crystals due to the PLP bound to Lys$^{209}$ and displayed a large prismatic shape (up to 0.5 mm in the longest direction). By contrast, 5MxCIPK24/SOS2 hits were colourless small thin plates initially identified as protein crystals using Izit crystal dye (Hampton Research). Subsequent crystallographic analysis confirmed in both cases that the hits were protein crystals.

### Data collection and structure resolution

Crystals were mounted in a fibre loop, transferred to the cryoprotectant (20 % glycerol on the crystal mother liquor solution) and flash-frozen at 100 K in a nitrogen gas steam. AGT-HEM and AGT-RHEAM crystals displayed good-quality diffraction patterns and a diffraction data set was collected using an ADSC Q4 CCD (charge-coupled device) detector at the ID14.4 beamline of the European Synchrotron Radiation Facility (Grenoble). Diffraction data were processed with XDS [31] and scaled with SCALA from the CCP4 package. AGT-HEM diffracted to a 2.5 Å resolution and belonged to space group $P4_12_12$, with unit-cell parameters $a = 90.2$, $b = 90.2$ and $c = 141.0$. AGT-RHEAM diffracted to a 2.4 Å resolution and belonged to space group $P4_12_12$, with unit-cell parameters $a = 89.6$, $b = 89.6$ and $c = 142.4$. Both structures were solved by molecular replacement using Phaser [32] with the co-ordinates from AGT-I340M (PDB code 2YOB). Several cycles of restrained refinement with PHENIX [33] and iterative model building with COOT [34] yielded the final model with an $R/R_{free}$ of 22.0/26.8 for AGT-HEM and 22.0/25.7 for AGT-RHEAM. The water structure was also modelled. A summary of the diffraction protocol, data collection and refinement statistics is given in Supplementary Table S1 (http://www.biochemj.org/bj/462/bj4620453add.htm).

A diffraction data set from 5MxCIPK24/SOS2 crystals was collected using an ADSC Q4 CCD detector at the Petra-III synchrotron, Hamburg, beamline P13. The data set was processed using XDS [31] and scaled using SCALA [35] from the CCP4 package [36]. A summary of the data collection statistics is given in Supplementary Table S1.

The stereochemistry of the models was verified with MolProbity [37]. Ribbon figures were produced using PyMOL (http://www.pymol.org). The co-ordinates and structure factor amplitudes have been deposited in the PDB under codes 4CBR and 4CBS for AGT-HEM and AGT-RHEAM respectively.

### Electrostatic calculations

The calculations of the energy of charge–charge interactions ($E_{q\text{-}q}$) were performed using the solvent-accessibility-corrected Tanford–Kirkwood model [39]. A detail description of this procedure can be found in [40]. The input for these calculations were the atomic co-ordinates of dimeric AGT-WT [41] and AGT-HEM and AGT-RHEAM (the present study), and the calculations were performed using home-built software at pH 7.4 and 200 mM ionic strength to mimic the conditions used in the DSC experiments. The output of these analyses provide the energy of charge–charge interactions of a given ionizable residue with all the ionizable residues in the protein dimer ($E_{q\text{-}q}$).

### AGT expression in cell cultures

CHO (Chinese-hamster ovary) cells were transfected using AGT cDNA variants subcloned in pCIneo plasmids (Promega). Details for cell growth, transfection, Western blot, activity measurements and immunofluorescence microscopy analyses have been described recently [24].

## RESULTS

### The consensus-based approach improves AGT stability and activity *in vitro*

First, we aligned the human AGT and 21 AGT sequences found in vertebrates (Figure 1A), and then calculated the ratio of frequencies between the most common amino acid found in a certain position in the multiple alignment and the frequency of the amino acid found in the human sequence (Figure 1B). On the basis of the consensus approach [9], mutations with ratios higher than 1 should increase protein stability. It is worth noting that these single point mutations suggested by the sequence alignment from vertebrates are also predicted to be stabilizing using a wider sequence alignment performed using human AGT and 88 sequences from eukaryotes (Supplementary Table S2 at http://www.biochemj.org/bj/462/bj4620453add.htm).

To test the consensus-based approach, we have first studied six of the eight single mutants suggested by the consensus approach and determined their thermal stability by DSC (Figure 2 and Supplementary Table S3 at http://www.biochemj.org/bj/462/bj4620453add.htm). The T9P and R197Q mutants were not investigated since we could not obtain good expression yields under different expression conditions. Thermal denaturation of the remaining six variants (as well as the WT protein) was found to be irreversible and strongly scan-rate dependent (for representative DSC scans, see Supplementary Figure S1 at http://www.biochemj.org/bj/462/bj4620453add.htm). As previously shown for WT-AGT [26], a two-state irreversible model with first-order kinetics satisfactorily describes thermal unfolding of all holo-AGT variants (Supplementary Figure S1 and Supplementary Table S4 at http://www.biochemj.org/bj/462/bj4620453add.htm). Five of the six single consensus-based variants show enhanced thermal stability ($\Delta T_m \sim 2$–$3$ °C; Figure 2A and Supplementary

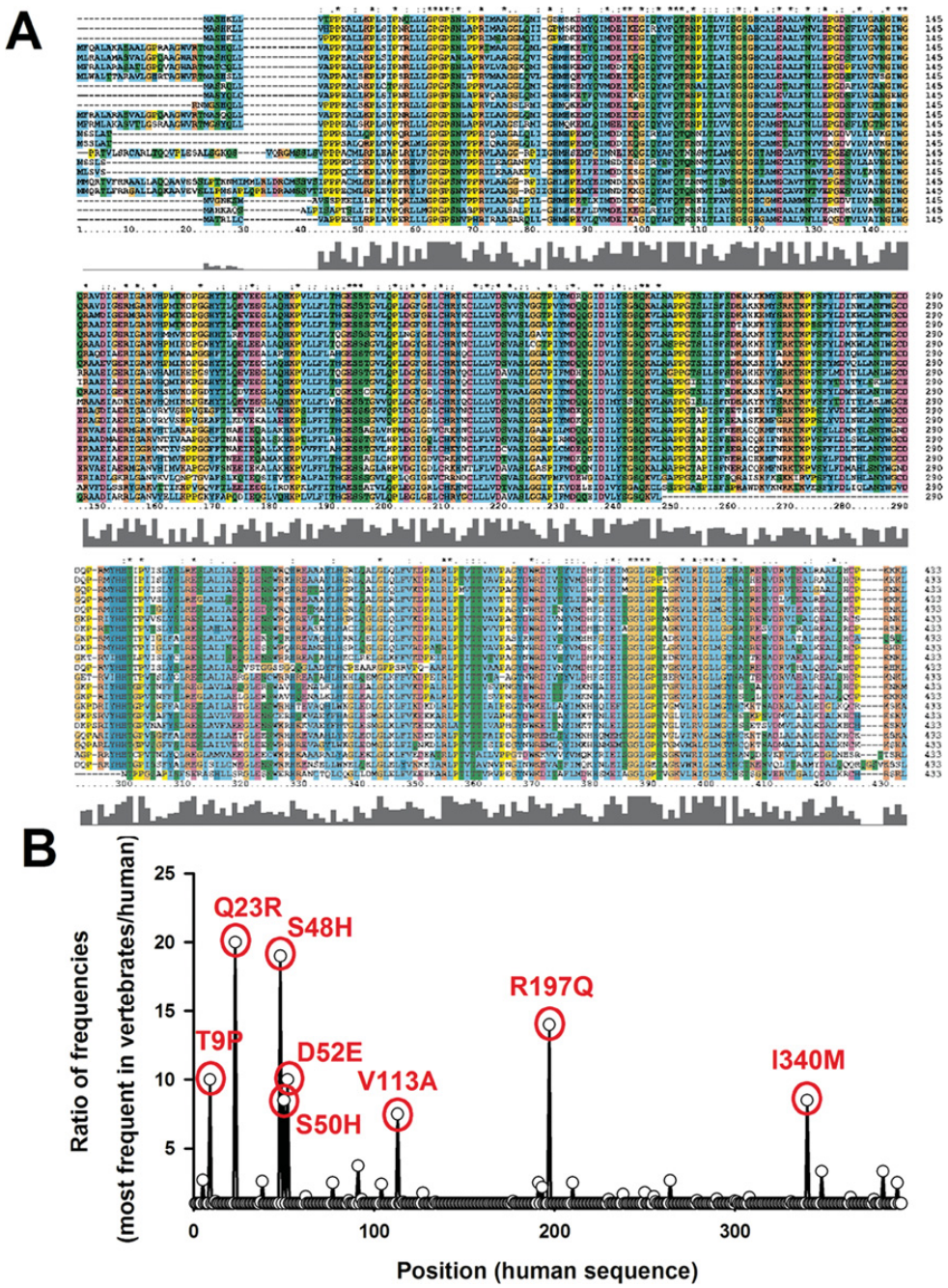


**Figure 1 Consensus-based analysis of AGT protein sequence alignment**

(**A**) Sequence alignment of AGT proteins from vertebrates. GI codes are (top to bottom): 4557289 (human AGT), 197101617, 109101707, 176612, 194211512, 1174432, 156121353, 164579, 2239081, 56972365, 73994226, 6180226, 55741900, 50417404, 45709105, 229366512, 47212538, 225706550, 209738520, 115941749, 196007736 and 224059805. (**B**) Consensus-based analyses: ratio of frequencies of the most frequent amino acid in the sequence alignment and of the amino acid found in the human AGT sequences expressed along the human AGT protein sequence.

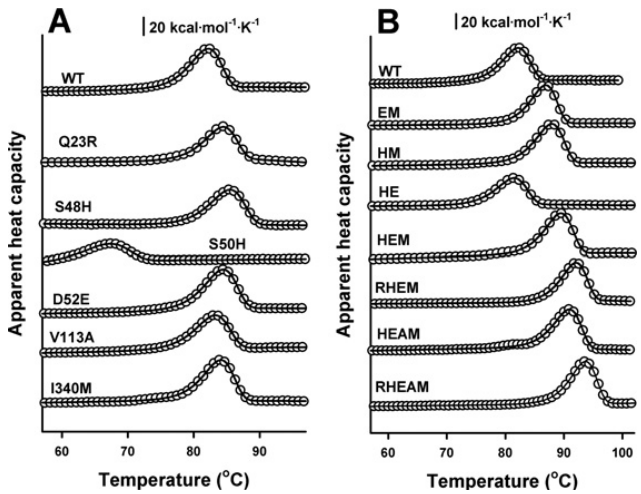


**Figure 2 Thermal denaturation profiles of WT and consensus-based variants of AGT as holo-proteins**

Single (**A**) and multiple (**B**) mutants. Lines are fits to a two-state irreversible denaturation model with first-order kinetics [29]. Experiments were carried out at 3 °C/min.

Table S3) compared with the WT protein. When some of these mutations are combined, the stabilizing effect was found to be additive, leading to increases in thermal stability in the range of 9–12 °C for the quadruple and quintuple mutants (Figures 2B and 3A and Supplementary Table S3). It is worth noting that the increase in thermal stability (as a $T_m$ up-shift) is exponentially translated to an increase in kinetic stability extrapolated to physiological temperature, yielding increases in protein kinetic stability for the 4-fold (HEAM and RHEM) and 5-fold (RHEAM) variants of ∼3–4 orders of magnitude compared with WT-AGT (Figures 3A and 3B, and Supplementary Figure S2A and Table S3 at http://www.biochemj.org/bj/462/bj4620453add.htm). Activity measurements also show enhanced specific activity for several of the consensus variants, mostly due to the activity enhancement provided by the S48H substitution (Supplementary Figure S2B and Table S3); thus, showing a weak overall correlation between stability (as the $T_m$ value) and activity (Supplementary Figure S2B).

We must note that the kinetic stability analyses reported in the previous paragraph for the holo-variants require very long extrapolations in the Arrhenius plots to be accurate. Nevertheless, similar kinetic analysis on the WT and RHEAM variants as apo-proteins (displaying ∼25 °C lower $T_m$ values compared with the holo-forms, see Figure 3A, and requiring shorter extrapolations to reach physiological temperature) reveal a consistent kinetic stabilization of the RHEAM variant, with a ∼440-fold increase in protein kinetic stability at 37 °C compared with the apo-WT (Figure 3B and Table 1).

### Functional characterization of AGT-RHEAM

We have studied in more detail the functional properties of the highly stable and active RHEAM quintuple variant. Enzyme kinetic analyses based on a double-displacement mechanism revealed an 125 % increase in $V_{max}$ in RHEAM compared with the WT, whereas the $K_m$ values for L-alanine and glyoxylate are only slightly different (Figure 3C and Table 2). We have further investigated the microenvironment surrounding the bound PLP and PMP cofactors in the WT and RHEAM variants by spectroscopic analyses (Figures 3D and 3E). PLP bound to the active sites of these two variants reveal the same spectroscopic signature, with strong absorption and positive dichroic signals centred around 420 and 430 nm respectively, indicating the microenvironment of the cofactor is not affected in this consensus-based mutant. Similarly, PMP bound to both variants shows a characteristic dichroic signal in the near-UV (centred around 320 nm), suggesting no perturbation of the PMP bound in the mutant either. PLP binding affinity determined by quenching of AGT intrinsic tryptophan fluorescence yields similarly strong binding affinities in both variants for PLP (Figure 3F and Table 2).

### Structural basis of the consensus-based enhancement of AGT function and stability: the crystal structures of the AGT-HEM and AGT-RHEAM

The structures of the AGT-HEM and AGT-RHEAM mutants with enhanced thermal stability have been determined by X-ray crystallography at 2.5 Å and 2.4 Å respectively (Figure 4 and Supplementary Table S1). All of the mutations are unambiguously modelled into the electron density. Both structures and the previously reported AGT-WT [41] and AGT-I340M [24] are nearly identical with changes confined at the point mutation positions (Supplementary Figure S3 at http://www.biochemj.org/bj/462/bj4620453add.htm) showing the RMSD between equivalent C$\alpha$ atoms that range from 0.13 Å to 0.28 Å (0.28, 0.13 and 0.27 for AGT-I340M, AGT-HEM and AGT-RHEAM respectively). The AGT structure is a homodimer, each protomer is formed by two domains. Of them, the N-terminal domain is involved in the dimerization and contains residues forming the PLP-binding site.

The analyses of the structures reveal that four of the five point mutations, Q23R, S48H, D52E and I340M, cluster in the same area of the macromolecule at the dimerization interface (Figure 4). The increased size of the mutated-residue side chains produces an increase of the intra- or inter-molecular hydrophobic interactions and/or in the number of hydrogen bonds with respect to the AGT. This yields a small, but noticeable, increase of approximately 2 % in the accessible surface buried in the dimer interface

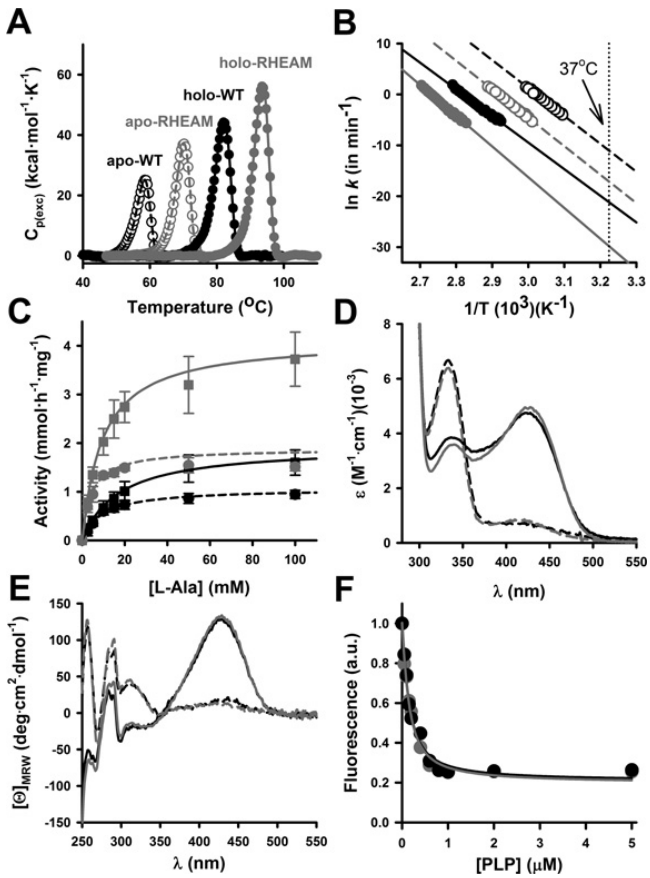


**Figure 3 Comparative stability and functional studies of WT and RHEAM AGT**

(**A** and **B**) DSC scans (**A**; at 3 °C/min) and Arrhenius plots (**B**) for the thermal denaturation of AGT-WT (black symbols) and RHEAM (grey symbols) proteins in their holo- (closed symbols) and apo- (open symbols) forms. (**C**) Enzyme activity dependence of AGT WT (black symbols) and RHEAM (grey symbols) on L-alanine concentration at 0.25 mM (circles) and 2 mM (squares) glyoxylate. Results are means ± S.D. from four independent experiments. Lines show global fittings to a double-displacement mechanism. (**D** and **E**) Absorption (**D**) and CD (**E**) spectra of holo-WT (black lines) and holo-RHEAM (grey lines) incubated with (broken lines) or without (continuous lines) 500 mM L-alanine for 10 min at 25 °C. (**F**) Titration of AGT WT (black) and RHEAM (grey) with PLP monitored by fluorescence.

(calculated using NACCESS from LIGPLOT [42]). Interestingly, the Q23R, D52E and I340M modifications increase the number of stabilizing interactions between protomers. The native hydrogen bond between Gln$^{23}$ OE1 and Gly$^{47}$ NH is reinforced by the AGT-HEM point mutations (Figure 4). The native Gln$^{23}$ NE1 was found to hydrogen bond to the mutated Glu$^{53}$ OE1, whereas the mutated Met$^{340}$ side chain and native Pro$^{21}$ sandwich the hydrophobic moiety of the Gln$^{23}$ side chain. The additional point mutation on the AGT-RHEAM variant Q23R enlarges the side chain at this position, but maintains the same pattern of interactions observed in AGT-HEM by the inclusion of a structural water molecule. In addition, the mutated Arg$^{23}$ NZ1 and native Asp$^{334}$ OD1 forms a salt bridge and the larger hydrophobic moiety of the arginine side chain stabilizes further the Pro$^{21}$/Arg$^{23}$/Met$^{340}$ cluster. The mutated His$^{48}$ side chain fills a solvent accessible cavity contiguous to the interprotomer interface and forms new hydrophobic interactions with Pro$^{30}$ of the adjacent protomer.

The active site architecture is maintained in both mutated proteins when compared with WT-AGT, in agreement with the spectroscopic analysis. However, the side chain of residue 48 lines the surface of the cavity towards the active site. The mutation S48H increases the nucleophilicity and narrows the access to the PLP, which may explain the differences in activity when WT and HEM/RHEAM are compared. Interestingly, these differences do not lead to large changes in the binding affinity for PLP, glyoxylate or L-alanine (Table 2).

The structural analyses described above strongly support a role of electrostatic interactions in the enhancement of conformational stability in the AGT-RHEAM. To provide a quantitative estimation of the electrostatic contribution to the high kinetic stability of HEM and RHEAM variants, we have used a simple Tanford–Kirwood algorithm described previously [40]. Comparison of the energy arising from charge–charge interactions ($E_{q\text{-}q}$) between the structures of HEM- and WT-AGT yield almost negligible net changes, whereas AGT-RHEAM shows a remarkable enhancement in the favourable electrostatic interactions (negative values $\Delta E_{q\text{-}q}$ in Supplementary Figure S4A at http://www.biochemj.org/bj/462/bj4620453add.htm). This stabilizing electrostatic contribution in RHEAM is due to the introduction of the Q23R and D52E mutations which seem to

**Table 1 Comparative analysis of AGT-WT and RHEAM variants thermal denaturation as holo- and apo-proteins determined by DSC**

Data were fitted using a two-state irreversible unfolding model including non-first-order kinetics as described in [26,28].

| Protein | $T_m$ (°C)* | $\Delta H$ (kcal/mol)† | $\mu$† | $E_a$ (kcal/mol)† | $k_{37°C}$ ($min^{-1}$)‡ | $k_{WT}/k_{RHEAM}$ (at 37°C) |
|---|---|---|---|---|---|---|
| Holo | | | | | | |
| WT | 82.1 | 274 ± 2 | 0.95 ± 0.02 | 109 ± 5 | $6.4 \times 10^{-10}$ | 1 |
| RHEAM | 93.7 | 319 ± 16 | 0.91 ± 0.04 | 125 ± 6 | $1.1 \times 10^{-13}$ | $2.1 \times 10^{3}$ |
| Apo | | | | | | |
| WT | 58.4 | 127 ± 8 | 1.30 ± 0.02 | 111 ± 10 | $1.6 \times 10^{-5}$ | 1 |
| RHEAM | 70.2 | 196 ± 10 | 0.89 ± 0.01 | 116 ± 7 | $3.6 \times 10^{-8}$ | $4.4 \times 10^{2}$ |

*From experiments at 3°C/min.
†From experiments at three different scan rates (means ± S.D.).
‡From Arrhenius plots.

**Table 2 Functional characterization of AGT-WT and RHEAM enzymes**

| Protein | $V_{max}$ (mmol·$h^{-1}$·$mg^{-1}$) | $K_{M(alanine)}$ (mM) | $K_{M(glyoxilate)}$ ($\mu$M) | $K_{d(PLP)}$ (nM) | $K_{d(Pex5p)}$ ($\mu$M) |
|---|---|---|---|---|---|
| WT | 2.22 ± 0.09 | 19.5 ± 1.4 | 245 ± 29 | 99 ± 20 | 1.4 ± 0.2 |
| RHEAM | 4.99 ± 0.22 | 11.5 ± 1.2 | 412 ± 45 | 101 ± 18 | 1.9 ± 0.4 |

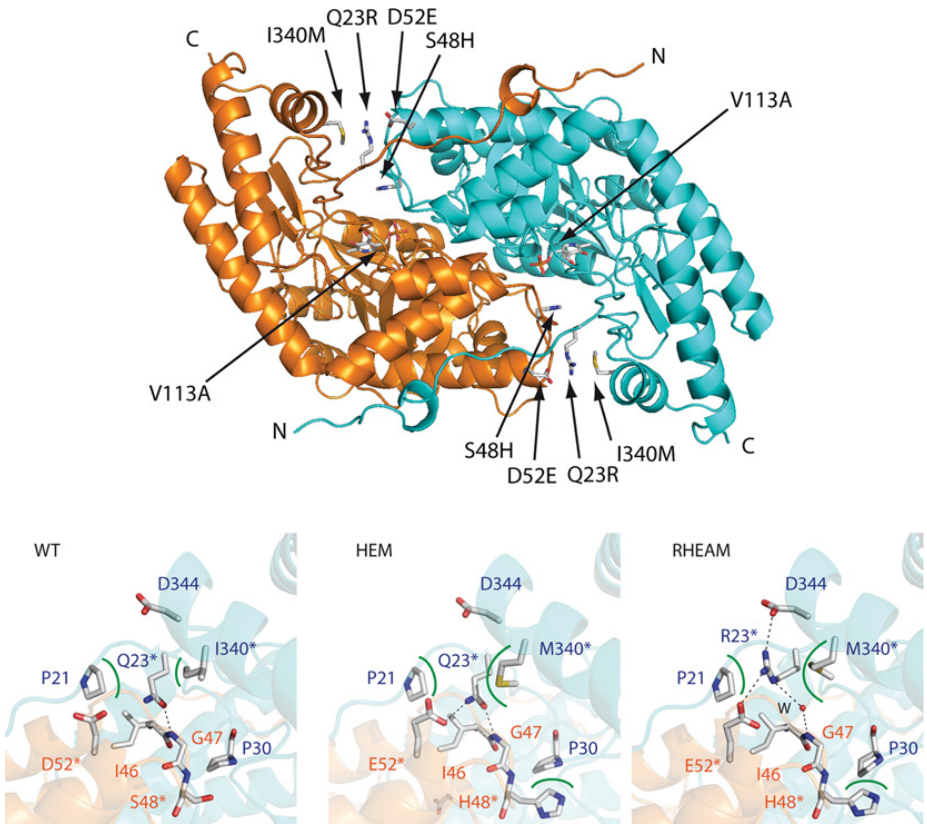


**Figure 4 Structural comparison between AGT-WT, HEM and RHEAM variants**

Upper panel: structural model for AGT-RHEAM dimer at 2.4 Å resolution. Individual mutations are labelled as well as the N- and C-termini. Lower panel, monomer–monomer interface in the vicinity of residues 23, 52 and 340 for AGT-WT, HEM and RHEAM variants. Hydrogen bonds are shown as dashed lines and hydrophobic interactions as green curved lines.

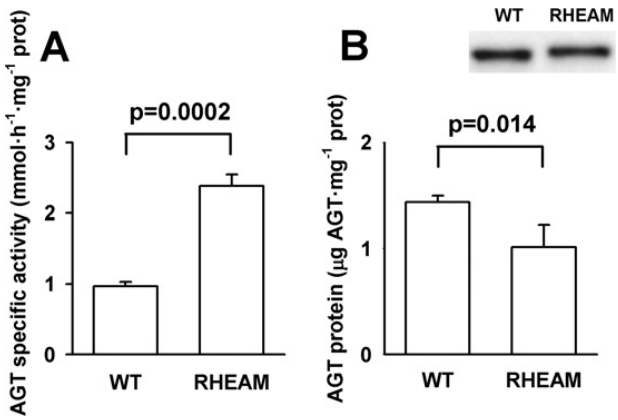


**Figure 5 AGT activity and immunoreactive protein levels in transiently transfected CHO cells**

(**A**) Specific AGT activity. Results are means ± S.D. for six independent experiments. (**B**) Immunoreactive AGT protein levels. Results are means ± S.D. for four independent experiments. Inset: representative Western blot analyses for AGT WT and RHEAM. Statistical differences are calculated using one-tailed unpaired Student's *t* tests.

develop new favourable charge–charge interactions with $Arg^{175}$, $Arg^{197}$, $Arg^{333}$ and $Asp^{344}$ (Supplementary Figure S4B). The overall increase in favourable electrostatic interactions amounts to several kcal/mol per AGT monomer (Supplementary Figure S4B), which is likely to contribute to the 4–5 kcal/mol increase in kinetic stability displayed by RHEAM compared with WT-AGT. These analyses also suggest that at least a significant fraction of the favourable electrostatic interactions formed in the native RHEAM protein are weakened or disrupted in the relevant transition state for the irreversible denaturation of AGT. Our data also show that the optimization of the kinetic stability occurs through the development of electrostatic networks between a few mutations (Q23R and D52E in RHEAM) and ionizable residues already present in the target protein ($Arg^{175}$, $Arg^{197}$, $Arg^{333}$ and $Asp^{344}$).

The effect of V113A is difficult to rationalize in structural terms. The side chain of $Val^{113}$ is partly exposed to the solvent, and its replacement by alanine reduces the amount of solvent-accessible hydrophobic surface by 40 %. Although this reduction is very small compared with the total solvent-accessible hydrophobic surface (less than 1 %), it might enhance the stability of the AGT dimer.

Our analysis shows that the stabilizing effect of consensus-based mutations on AGT are related to new favourable interactions in the native dimer, particularly of electrostatic and hydrophobic nature. Most of the new favourable interactions found in AGT-RHEAM occur at the monomer–monomer interface, a similar scenario to the one described for a tetrameric glucose dehydrogenase [43], in which a structure-guided consensus approach was found to enhance its kinetic stability by targeting the subunit–subunit interface.

### Expression of AGT-RHEAM in cell culture

To test whether the stability and activity of the AGT-RHEAM measured *in vitro* correlate with the same behaviour inside cells, as well the effects of the mutations on the foldability and intracellular transport into peroxisomes, we have transiently transfected CHO cells with plasmids carrying the cDNAs of WT and RHEAM AGTs. The specific activity (per mg of AGT) measured in AGT-RHEAM soluble extracts is 2.5 ± 0.3 fold higher than for WT-AGT (Figure 5A), which correlates excellently with the *in vitro* measurements. Regarding the expression levels, AGT-RHEAM is expressed at slightly lower levels than WT (72 ± 15 % of WT levels; Figure 5B). Moreover, the specific activities of both AGT enzymes compare reasonably well with those measured using recombinant enzyme (Table 2), thus supporting that most of the soluble AGT enzyme is forming active dimers. AGT-RHEAM exists mainly in the soluble fraction (~90 % of the total AGT protein) similar to the WT-AGT.

The intracellular targeting of the AGT-RHEAM has been also analysed in CHO cells by immunofluorescence confocal microscopy (Supplementary Figure S5 at http://www.biochemj.org/bj/462/bj4620453add.htm), and found at the correct peroxisomal location, as for WT-AGT. Thus the five amino acid changes introduced in the RHEAM sequence do not have negative consequences in terms of subcellular localization of the enzyme. This outcome is consistent with AGT-RHEAM properly folding to dimers and interacting with the Pex5p receptor with normal affinity (Table 2).

### The consensus approach improves the solubility and crystallization of AGT and CIPK24/SOS2 proteins

During the purification and crystallization trials of AGT HEM and RHEAM, we observed an increase in the ability of AGT consensus variants to crystallize and to increase the protein yield upon expression in *E. coli* cells (Figures 6A and 6B). These results suggest that thermostability enhancement of AGT provided by consensus mutations also increases their crystallization propensity and solubility. We thus propose that the consensus-based approach may be a simple strategy to increase the crystallization success and protein yields in challenging protein systems. To further support this notion, we have applied the consensus approach to the kinase domain of the plant CIPK24/SOS2 protein. WT CIPK24/SOS2 is a very unstable protein with very poor yields as a soluble protein upon expression in *E. coli* under several experimental conditions (Figure 6C and results not shown), therefore impeding its characterization by biophysical methods and X-ray crystallography. By applying the consensus approach to CIPK24/SOS2, we have chosen

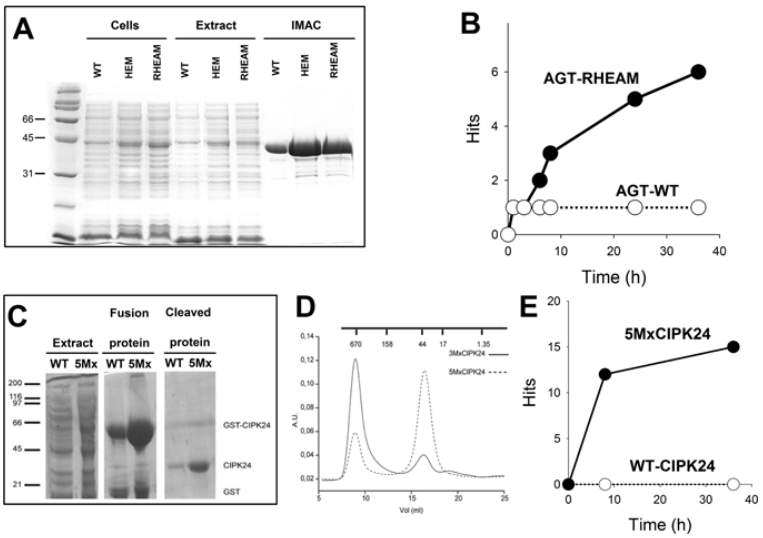


**Figure 6 Application of the consensus approach for crystallization studies**

(**A**) Increased levels of soluble protein by the consensus HEM and RHEAM variants compared with WT AGT upon expression in *E. coli* cultures. The cells were induced for 6 h at 4 °C. The amount of purified protein for the consensus variants is approximately 5-fold higher than for WT AGT, and all of them are approximately 90 % dimeric as judged by SEC (size-exclusion chromatography; results not shown). (**B**) The consensus RHEAM variant shows a higher crystallization propensity (number of hits) than the WT protein under a variety of conditions. (**C**) Increased levels of soluble protein for the consensus 5MxCIP24/SOS2 protein compared with the WT protein upon expression in *E. coli* cells at 16 °C. (**D**) SEC analyses shows that 5MxCIPK24 exists mainly as a monomer, whereas 3MxCIPK24 aggregates to a larger extent. (**E**) The consensus 5MxCIPK24/SOS2 variant shows a higher crystallization propensity (number of hits) than the WT protein under a variety of conditions. A.U., absorbance units.

five mutations (P81K, E107K, S109D, C127S and L266K) that should stabilize CIPK24/SOS2 (Supplementary Figure S6A at http://www.biochemj.org/bj/462/bj4620453add.htm). The resulting consensus multiple variants (3MxCIPK24 and 5MxCIPK24) significantly improved the foldability and solubility upon expression in *E. coli* cells (Figure 6C), increasing the yield as soluble protein several-fold and decreasing the tendency to aggregate along the purification process (notably 5MxCIPK24; Figure 6D), allowing its further characterization. The 5MxCIPK24 variant show a slightly higher $T_m$ value than that for 3MxCIPK24 (Supplementary Figure S6B), suggesting that the mutations P81K and L266K target the solubility of this protein rather than its thermostability. 5MxCIPK24 produced 3.5 Å resolution diffracting crystals (Supplementary Table S4), whereas no crystallization hits were found for WT CIPK24 (Figure 6E). These results suggest that consensus mutations may increase thermostability, solubility and crystallizability, even though the correlation between the improvement of these three properties may depend on the selected mutations and the protein target.

## DISCUSSION

In the present study we have successfully applied the consensus-based approach to protein stability to generate, through five single point mutations (AGT-RHEAM), an enhanced of version of human AGT, a protein of biomedical interest. This approach yielded a mutant enzyme with high kinetic stability (approximately 1000-fold longer half-life than the WT enzyme extrapolated to 37 °C) and with improved catalytic performance. The three dimensional structure of the AGT-RHEAM shows how the consensus based mutations target the stability of the native AGT dimer, which is the co-operative unit determining the kinetic stability of human AGT *in vitro* (i.e. denaturation follows first-order kinetics), partly by a remarkable optimization of electrostatic interactions in the protein native state. Thus the consensus-based approach can be efficiently applied to improve protein stability and solubility *in vitro* and possibly *in vivo*, which are known to strongly depend on protein thermodynamic and kinetic stabilities [44–46]. Our approach may thus have application to different fields of protein biotechnology aimed to reduce the degradation and aggregation rates of protein-based pharmaceutical drugs [47] and also, possibly, in preparations for ERT. One of the advantages of the consensus-based approach is that it does not require prior structural analysis of the target protein or exhaustive experimental/computational work, in contrast with other protein engineering technologies, such as structure-based protein engineering [17,20,48,49], ancestral resurrection approaches [12,20,50] and biophysical screenings in cells [51].

The production of large amounts of high-quality protein samples still is the bottleneck for many biophysical and structural studies. Although the consensus-based approach for stabilizing proteins is largely employed in biotechnology, to our knowledge there are no reports on the ability of these mutant proteins to improve recombinant protein production for structural characterization. Importantly, the consensus-based variants maintain the protein fold virtually unchanged, improving the success of the crystallization process and increasing protein solubility (as seen in the present study for the AGT and CIPK24/SOS2 proteins). Protein homogeneity and solubility are linked to successful crystallization which can be improved by removing unstructured regions or by the introduction of localized surface entropy-reducing mutations [52,53] and protein stabilization by chemical ligands and solvent conditions [22,23]. Accordingly, the consensus-based approach seems to be a simple

and generic strategy to identify those mutations that increase the chance of crystallization in hard-to-get protein targets and, thus, it may be complementary to current approaches based on protein stabilization to increase the success in obtaining high-quality protein crystals for high-resolution structural determinations [22,23,51].

Another very attractive application for the consensus-based approach in proteins of biomedical interest (such as human AGT) would be to develop more efficient systems for GT and ERT. Specific amino acid substitutions that result in enhanced activity of coagulation Factor VIIa [2] have been introduced in gene replacement vectors in order to improve the therapeutic effect obtained with the lowest viral dose possible in mouse models of haemophilia [1]. An increase in protein stability would allow reduction of the dose and/or the frequency of vector/protein administration. However, we must note that the enhanced *in vitro* stability of AGT-RHEAM does not translate into higher resistance towards degradation or increased steady-state levels or solubility inside eukaryotic cells (as seen by the similar expression levels of AGT-WT and RHEAM; Figure 5). One possible explanation for this is that the kinetically relevant state *in vitro* (leading to protein aggregation upon thermal denaturation) might differ from the state sensitive to cytosolic and/or peroxisomal degradation by the corresponding protein quality control systems. Degradation of AGT in the cytosol may likely occur through a proteasome ATP-dependent mechanism [54]. However, the degradation pathway of AGT in peroxisomes is unknown, but possibly involves a Lon protease that selectively degrades oxidatively damaged proteins [55]. Alternatively, it is plausible that evolution has provided the human AGT WT enzyme with an adequate balance between conformational stability and turnover to perform its function efficiently in a relatively harsh scenario such as the peroxisomes (with a high production of reactive oxygen species). Nevertheless, it is plausible that increasing the conformational stability will lead to higher transduction levels upon gene transfer and/or longer half-lives of injectable ERT preparations in the bloodstream, thus benefiting from the use of the consensus-based approach to enhance protein stability and activity.

In summary, we have shown in the present study a simple approach to improve protein kinetic stability and catalytic performance in proteins of biomedical interest, with potential applications in GT and ERT for a wide range of inborn errors of metabolism. The present study clearly shows that the consensus-based approach optimizes interactions in the AGT native structure that efficiently increase the kinetic free energy barrier for protein denaturation. Since the introduction of consensus mutations maintain the overall protein fold and enzyme function while increasing protein stability and/or solubility and speeding up the crystallization process in two unrelated proteins (AGT and SOS2), we also propose that this approach could be used as a simple and generic strategy to obtain high resolution structural models of proteins for which crystallization is a challenging task due to protein stability issues.

## AUTHOR CONTRIBUTION

Noel Mesa-Torres expressed and purified most of the AGT proteins, performed and analysed the activity, spectroscopic and calorimetric experiments and electrostatic calculations. Cristina Yunta crystallized the AGT proteins and determined their three-dimensional structures. Israel Fabelo-Rosa expressed some AGT proteins, performed some activity and calorimetric experiments and contributed to the AGT expression in eukaryotic cells. Juana María Gonzalez-Rubio performed cloning, site-directed mutagenesis, and expression and crystallization of the CIPK24/SOS2 proteins. José Manuel Sánchez-Ruiz contributed with analytical tools for the consensus- and structure-based electrostatic analyses. Eduardo Salido performed site-directed mutagenesis on the AGT enzymes, and performed and analysed expression studies in eukaryotic cells. Armando Albert performed consensus-based analyses on the CIPK24/SOS2 proteins, supervised and contributed to structural determination and analyses on AGT and CIPK24/SOS2 proteins. Angel L. Pey expressed and purified some AGT and Pex5p proteins, performed calorimetric studies, analysed data and supervised the *in vitro* experiments on AGT proteins. Eduardo Salido, Armando Albert and Angel L. Pey designed the research and wrote the paper. All authors discussed the results and approved the final version of the paper.

## ACKNOWLEDGEMENT

We thank Dr Beatriz Ibarra-Molero for her help and advice with electrostatic calculations.

## FUNDING

This work was supported by the MINECO (Ministerio de Economía y Competitividad) [grant numbers BIO2012-34937 and CSD-2009-00088 (to J.M.S.-R.), BFU2011-25384 and CSD2006-00015 (to A.A.) and SAF2011-23933 (to E.S.)], the Comunidad de Madrid [grant number S2010/BMD-2457 (to A.A.)], the Junta de Andalucia [grant number P11-CTS-07187 (to A.L.P)], the European Union [grant number FP7-REGPOT-CT2012-31637-IMBRAIN (to E.S.)] and FEDER (Fondo Europeo de Desarrollo Regional) (to J.M.S-R.). A.L.P. is supported by a Ramón y Cajal research contract from MINECO [grant number RYC-2009-04147]. N.M.-T. and I.F.-R are supported by FPI pre-doctoral fellowships from MINECO.

## REFERENCES

1 Margaritis, P., Roy, E., Faella, A., Downey, H. D., Ivanciu, L., Pavani, G., Zhou, S., Bunte, R. M. and High, K. A. (2011) Catalytic domain modification and viral gene delivery of activated factor VII confers hemostasis at reduced expression levels and vector doses *in vivo*. Blood **117**, 3974–3982 CrossRef PubMed
2 Persson, E., Kjalke, M. and Olsen, O. H. (2001) Rational design of coagulation factor VIIa variants with substantially increased intrinsic activity. Proc. Natl. Acad. Sci. U.S.A. **98**, 13583–13588 CrossRef PubMed
3 Cochat, P. and Rumsby, G. (2013) Primary hyperoxaluria. N. Engl. J. Med. **369**, 649–658 CrossRef PubMed
4 Salido, E., Pey, A. L., Rodriguez, R. and Lorenzo, V. (2012) Primary hyperoxalurias: disorders of glyoxylate detoxification. Biochim. Biophys. Acta **1822**, 1453–1464 CrossRef PubMed
5 Pey, A. L., Albert, A. and Salido, E. (2013) Protein homeostasis defects of alanine-glyoxylate aminotransferase: new therapeutic strategies in primary hyperoxaluria type I. Biomed. Res. Int. **2013**, 687658 CrossRef PubMed
6 Bobrowski, A. E. and Langman, C. B. (2006) Hyperoxaluria and systemic oxalosis: current therapy and future directions. Expert Opin. Pharmacother. **7**, 1887–1896 CrossRef PubMed
7 Roncador, A., Oppici, E., Montioli, R., Maset, F. and Cellini, B. (2013) TAT-mediated delivery of human alanine:glyoxylate aminotransferase in a cellular model of primary hyperoxaluria type I. Int. J. Pep. Res. Ther. **19**, 175–184 CrossRef
8 Salido, E., Rodriguez-Pena, M., Santana, A., Beattie, S. G., Petry, H. and Torres, A. (2011) Phenotypic correction of a mouse model for primary hyperoxaluria with adeno-associated virus gene transfer. Mol. Ther. **19**, 870–875 CrossRef PubMed
9 Steipe, B. (2004) Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes. Methods Enzymol. **388**, 176–186 CrossRef PubMed
10 Bershtein, S., Goldin, K. and Tawfik, D. S. (2008) Intense neutral drifts yield robust and evolvable consensus proteins. J. Mol. Biol. **379**, 1029–1044 CrossRef PubMed
11 Akanuma, S., Nakajima, Y., Yokobori, S., Kimura, M., Nemoto, N., Mase, T., Miyazono, K., Tanokura, M. and Yamagishi, A. (2013) Experimental evidence for the thermophilicity of ancestral life. Proc. Natl. Acad. Sci. U.S.A. **110**, 11067–11072 CrossRef PubMed
12 Risso, V. A., Gavira, J. A., Mejia-Carmona, D. F., Gaucher, E. A. and Sanchez-Ruiz, J. M. (2013) Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. J. Am. Chem. Soc. **135**, 2899–2902 CrossRef PubMed
13 Risso, V. A., Gavira, J. A., Gaucher, E. A. and Sanchez-Ruiz, J. M. (2014) Phenotypic comparisons of consensus variants versus laboratory resurrections of Precambrian proteins. Proteins **82**, 887–896 CrossRef PubMed
14 Wang, Q., Buckle, A. M., Foster, N. W., Johnson, C. M. and Fersht, A. R. (1999) Design of highly stable functional GroEL minichaperones. Protein Sci. **8**, 2186–2193 CrossRef PubMed
15 Nikolova, P. V., Henckel, J., Lane, D. P. and Fersht, A. R. (1998) Semirational design of active tumor suppressor p53 DNA binding domain with enhanced stability. Proc. Natl. Acad. Sci. U.S.A. **95**, 14675–14680 CrossRef PubMed
16 Rath, A. and Davidson, A. R. (2000) The design of a hyperstable mutant of the Abp1p SH3 domain by sequence alignment analysis. Protein Sci. **9**, 2457–2469 CrossRef PubMed

17 Pey, A. L., Rodriguez-Larrea, D., Bomke, S., Dammers, S., Godoy-Ruiz, R., Garcia-Mira, M. M. and Sanchez-Ruiz, J. M. (2008) Engineering proteins with tunable thermodynamic and kinetic stabilities. Proteins **71**, 165–174 CrossRef PubMed
18 Rodriguez-Larrea, D., Perez-Jimenez, R., Sanchez-Romero, I., Delgado-Delgado, A., Fernandez, J. M. and Sanchez-Ruiz, J. M. (2010) Role of conservative mutations in protein multi-property adaptation. Biochem. J. **429**, 243–249 CrossRef PubMed
19 Lehmann, M., Pasamontes, L., Lassen, S. F. and Wyss, M. (2000) The consensus concept for thermostability engineering of proteins. Biochim. Biophys. Acta **1543**, 408–415 CrossRef PubMed
20 Wijma, H. J., Floor, R. J. and Janssen, D. B. (2013) Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. Curr. Opin. Struct. Biol. **23**, 588–594 CrossRef PubMed
21 Sanchez-Barrena, M. J., Martinez-Ripoll, M. and Albert, A. (2013) Structural biology of a major signaling network that regulates plant abiotic stress: the CBL-CIPK mediated pathway. Int. J. Mol. Sci. **14**, 5734–5749 CrossRef PubMed
22 Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal. Biochem. **357**, 289–298 CrossRef PubMed
23 Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, Jr, P. J., Wasney, G. A., Yeung, R., Arrowsmith, C., Ball, L. J., Berglund, H. et al. (2006) Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. Proc. Natl. Acad. Sci. U.S.A. **103**, 15835–15840 CrossRef PubMed
24 Mesa-Torres, N., Fabelo-Rosa, I., Riverol, D., Yunta, C., Albert, A., Salido, E. and Pey, A. L. (2013) The role of protein denaturation energetics and molecular chaperones in the aggregation and mistargeting of mutants causing primary hyperoxaluria type I. PLoS ONE **8**, e71963 CrossRef PubMed
25 Cellini, B., Bertoldi, M., Montioli, R., Paiardini, A. and Borri Voltattorni, C. (2007) Human wild-type alanine:glyoxylate aminotransferase and its naturally occurring G82E variant: functional properties and physiological implications. Biochem. J. **408**, 39–50 CrossRef PubMed
26 Pey, A. L., Salido, E. and Sanchez-Ruiz, J. M. (2011) Role of low native state kinetic stability and interaction of partially unfolded states with molecular chaperones in the mitochondrial protein mistargeting associated with primary hyperoxaluria. Amino Acids **41**, 1233–1245 CrossRef PubMed
27 Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci. **4**, 2411–2423 CrossRef PubMed
28 Sanchez-Ruiz, J. M. (1992) Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. Biophys. J. **61**, 921–935 CrossRef PubMed
29 Sanchez-Ruiz, J. M., Lopez-Lacomba, J. L., Cortijo, M. and Mateo, P. L. (1988) Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin. Biochemistry **27**, 1648–1652 CrossRef PubMed
30 Rumsby, G., Weir, T. and Samuell, C. T. (1997) A semiautomated alanine:glyoxylate aminotransferase assay for the tissue diagnosis of primary hyperoxaluria type 1. Ann. Clin. Biochem. **34**, 400–404 CrossRef PubMed
31 Kabsch, W. (2010) Xds. Acta Crystallogr. D Biol. Crystallogr. **66**, 125–132 CrossRef PubMed
32 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. and Read, R. J. (2007) Phaser crystallographic software. J. Appl. Crystallogr. **40**, 658–674 CrossRef PubMed
33 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. **66**, 213–221 CrossRef PubMed
34 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60**, 2126–2132 CrossRef PubMed
35 Evans, P. (2006) Scaling and assessment of data quality. Acta Crystallogr. D Biol. Crystallogr. **62**, 72–82 CrossRef PubMed
36 The CCP4 suite: programs for protein crystallography (1994) Acta Crystallogr. D Biol. Crystallogr. **50**, 760–763 CrossRef PubMed
37 Chen, V. B., Arendall, III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. **66**, 12–21 CrossRef PubMed
38 Reference deleted
39 Tanford, C. and Kirkwood, J. G. (1957) Theory of protein titration curves. I. General equations for impenetrable spheres. J. Am. Chem. Soc. **79**, 5333–5339 CrossRef
40 Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I. and Sanchez-Ruiz, J. M. (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge–charge interactions to protein stability. Biochemistry **38**, 8138–8149 CrossRef PubMed
41 Zhang, X., Roe, S. M., Hou, Y., Bartlam, M., Rao, Z., Pearl, L. H. and Danpure, C. J. (2003) Crystal structure of alanine:glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1. J. Mol. Biol. **331**, 643–652 CrossRef PubMed
42 Wallace, A. C., Laskowski, R. A. and Thornton, J. M. (1995) LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. Protein Eng. **8**, 127–134 CrossRef PubMed
43 Vazquez-Figueroa, E., Chaparro-Riggers, J. and Bommarius, A. S. (2007) Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. ChemBioChem **8**, 2295–2301 CrossRef PubMed
44 Niwa, T., Ying, B. W., Saito, K., Jin, W., Takada, S., Ueda, T. and Taguchi, H. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. Proc. Natl. Acad. Sci. U.S.A. **106**, 4201–4206 CrossRef PubMed
45 Bai, S., Manning, M. C., Randolph, T. W. and Carpenter, J. F. (2011) Aggregation of recombinant human botulinum protein antigen serotype C in varying solution conditions: implications of conformational stability for aggregation kinetics. J. Pharm. Sci. **100**, 836–848 CrossRef PubMed
46 Agostini, F., Vendruscolo, M. and Tartaglia, G. G. (2012) Sequence-based prediction of protein solubility. J. Mol. Biol. **421**, 237–241 CrossRef PubMed
47 Jiskoot, W., Randolph, T. W., Volkin, D. B., Middaugh, C. R., Schoneich, C., Winter, G., Friess, W., Crommelin, D. J. and Carpenter, J. F. (2012) Protein instability and immunogenicity: roadblocks to clinical application of injectable protein delivery systems for sustained release. J. Pharm. Sci. **101**, 946–954 CrossRef PubMed
48 Schweiker, K. L. and Makhatadze, G. I. (2009) Protein stabilization by the rational design of surface charge–charge interactions. Methods Mol. Biol. **490**, 261–283 CrossRef PubMed
49 Sanchez-Romero, I., Ariza, A., Wilson, K. S., Skjot, M., Vind, J., De Maria, L., Skov, L. K. and Sanchez-Ruiz, J. M. (2013) Mechanism of protein kinetic stabilization by engineered disulfide crosslinks. PLoS ONE **8**, e70013 CrossRef PubMed
50 Perez-Jimenez, R., Ingles-Prieto, A., Zhao, Z. M., Sanchez-Romero, I., Alegre-Cebollada, J., Kosuri, P., Garcia-Manyes, S., Kappock, T. J., Tanokura, M., Holmgren, A. et al. (2011) Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nat. Struct. Mol. Biol. **18**, 592–596 CrossRef PubMed
51 Asial, I., Cheng, Y. X., Engman, H., Dollhopf, M., Wu, B., Nordlund, P. and Cornvik, T. (2013) Engineering protein thermostability using a generic activity-independent biophysical screen inside the cell. Nat. Commun. **4**, 2901 CrossRef PubMed
52 Derewenda, Z. S. and Vekilov, P. G. (2006) Entropy and surface engineering in protein crystallization. Acta Crystallogr. D Biol. Crystallogr. **62**, 116–124 CrossRef PubMed
53 Ng, L. M., Soon, F. F., Zhou, X. E., West, G. M., Kovach, A., Suino-Powell, K. M., Chalmers, M. J., Li, J., Yong, E. L., Zhu, J. K. et al. (2011) Structural basis for basal activity and autoactivation of abscisic acid (ABA) signaling SnRK2 kinases. Proc. Natl. Acad. Sci. U.S.A. **108**, 21259–21264 CrossRef PubMed
54 Coulter-Mackie, M. B. and Lian, Q. (2006) Consequences of missense mutations for dimerization and turnover of alanine:glyoxylate aminotransferase: study of a spectrum of mutations. Mol. Genet. Metab. **89**, 349–359 CrossRef PubMed
55 Bartoszewska, M., Williams, C., Kikhney, A., Opalinski, L., van Roermund, C. W., de Boer, R., Veenhuis, M. and van der Klei, I. J. (2012) Peroxisomal proteostasis involves a Lon family protein that functions as protease and chaperone. J. Biol. Chem. **287**, 27380–27395 CrossRef PubMed


Received 25 February 2014/4 June 2014; accepted 24 June 2014
Published as BJ Immediate Publication 24 June 2014, doi:10.1042/BJ20140250

# APPENDIX C

# The lower limits for protein stability and foldability in primary hyperoxaluria type I.

**Noel Mesa-Torres[1], Eduardo Salido[2] and Angel L. Pey[1*].**

[1] Departamento de Química-Física, Facultad de Ciencias, Universidad de Granada, E-18071, Granada (Spain).
[2] Hospital Universitario de Canarias, Universidad La Laguna, and Centre for Biomedical Research on Rare Diseases (CIBERER), Tenerife E-38320, (Spain).
*To whom correspondence should be addressed: Correspondence to: Angel L. Pey, Departamento de Química-Física, Facultad de Ciencias, Universidad de Granada, Av. Fuentenueva s/n, Granada 18071, Spain. Tel: +34-958240436; Fax: +34-958242747; E-mail: angelpey@ugr.es.

**Keywords**.- Protein kinetic stability; Protein denaturation energetics; Protein foldability; misfolding disease; differential scanning calorimetry; molecular chaperones.

**Abbreviations**.- AGT.- alanine:glyoxylate aminotransferase; PH1.- primary hyperoxaluria type I; PLP.- pyridoxal 5´-phosphate; $T_m$.- maximum of the thermal denaturation transition; $E_a$.- activation energy; $\Delta H$.- denaturation enthalpy; ΔASA, change in accessible surface area upon denaturation.

## Abstract

Mutational effects on protein stability and foldability are important to understand conformational diseases and protein evolution. In this work, we perform a comprehensive investigation on the energetic basis underlying mutational effects on the stability of human alanine:glyoxylate aminotransferase (AGT) and their foldability upon expression in eukaryotic cells. We study twenty two variants whose kinetic stabilities span over eleven orders of magnitude and classified in two groups: i) ten naturally-occurring variants, including the most common mutations causing primary hyperoxaluria type I (PH1); ii) twelve consensus variants obtained from sequence-alignment statistics. We show that AGT dimer stability determines denaturation rates, and mutations modulate stability by changes in the effective thermodynamic stability, the aggregation propensity of partially/globally unfolded states and subtle energetic changes in the rate-limiting denaturation step. In combination with expression analyses in eukaryotic cells, we propose the existence of two lower limits for AGT stability, one linked to optimal folding efficiency (close to the *major allele* stability) and the other setting a minimal efficiency compatible with glyoxylate detoxification in vivo (close to *the minor allele* stability). These lower limits could explain the high prevalence of misfolding as a disease mechanism in PH1 and support the use of pharmacological ligands aimed to increase AGT stability as therapies for this disease.

## 1. Introduction

The intracellular levels of folded and active proteins are determined to a large extent by their conformational (kinetic and thermodynamic) stability and the interaction of folded and partially folded states with folding, degradation and intracellular trafficking machineries (e.g. protein homeostasis networks; [1-3]).

Protein stability and homeostasis networks have been adapted along molecular evolution to provide an adequate balance between stability, flexibility, functionality and evolvability [3-6]. Protein stability has been particularly important to develop new protein functions and substrate/cofactor specificities, and for this, molecular chaperones may have also played a dominant role in promoting functional diversity and evolvability [7,8]. In many human misfolding diseases, mutations may often reduce protein thermodynamic and kinetic stabilities, leading to enhanced protein degradation, aggregation and, in some cases, aberrant intracellular trafficking [9-12]. However, comprehensive studies on the relationship between mutation induced destabilization and protein foldability in eukaryotic cells are scarce.

Primary hyperoxaluria type I (PH1) is a human metabolic disease inherited in a autosomal recessive pattern and caused by mutations in the hepatic and peroxisomal alanine:glyoxylate aminotransferase (AGT) enzyme. AGT catalyzes the transamination of L-ala to pyruvate and of glyoxylate to glycine using pyridoxal 5-phosphate (PLP) as coenzyme, and its main role is to detoxify glyoxylate in humans. A defective AGT causes accumulation of glyoxylate, which oxidizes to oxalate and eventually precipitates. This leads to renal failure, widespread oxalate precipitation and oxalosis, which is a life-threatening condition unless liver and kidney transplation are successful [13]. Mutations generally affect the ability of AGT to fold properly, either leading to peroxisomal aggregation, mitochondrial mistargeting or accelerated protein turnover [11,14-20].

Human AGT forms active dimers, with the two monomers being necessary to form the two active sites, and each active site contains a PLP molecule forming a Schiff base with Lys209. The AGT protein exists as two polymorphic variants, named the *major* allele (or WT enzyme) and the *minor* allele (named in this work as AGT LM), which contains two amino acid substitutions (P11L and I340M)[13]. Despite the minor allele does not cause PH1 by itself, it is known to sensitize AGT protein towards additional deleterious mutations [16,17,19,20, 21] and it is found with higher frequency in PH1 patients that in control subjects (50% vs 20%, respectively; [22]). More than 150 mutations have been described in PH1 patients, and they affect residues distributed all over the AGT structure [22]. The molecular basis of mitochondrial mistargeting and protein aggregation in PH1 are not fully understood, but disease-causing variants show a decrease in their resistance towards thermal and chemical denaturation, and importantly, reduced native state kinetic stability, features which may play a role in these two pathogenic mechanisms [11, 13, 16, 19]. Importantly, a low protein stability seems to roughly correlate with enhanced interaction of partially unfolded states molecular chaperones [14, 16, 19], which may represent a checkpoint along folding of AGT and partitioning between folding into dimers and peroxisomal import, aggregation and mistargeting to mitochondria [11].

The *consensus approach* is a simple strategy to improve protein stability that uses statistical information contained in protein sequence alignments ([23, 24] and references therein). Originally, this stabilizing effect was explained by the *pseudo-equilibrium* hypothesis, which implies that in an evolutionary time-scale, neutral mutations are fixed in protein sequences with probabilities scaling with their contribution to protein stability ([25] and references therein). More recently, it has been proposed that consensus mutations reflect to some extent the thermophilicity of ancestral proteins [26,27]. Indeed, kinetic and thermodynamic studies on *consensus*

mutations have suggested that evolution may have optimized (at least in some cases) kinetic rather than thermodynamic stability [25]. We have recently shown that *consensus* mutations significantly enhance the kinetic stability of human AGT, which might be useful to improve current gene- and enzyme-replacement therapies for PH1 [24].

In this work, we have performed a comprehensive characterization of the conformational stability of twenty two AGT variants in their holo-form (saturated with PLP) by a combination of thermal and chemical denaturation studies, including ten naturally-occurring variants and associated to PH1 [19]) and twelve consensus variants that often enhance AGT kinetic stability [24]. Our results suggest a delicate balance between protein energetics, thermodynamic stability and aggregation propensities determining the wide range of kinetic stability observed (spanning from hours to millions of years). The correlation between AGT kinetic stability *in vitro* and *foldability* in eukaryotic cells for twelve different variants suggests the existence of lower limits in AGT stability and foldability in health and disease.

## 2. Methods

### 2.1 Protein expression and purification.

Site-directed mutagenesis, expression and purification of holo-AGT dimers has been recently described [19,24]. Proteins were kept at -80$^{o}$C until use, and protein concentration was measured spectrophotometrically using a $\varepsilon_{280nm}$=1.069 $mg^{-1}\cdot ml\cdot cm^{-1}$.

### 2.2 Differential scanning calorimetry (DSC).

Preparation of protein samples in the absence of urea, analyses and fitting procedures based on a two-state irreversible denaturation model (N→F, where N and F stand for the native and the irreversible denatured states, respectively) have been recently described in detail [16,19]. For calorimetric analysis in the presence of urea, urea solutions were daily prepared and concentrations were determined by refractive index measurements.

### 2.3 Structure-energetic relationships and calculation of activation energetics parameters.

Mutational effects on the activation free energy ($\Delta\Delta G^{\neq}$), enthalpy ($\Delta\Delta H^{\neq}$) and entropy ($\Delta\Delta S^{\neq}$) where determined on the basis of the transition state theory as described in [28,29]. The values of $\Delta\Delta H^{\neq}$ and $\Delta\Delta S^{\neq}$ where considered to be constant owing to the highly linear Arrhenius plots found in all cases. However, the temperature dependence of these two parameters are shown to modestly affect the outcome of the following calculations (see Figure 4 for two widely different temperatures). The effect of mutations on activation energetic parameters were determined using the following expressions (equations 1-3):

$$\Delta\Delta G^{\neq} = -R\cdot T\cdot \ln\left(\frac{k_{(37^{o}C)}(mut)}{k_{(37^{o}C)}(WT)}\right) \qquad \text{(Equation 1)}$$

$$\Delta\Delta H^{\neq} = E_a(mut) - E_a(WT) \quad \text{(Equation 2)}$$

$$-T\Delta\Delta S^{\neq} = \Delta\Delta G^{\neq} - \Delta\Delta H^{\neq} \quad \text{(Equation 3)}$$

Kinetic $m$ values ($m^{\ddagger}$) were determined from the urea concentration dependence of the $T_m$ and $E_a$ values for each variant according to [28], using the following expression:

$$m^{\neq} = -\frac{E_a}{T_m}\left(\frac{dT_m}{d[urea]}\right) - RT_m\left(\frac{d\ln\left(\frac{E_a}{RT_m^2}\right)}{d[urea]}\right) \quad \text{(Equation 4)}$$

For each experimental scan rate, $T_m$ and $E_a$ values in the absence of urea were used, as well as the corresponding $dT_m/d[urea]$ and $d\ln(E_a/RT_m^2)/d[urea]$ values obtained from their dependence on urea concentrations. The reported $m^{\ddagger}$ are the means±s.d. from three experimental scan rates.

The contributions from *unfolding* ($\Delta H_{UNF}^{\ddagger}$) and *solvation barriers* ($\Delta H^*$) to activation energies (≈activation enthalpies) were determined following the procedures described in [28,30]. Briefly, $\Delta H_{UNF}^{\ddagger}$ and $\Delta H^*$ were determined using the following expressions:

$$\Delta H_{UNF}^{\neq} = \Delta H \cdot \frac{m^{\neq}}{m_{eq}} \quad \text{(Equation 5)}$$

$$\Delta H^* = E_a - \Delta H_{UNF}^{\ddagger} \quad \text{(Equation 6)}$$

The contributions of $\Delta H_{UNF}^{\ddagger}$ and $\Delta H^*$ has been determined for each variant using two different methods to account for uncertainties inherent to the structure-energetic parametrizations. We have thus used either experimental $\Delta H$ for each variant or the theoretical $\Delta H$ values calculated at the $T_m$ from the parameterization of Robertson and Murphy [31] based on changes in total change in accessible surface are (ΔASA) from the crystal structure of WT AGT (PDB:1H0C) and model tripeptide Gly-X-Gly for the unfolded state (see [30] for further details). The structural changes associated to the transition between the native and transition states are calculated from the average values for $\Delta H_{UNF}^{\ddagger}$, and $\Delta H^*$ and their relation with $\Delta ASA_{total}$ previously described [28,29].

**2.4 Urea induced unfolding.**

Concentrated urea stocks solutions were daily prepared in 20 mM HEPES, 200 mM NaCl, pH 7.4, and urea concentrations were measured by refractive index. Protein unfolding was monitored by measuring ellipticity at 222 nm in a Jasco J-710 spectropolarimeter using 1-mm path-length cuvettes thermostatized with a Peltier element. Urea solutions (7-9 M) containing a final concentration of 1 mM Tris(2-carboxyethyl)phosphine (TCEP) were incubated at the target temperature (25-70$^{o}$C) for 10 min in a 1 mm path-length cuvette and then, concentrated protein solutions (customarily 50 μM in AGT monomer with a 10-fold excess of PLP) were diluted 1:10, manually mixed and the time dependence of ellipticity was registered at 222

nm. The dead time of the experiments was 15-20 s. These kinetic traces were always excellently described by a single exponential function that yields an apparent rate constant *k*. The dependence of *k* on urea concentration was found to be linear and analyzed using the following expression:

$$\ln k = \ln k_{unf(0M)} + \frac{m^{\neq}}{RT} \cdot [Urea] \qquad \text{(Equation 7)}$$

Where $\ln k_{unf(0M)}$ is the natural logarithm of the rate constant extrapolated to the absence of urea and $m^{\neq}$ is the unfolding kinetic *m* value. The errors reported are fitting errors.

The temperature dependence was analyzed using the Arrhenius equation:

$$k = A \cdot \exp\left[-\frac{E_a}{R \cdot T}\right] \qquad \text{(Equation 8)}$$

Where $E_a$ is the activation energy, *R* is the ideal gas constant and *T* is the absolute temperature in K.

**2.5 Sequence based prediction of aggregation propensity.**

The aggregation propensity of human AGT enzymes was evaluated using the Zyggregator algorithm (http://www-vendruscolo.ch.cam.ac.uk/zyggregator-all.php; [32,33]). The input for these analyses was the primary sequence of human AGT enzymes and calculations were performed at neutral pH. The Zyggregator algorithm provides two different scores for propensities towards protein aggregation ($Z_{agg}$, for amyloid formation, and, $Z_{tox}$, for non-native oligomer formation). The algorithm provides aggregation propensities at the residue level as well as the overall aggregation propensity of the entire sequence, in all cases considering an unfolded state as reference (thus neglecting the contribution from the stability in the native state).

**3. Results**

**3.1 Naturally-occurring and consensus AGT variants display widely different kinetic stabilities but share a common denaturation mechanism**

We have recently reported thermal denaturation and kinetic stability analyses on twenty-two variants of human AGT by differential scanning calorimetry (DSC, [19,24]). This set of variants includes: i) ten *naturally-occurring* variants of AGT, five of them are polymorphic variants: the major allele (WT), the minor allele (LM), the two polymorphisms forming the minor allele individually (P11L and I340M), and the R197Q change on the minor allele (a polymorphic variant recently found in PH1 patients; [19]), and the remaining five variants known to cause PH1 (H83R, G170R, I244T, A295T and A368T, all on the minor allele). All these variants are correctly targeted to peroxisomes, with the exception of F152I and G170R on the minor alleles, which are mistargeted to mitochondria, and often decrease the yield in functional dimers when expressed in eukaryotic cells [19]; ii) twelve *consensus* variants previously generated from sequence alignment statistics, which were shown to increase in most of the cases AGT protein kinetic stability, including five single point mutants (Q23R, S48H, S50H, D52E, V113A), three double mutants (HE, S48H/D52E; HM, S48H/I340M; EM, D52E/I340M), a triple

mutant (HEM, S48H/D52E/I340M), two four-fold mutants (RHEM, Q23R/S48H/D52E/I340M and HEAM, S48H/D52E/V113A/I340M) and a quintuple mutant (RHEAM, Q23R/S48H/D52E/V113A/I340M) [24]. The consensus variant RHEAM is expressed at WT-like levels as functional dimers in eukaryotic cells and normally targeted to peroxisomes [24], suggesting that these consensus mutations do not alter intracellular traffic of AGT. As we show in Figure 1, the location of the residues mutated in these twenty two variants scatter over the structure of dimeric AGT, although some consensus variants reported to stabilize AGT (Q23R, D52E and I340M) generate a new cluster of stabilizing interactions at the dimer interface [24].

Figure 2 shows representative DSC analyses for three AGT variants: WT, P11L (a polymorphic variant with lower thermal stability) and the triple mutant HEM (a consensus variant with enhanced thermal stability). All the AGT variants studied are described well by a simple two-state irreversible denaturation model with first-order kinetics, as supported by: i) good fittings to the experimental calorimetric traces (Figure 2); ii) consistent values of the rate constants obtained at different scan rates (see Arrhenius plots in Fig S1); iii) very similar activation energy ($E_a$) values obtained by applying the consistency tests proposed by [34](see Figure S1 and Table 1); iv) first-order kinetics is supported by two independent tests: the $T_m$ values are independent of protein concentration (see Figure S2), and, fittings to a two-state denaturation model including the reaction order ($1/\mu$) as a fitting parameter showed negligible deviations from first-order kinetics (experimental $\mu$ values in the range ~0.9-1.0, see Table S1; Note that $\mu$=1 implies first-order kinetics).

As shown in Table 1, the AGT variants studied cover a wide range of $T_m$ values as well as kinetic stabilities extrapolated to 37°C [19,24]. The $T_m$ values range from 58-92°C (at 3°C/min), therefore displaying $T_m$ values which are covering a range of ~34°C. The extrapolated values for $k$ at 37°C range from $2.9 \cdot 10^{-3}$ $min^{-1}$ to $9.5 \cdot 10^{-14}$ $min^{-1}$, which yield denaturation half-lives at this temperature from ~4 h to ~14 million years, thus spanning over eleven orders of magnitude. Despite the inherent uncertainty associated to the long extrapolations to 37°C (mostly due to the estimation of activation energies), the excellent agreement between $T_m$ values and kinetic stabilites at 37°C (Figure S3) support that overall these results are robust, and thus, that the differences in kinetic stabilities observed reflect the free energy difference between the dimeric native and denaturation transition states at physiological temperature.

### 3.2 AGT variants show different structural and energetic differences between the native and denaturation transition states.

In Figure 3, we show the denaturation enthalpies ($\Delta H$; panel A) and $E_a$ (panel B) for the denaturation of this set of AGT enzymes as a function of their $T_m$ values. Denaturation enthalpies are strongly dependent on $T_m$ values, yielding a slope (denaturation $\Delta C_p$) of 11.3±0.8 $kcal \cdot mol^{-1} \cdot K^{-1}$. This value compares well with the theoretical $\Delta C_p$ of 14.8 $kcal \cdot mol^{-1} \cdot K^{-1}$ obtained using structure-energetics parameterizations [31] and the changes in accessible surface area (ΔASA) from the native state using the crystal structure of AGT WT (PDB code:1H0C; [35]) and a model of the unfolded states based on Gly-X-Gly tripeptides [30]. These data indicate that most of the native structure is disrupted in the thermally denatured state

of these twenty-two AGT variants, and this situation holds for the entire set of variants. The dependence of $E_a$ values on the $T_m$ of these AGT variants is much weaker, yielding a slope of 1.85±0.15 kcal·mol$^{-1}$·K$^{-1}$. If we consider this slope as an activation heat capacity change ($\Delta C_p^{\ddagger}$, that is the difference in heat capacity between the native and transition states), this may imply that the denaturation transition state is close to the native state, in terms of solvent exposure. Accordingly, determination of the kinetic *m* values (Figure 3C) in the presence of low non-denaturating urea concentrations provide relatively small values (from ~0.4 to 1.2 kcal·mol$^{-1}$·M$^{-1}$, with an average value of 0.77±0.21 kcal·mol$^{-1}$·M$^{-1}$) compared to the theoretical equilibrium *m* value estimated for AGT from its crystal structure (11.5 kcal·mol$^{-1}$·M$^{-1}$; using the parameterizations of [36]) indicating that: i) the denaturation transition state shows less than 10% of the native state surface area exposed to the solvent; ii) no clear trend for the kinetic *m* values is found for this set of AGT variants regardless of their stability. Therefore, we may conclude that the structure of the denaturation transition state is not largely altered by the mutations studied in terms of structure when the data are analyzed considering their effects on kinetic stability.

### 3.3 Enthalpy/entropy compensation occurs in denaturation free energy barriers of AGT variants.

Mutations affecting protein kinetic stability may produce comparative small changes in denaturation free energies due to the existence of large changes in the activation enthalpic and entropic contributions that almost cancel out [19,28-30]. In Figure 4A, we show the changes in activation enthalpies and entropies vs. the changes in activation free energy for the set of twenty two AGT variants extrapolated to 37$^o$C. These results clearly show that changes in activation free energies are comparative small compared to changes in activation enthalpies and entropies for this set of AGT enzymes. However, these calculations required a comparative long extrapolation to 37$^o$C, since denaturation of most AGT variants occurs at higher temperatures ($T_m \geq 70^oC$; Table 1). Nevertheless, similar calculations performed at a temperature equal to the average $T_m$ value of this set of AGT enzymes (81$^o$C) show the same behavior, despite the changes in activation free energies are somewhat smaller. We can therefore conclude that these and further analysis on activation energetics for this set of AGT mutants are quite robust towards the reference temperature used, and that changes in kinetic stability are associated to different contributions from enthalpy and entropy to denaturation barriers that are highly correlated and almost cancel out.

### 3.4 *Unfolding* and *solvation* barrier contributions to AGT kinetic stability

Previous works have proposed that denaturation activation energies (~enthalpies) contain two different contributions at the structural and energetic levels: i) contributions from unfolding, arising from the unfolding and solvation of a part of the native state surface area in the denaturation transition state (which is thus proportional to the kinetic *m* values reported in Figure 3C); ii) contribution from *solvation barriers*, networks of broken internal contacts in the denaturation transition state but not yet solvated [28,30]. We have estimated both contributions

for the denaturation of our set of AGT variants using previously described procedures [28] (see Figure 5). These results indicate that both unfolding and solvation barriers significantly contribute to the denaturation activation enthalpy of the AGT variants with no clear difference between natural and consensus-based variants (Figure 5A). We also observed a weak increase in both contributions as the $T_m$ value (a parameter that correlate excellently with the corresponding kinetic stability; Figure S3) of the AGT is increased (Figure 5B). We must note that in the calculations showed in Figure 5, the average kinetic *m* value for all AGT variants ($0.77\pm0.21$ kcal·mol$^{-1}$·M$^{-1}$) was used, even though the results obtained using the individual kinetic *m* values are quite similar but more scattered (Figure S4).

The unfolding and solvation barrier contributions originate from different types of structural changes between the native and denaturation transition states [28-30]. Using well-known structure-energetic relationships [31,37], we have used the data obtained from the dissection of the activation enthalpies into unfolding and solvation barrier contributions (Figure 5A) to estimate such structural changes (Figure 6). We must note: i) These structural changes are always small in structural terms compared to the total ΔASA for unfolding (about 13% on average; Table 2 and Figure 6A); ii) Structurally, the changes associated to solvation barriers are larger (about 2-fold; Table 2) than those found for unfolding; iii) Overall, no significant differences ($p > 0.05$, using a t-test) for the structural effect are found between natural and consensus-based variants; iv) Despite the uncertainty associated to these calculations, it seems that the structural changes associated with solvation barriers increase as the kinetic stability is raised (simply considering the $T_m$ values), while the trend for the unfolding contribution seems to be of opposite sign and weaker (Figure 6B).

### 3.5 AGT kinetic stability is not rate-limited by global unfolding kinetics

We have shown that a comprehensive experimental analysis based on a phenomenological two-state kinetic model (N→F) provides a good description of thermal denaturation of AGT proteins and yields relevant structural and energetic insight into the denaturation process. However, in our case, the two state kinetic model often requires long extrapolations to physiological temperature and do not shed light into the nature of the denaturation rate-limiting step. We must keep in mind that the two-state kinetic model is a particular case of a more general mechanism represented by:

$$N \leftrightarrow U \rightarrow F$$

in which, a (partially) unfolded state U is in equilibrium with N, but it also undergoes irreversible denaturation to the final state F. In this scenario, two-limiting scenarios are found: i) U→F is much faster than U→N, and thus, the overall denaturation rate is limiting by unfolding kinetics (N→U); ii) U→F is much slower than U→N, and hence, both the equilibrium constant for N↔U and the rate constant for the irreversible step U→F determine the overall denaturation rate [11, 38]. To discriminate between these two scenarios, we have determined the kinetics for global unfolding of selected AGT variants (WT, the highly stable natural polymorphism I340M, the very highly stable consensus HEM variant and the destabilizing disease-causing A295T variant) at high urea concentrations and different temperatures (Figure S5).

The HEM variant is kinetically resistant to high (9M) urea at physiological temperature, while I340M is 3-fold more stable than WT and A295T denatures about 4-fold faster than WT (Figure S5A). We have thus performed experiments at different urea concentrations (Figure S5B) in order to obtain the urea denaturation rate constants extrapolated to the absence of denaturant at different temperatures (Figure S5C). Control experiments showed no protein concentration dependence of denaturation rates at different temperatures and urea concentrations (Figure S5D), supporting that no dimer dissociation occurs prior to the rate limiting step of urea denaturation, in agreement with the analyses of thermal denaturation experiments (Figure S2 and Table S1). The kinetic *m* values derived from urea-induced unfolding kinetics are comparatively small and nearly temperature independent (average $m^{\ddagger}$ at different temperatures are: WT.- 0.56±0.07; A295T.- 0.32±0.02; HEM.- 0.51±0.07; I340M.- 0.69±0.14; in $kcal \cdot mol^{-1} \cdot M^{-1}$) while the activation energies are 22±12 (WT), 22±4 (A295T), 96±9 (HEM) and 130±5 (I340M) $kcal \cdot mol^{-1}$.

The comparison between rates extrapolated at 37°C obtained from irreversible denaturation by temperature and by urea-induced unfolding are shown in Figure S5E. In all cases, the rates for irreversible denaturation at 37°C extrapolated from DSC experiments are lower than those obtained from urea-induced denaturation, even though these differences widely vary among them, from ~20-fold for I340M to $3.6 \cdot 10^5$-fold for A295T. This supports that, in the context of the N↔U→F mechanism, the N→U step is not rate-limiting, and thus, the overall denaturation rate should depend on both the thermodynamic stability (i.e. the equilibrium constant for the N↔U step) and the rate of the irreversible step (U→F). Thus, these results suggest that the changes observed in kinetic stability in natural as well as in consensus-based mutants stem from either changes in thermodynamic stability or the aggregation rate constant (or aggregation propensity) of *unfolded* states.

### 3.6 Sequence-based analyses support that changes in aggregation propensities from *unfolded* states partly correlate with changes in kinetic stability

We have studied the aggregation propensity of our set of AGT variants by a widely used algorithm which models protein aggregation from unfolded protein regions [32,33] yielding two scores: $Z_{agg}$ (Figure 7), which measures the propensity to form high molecular weight aggregates (or amyloids), and $Z_{tox}$ (Figure S6), which measures the tendency to form low molecular weight aggregates (or amyloid pre-fibrils/oligomers). As shown, both methods provide qualitatively similar results. Despite these algorithms were originally developed to study amyloid type of aggregation, they may also provide information on the formation of other types of aggregates since they may share some of the molecular mechanisms [39].

In general, the studied mutations do not change the overall aggregation propensities (wt *plus* naturally occurring variants; 0.767 ±0.009, range: 0.753-0.784; wt *plus* consensus variants; 0.776 ±0.004, range: 0.768-0.781) but seem to affect the local aggregation propensities (Figure 7). Naturally-occurring variants mostly affect regions with some aggregation propensity (positive $Z_{agg}$ values; Figure 7A). Interestingly, the two polymorphisms forming the minor allele (P11L and I340M)

affect the aggregation propensity in a manner that correlates with their effect on protein kinetic stability. However, with the exception of LM-H83R, that decreases holo-protein kinetic stability due to aberrant PLP binding [19], and R197Q, which is a polymorphism not unambiguously associated with disease, the remaining mutations show effects on aggregation propensity difficult to reconcile with their effect on protein stability. This might imply, either that the algorithm does not properly capture their effects on protein kinetic stability, or that the local stability of these regions (which is not included in the algorithm) strongly affects their kinetic stability *in vitro*.

The consensus mutations often occur in regions with moderate to mild propensity towards aggregation (Figure 7C), and, in contrast to natural variants, all the consensus mutations strongly increasing protein kinetic stability show a significant reduction in the aggregation propensity considering the unfolded state as a reference (Q23R, D52E, V113A and I340M; Figure 7D). These results suggest that consensus mutations may increase protein kinetic stability by decreasing the aggregation propensity of partially unfolded states, and in the case of RHEAM, its very high kinetic stability may arise from the contribution of several consensus mutations to reduce its aggregation propensity. Additionally, the presence of stabilizing interactions observed in the crystal structure of AGT RHEAM (mostly of electrostatic nature) suggest that most of the kinetic stabilization induced by consensus mutations may arise from favorable interactions in the native state partially disrupted in the relevant denaturation transition state [24]. Therefore, consensus mutations may enhance kinetic stability by selective stabilization of the native state and also by decreasing the aggregation propensity of *partially* unfolded states.

## 4. Discussion

In this work, we have performed comprehensive mechanistic and energetic studies on a set of variants of human AGT with kinetic stabilities spanning over 11 orders of magnitude (implying changes in activation free energies up to ~15 $kcal \cdot mol^{-1}$) and including naturally-occurring (associated to PH1) and consensus variants (obtained from sequence alignment statistics). We show that the large difference in the kinetic stability stems from large and highly compensating changes in the enthalpic/entropic contributions to denaturation barriers. Interestingly, these changes in kinetic stability seem to arise from relatively small changes the structure of the relevant denaturation transition state that retains a native-like overall structure and mostly involve changes in *solvation* barriers. We cannot either exclude some local changes in structure and/or dynamics of the native state might occur upon mutation, but it is unlikely that they could explain the difference in several thousand $Å^2$ found for instance in the contributions from solvation barriers (Figure 6B), since the variants studied here showed similar hydrodynamic behavior as native dimers [19,24], comparatively similar activity [19,24], denaturation energetics consistent with similar degree of unfolding ($\Delta H_{cal}$ and $\Delta C_p$ values) and in some cases, almost superimposable crystal structures [19,24]. All these results suggest that the kinetic stability of AGT depends on a very delicate balance between the structure, energetics and dynamics of the native and denaturation transition states.

A comparison of the irreversible thermal denaturation and urea-induced global unfolding kinetics support that thermodynamic destabilization of the native state and enhanced aggregation propensity of the unfolded state modulate AGT

kinetic stability, since the irreversible denaturation step U→F from a global or (partially) unfolded state is not rate-limiting. Due to the strong correlation between the stability of holo- and apo-forms (Figure S7), this might also hold for apo-proteins. Direct assessment of thermodynamic stability cannot be obtained from equilibrium denaturation studies using urea or guanidium hydrochloride due to the irreversibility of the chemically induced unfolding [16,40]. However, bioinformatics analyses on the aggregation propensity of AGT variants from unfolded states (Figure 7 and S6) may provide an estimation of the mutational effects on the kinetics of the irreversible step U→F, and therefore, our results support that the aggregation/irreversible step might be slowed down by consensus-based mutations while no clear trend is found for naturally-occurring and destabilizing variants. This might imply than disease-associated variants mostly affect the thermodynamic stability of the native AGT, thus providing a mechanistic link for the low kinetic stability of disease-causing variants and their reduced foldability upon expression in eukaryotic cells (Figure 8).

The consensus mutations clearly enhance the kinetic stability of AGT, apparently mostly through changes in solvation barriers that modestly affect the structure/dynamics of the native and/or the denaturation transition state. Remarkably, consensus mutations target the stability of the native state by developing new favorable interaction in the AGT [19,24]. Noteworthy, consensus-mutations also seem to reduce the aggregation propensity of partially unfolded states as supported by bioinformatic algorithms. Nevertheless, consensus-variants seem to fail in improving AGT foldability inside eukaryotic cells, as shown for I340M and RHEAM variants [19,24]. Therefore, it is plausible that human AGT WT represents a stability threshold to provide sufficient stability for intracellular folding to active dimers and proper turnover inside peroxisomes (Figure 8). Thus, we speculate that the presence of consensus mutations in other vertebrate and eukaryotic AGT proteins (see Figure 9)may provide *additional* stability to compensate for the fixation of additional (destabilizing) mutations that are beneficial for AGT function in the fitness of these organisms (for instance, to change substrate specificity). However, to our knowledge, no comprehensive molecular characterization of AGT enzymes from vertebrates other than humans have been performed to test this hypothesis.

Naturally-occurring AGT variants show a very low degree of conservation among vertebrates (Figure 9), which based on the perspective of the *pseudo-equilibrium hypothesis* could explain their low fixation frequency due to their protein destabilizing effect. Interesting, the P11L and I340M polymorphisms, which are found together (as the *minor allele*) in about 20% of non-PH1 alleles, show a remarkable effect on protein stability and intracellular foldability, mostly due to the effect of P11L [19]. The minor allele is known to predispose human AGT towards additional deleterious effects of mutations, and thus, may represent a lower limit for AGT stability separating health and disease (Figure 8). Similarly, the major allele (WT AGT) may represent the lower limit for protein stability leading to optimal intracellular foldability. Some additional mutations (such as G170R and F152I) may lead to mitochondrial mistargeting possibly due to kinetic stability defects and strong interaction with molecular chaperones [16, 24]. The intracellular milieu can

modulate mitochondrial mistargeting possibly by tuning protein stability and interactions with chaperones inside cells, thus explaining why some stability issues (such as in I244T on the minor allele) may lead to aggregation and/or mitochondrial mistargeting depending on the expression system and conditions used [14,18,19]. By contrast, mitochondrial localization in vertebrate AGT proteins seems to rely mostly in the presence of a long and import competent N-terminal mitochondrial targeting sequence (MTS) [41] rather than in the presence of P11L variation (Figure 9). These results thus may support than mitochondrial localization of AGT in PH1 and along evolution in vertebrates operated through different molecular mechanisms.

Our results allow conclude that human AGT WT displays the adequate conformational stability to fold efficiently inside cells and to be targeted to peroxisomes, while the minor allele represent the threshold between health and disease, thus explaining the vast majority of PH1-associated mutations lead to misfolding by reducing protein stability [11,17,19,40, 42-44] below the LM threshold. Inside cells, AGT WT transiently interacts with molecular chaperones in order to reach its functional and import-competent state to peroxisomes [16,19]. The well-known action of molecular chaperones buffering destabilizing mutations [45], may allow disease-causing mutants to fold inside cells but to a lower extent [14,16,19,46], thus explaining loss-of-function in these mutations. Our results also explain why pharmacological therapies based on native state ligands (such as pyridoxine supplementation, which increases PLP intracellular levels), osmolytes and low temperatures, that may increase AGT stability over the value of the *minor allele*, are capable of rescuing AGT function in cells [18,47] and in patients [48,49].

**Acknowledgments.** We thank Prof. Jose Manuel Sánchez-Ruiz for his advice and support. This work was supported by grants from MINECO (BIO2012-34937 and CSD-2009-00088 and SAF2011-23933), Junta de Andalucia (P11-CTS-07187), European Union (FP7-REGPOT-CT2012-31637-IMBRAIN) and FEDER Funds. A.L.P. is supported by a Ramón y Cajal research contract from MINECO (RYC-2009-04147). N.M-T. is supported by FPI predoctoral fellowships from MINECO.

**References**

[1] W.E.Balch, R.I.Morimoto, A.Dillin, J.W.Kelly, Adapting proteostasis for disease intervention, Science 319 (2008) 916-919.

[2] E.T.Powers, R.I.Morimoto, A.Dillin, J.W.Kelly, W.E.Balch, Biological and chemical approaches to disease of proteostasis deficiency, Annu.Rev.Biochem. 78 (2009) 959-991.

[3] E.T. Powers, W.E. Balch, Diversity in the origins of proteostasis networks-a driver for protein function in evolution, Nat Rev Mol Cell Biol. 14 (2013) 237-248.

[4] G.Feller, C.Gerday, Psychrophilic enzymes: hot topics in cold adaptation, Nat.Rev.Microbiol. 1 (2003) 200-208.

[5] E.A.Gaucher, S.Govindarajan, O.K.Ganesh, Palaeotemperature trend for Precambrian life inferred from resurrected protein, Nature 451 (2008) 704-707.

[6] T.P.Schrank, D.W.Bolen, V.J.Hilser, Rational modulation of conformational fluctuations in adenylate kinase reveals a local unfolding mechanism for allostery and functional adaptation in proteins, Proc.Natl.Acad.Sci USA 106 (2009) 16984-16989.

[7] N.Tokuriki, D.S.Tawfik, Stability effects of mutations and protein evolvability, Curr.Opin.Struct.Biol. 19 (2009) 596-604.
[8] N.Tokuriki, D.S.Tawfik, Chaperonin overexpression promotes genetic variation and enzyme evolution, Nature 459 (2009) 668-673.
[9] C.M.Gomes, Protein misfolding in disease and small molecule therapies, Curr.Top.Med.Chem. 12 (2012) 2460-2469.
[10] N.Gregersen, P.Bross, S.Vang, J.H.Christensen, Protein misfoldng and human disease, Annu.Rev.Genomics.Hum.Genet.7 (2006) 103-124.
[11] A.L.Pey, Protein homeostasis disorders of key enzymes of amino acids metabolism: mutation-induced protein kinetic destabilization and new therapeutic strategies, Amino Acids 45 (2013) 1331-1341.
[12] G.Valentini, M.Maggi, A.L.Pey, Protein stability, folding and misfolding in human PGK1 deficiency, Biomolecules 3 (2013) 1030-1052.
[13] E.Salido, A.L.Pey, R.Rodriguez, V.Lorenzo, Primary hyperoxalurias: disorders of glyoxylate detoxification, Biochim.Biophys.Acta. 1822 (2012) 1453:1464.
[14] A.Santana, E.Salido, A.Torres, L.J.Shapiro, Primary hyperoxaluria type 1 in the canary islands: a conformational disease due to I244T mutation in the P11L-containing alanine:glyoxylate aminotransferase, Proc.Natl.Acad.Sci. USA 100 (2003) 7277-7282.
[15] M.B.Coulter-Mackie, Q.Lian, Consequences of missense mutation for dimerization and turnover of alanine:glyoxylate aminotransferase: study of a spectrum of mutations, Mol.Genet.Metabol. 89 (2006) 349-359.
[16] A.L.Pey, E.Salido, J.M.Sanchez-Ruiz, Role of low native state kinetic stability and interaction of partially unfolded states with molecular chaperones in the mitocondrial protein mistargeting associated with primary hyperoxaluria, Amino Acids 41 (2011) 1233-1245.
[17] B.Cellini, R.Montioli, C.B.Voltattorni, Human liver peroxisomal alanine:glyoxylate aminotransferase: characterization of the two allelic forms and their pathogenic variants, Biochim.Biophys.Acta 1814 (2011) 1577-1584.
[18] S.Fargue, J.Lewin, G.Rumsby, C.J.Danpure, Four of the most common mutations in primary hyperoxaluria type 1 unmask the cryptic mitocondrial targeting sequence of alanine:glyoxylate aminotransferase encoded by the polymorphic minor allele, J.Biol.Chem. 288 (2013) 2475-2484.
[19] N.Mesa-Torres, I.Fabelo-Rosa, D.Riverol, C.Yunta, A.Albert, E.Salido, A.L.Pey, The role of protein denaturation energetics and molecular chaperones in the aggregation and mistargeting of mutants causing primary hyperoxaluria type I, PLoS One 8 (2013) e71963.
[20] E.Oppici, A.Roncador, R.Montioli, S.Bianconi, B.Cellini, Gly161 mutations associated with primary hyperoxaluria type I induce the cytosolic aggregation and the intracelular degradation of the apo-form of alanine:glyoxylate aminotransferase, Biochim.Biophys.Acta 1832 (2013) 2277-2288.
[21] M.J.Lumb, C.J.Danpure, Functional synergism between the most common polymorphism in human alanine:glyoxylate aminotransferase and four of the most common disease-causing mutations, J.Biol.Chem. 275 (2000) 36415-36422.
[22] E.L.Williams, C.Acquaviva, A.Amoroso, F.Chevalier, M.Coulter-Mackie, C.G.Monico, D.Giachino, T.Owen, A.Robbiano, E.Salido, H.Waterham, G.Rumsby,

Primary hyperoxaluria type 1: update and additional mutation analysis of the AGXT gene, Hum.Mutat. 30 (2009) 910-917.
[23] B.Steipe, Consensus-based engineering of protein stability: from intrabodies to thermostable enzymes, Methods Enzymol. 388 (2004) 176-186.
[24] N.Mesa-Torres, C.Yunta, I.Fabelo-Rosa, J.M.Gonzalez-Rubio, J.M.Sanchez-Ruiz, E.Salido, A.Albert, A.L.Pey, The consensus-based approach for gene/enzyme replacement therapies and crystallization strategies: the case of human alanine:glyoxylate aminotransferase, Biochem.J. doi: 10.1042/BJ20140250.
[25] R.Godoy-Ruiz, F.Ariza, D.Rodriguez-Larrea, R. Perez-Jimenez, B. Ibarra-Molero, J.M.Sanchez-Ruiz, Natural selection for kinetic stability is a likely origin of correlations between mutational effects on protein energetics and frequencias of amino acid occurrences in sequence alignments, J.Mol.Biol, 362 (2006) 966-978.
[26] S.Akanuma, Y.Nakajima, S.Yokobori, M.Kimura, N.Nemoto, T.Mase, K.Miyazono, M.Tanokura, A.Yamagishi, Experimental evidence for the thermophilicity of ancestral life, Proc.Natl.Acad.Sci USA, 110 (2013) 11067-11072.
[27] V.A.Risso, J.A.Gavira, E.A.Gaucher, J.M.Sanchez-Ruiz, Phenotypic comparisons of consensus variants versus laboratory resurrections of precambrian proteins, Proteins 82 (2014) 887-896.
[28] D.Rodriguez-Larrea, S.Minning, T.V.Borchert, J.M.Sanchez-Ruiz, Role of solvation barriers in protein kinetic stability, J.Mol.Biol. 360 (2006) 715-724.
[29] A.L.Pey, N.Mesa-Torres, L.R.Chiarelli, G.Valentini, Structural and energetic basis of protein kinetic stability in human phosphoglycerate kinase 1 deficiency, Biochemistry 52 (2013) 1160-1170.
[30] M.Costas, D.Rodriguez-Larrea, L.De Maria, T.V.Borchert, A.Gomez-Puyou, J.M.Sanchez-Ruiz, Between-species variation in the kinetic stability of TIM proteins linked to solvation-barrier free energies, J.Mol.Biol. 385 (2009) 924-937.
[31] A.D.Roberston, K.P.Murphy, Protein structure and the energetics of protein stability, Chem.Rev. 97 (1997) 1251-1268.
[32] G.G.Tartaglia, A.P.Pawar, S.Campioni, C.M.Dobson, F.Chiti, M.Vendruscolo, Prediction of aggregation-prone regions in structured protein, J.Mol.Biol. 380 (2008) 425-436.
[33] G.G.Tartaglia, M.Vendruscolo, The Zyggregator method for predicting protein aggregation propensities, Chem.Soc.Rev. 37 (2008) 1395-1401.
[34] J.M.Sanchez-Ruiz, J.L.Lopez-Lacomba, M.Cortijo, P.L.Mateo, Differential scanning calorimetry of the irreversible thermal denaturation of thermolysin, Biochemistry 27 (1988) 1648-1652.
[35] X.Zhang, S.M.Roe, Y.Hou, M.Bartlam, Z.Rao, L.H.Pearl, C.J.Danpure, Crystal structure of alanine:glyoxylate aminotransferase and the relationship between genotype and enzymatic phenotype in primary hyperoxaluria type 1, J.Mol.Biol. 331 (2003) 643-652.
[36] J.K.Myers, C.N.Pace, J.M.Scholtz, Denaturant m values and heat capacity changes: relation to changes in accesible surface areas of protein unfolding, Protein Sci. 4 (1995) 2138-2148.
[37] V.J.Hilser, J.Gomez, E.Freire, The enthalpy change in protein folding and binding: refinement of parameters for structure-based calculation, Proteins 26 (1996) 123-133.

[38] I.M.Plaza del Pino, B.Ibarra-Molero, J.M.Sanchez-Ruiz, Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases, Proteins 40 (2000) 58-70.
[39] F.Rousseau, J.Schymkowitz, L.Serrano, Protein aggregation and amyloidosis: confusión of the kinds?, Curr.Opin.Struct.Biol. 16 (2006) 118-126.
[40] B.Cellini, A.Lorenzetto, R.Montioli, E.Opicci, C.B.Voltattorni, Human liver peroxisomal alanine:glyoxylate aminotransferase: different stability under chemical stress of the major allele, the minor allele, and its pathogenic G170R variant, Biochimie 92 (2010) 1801-1811.
[41] G.M.Birdsey, J.Lewin, J.D.Holbrook, V.R.Simpson, A.A.Cunningham, C.J.Danpure, A comparative analysis of the evolutionary relationship between diet and enzyme targeting in bats, marsupials and other mammals, Proc.Biol.Sci. 272 (2005) 833-840.
[42] E.Oppici, R.Montioli, A.Lorenzetto, S.Bianconi, C.Borri Voltattorni, B.Cellini, Biochemical analyses are instrumental in identifying the impact of mutations on holo and/or apo-forms and on the region(s) of alanine:glyoxylate aminotransferase variants associated with primary hyperoxaluria type I, Mol.Genet.Metab. 105 (2012) 132-140.
[43] A.M.Pittman, M.D.Lage, V.Poltoratsky, J.D.Vrana, A.Paiardini, A.Roncador, B.Cellini, R.M.Hughes, C.L.Tucker, Rapid profiling of disease alleles using a tunable reporter of protein misfolding, Genetics 192 (2012) 831-842.
[44] A.L.Pey, A.Albert, E.Salido, Protein homeostasis defects of alanine-glyoxylate aminotransferase: new therapeutic strategies in primary hyperoxaluria type I, Biomed.Res.Int. 2013 (2013) 687658.
[45] M.Soskine, D.S.Tawfik, Mutational effects and the evolution of new protein functions, Nat.Rev.Genet. 11 (2010) 572-582.
[46] A.Albert, C.Yunta, R.Arranz, A.Peña, E.Salido, J.M.Valpuesta, J.Martin-Benito, Structure of GroEl in complex with an early folding intermediate of alanine glyoxylate aminotransferase, J.Biol.Chem. 285 (2010) 6371-6376.
[47] M.J.Lumb, G.M.Birdsey, C.J.Danpure, Correction of an enzyme trafficking defect in hereditary kidney stone disease in vitro, Biochem.J. 374 (2003) 79-87.
[48] C.G.Monico, S.Rossetti, J.B.Olson, D.S.Milliner, Pyridoxine effect in type I primary hyperoxaluria is associated with the most common mutant allele, Kidney Int. 67 (2005) 1704-1709.
[49] C.S.van Woerden, J.W.Groothoff, F.A.Wijburg, C.Annink, R.J.Wanders, H.R.Waterham, Clinical implications of mutation analysis in primary hyperoxaluria type 1, Kidney Int. 66 (2004) 74

**Table 1. Parameters for thermal denaturation and kinetic stability of AGT enzymes derived from DSC analyses.**

| AGT variant | $T_m$ ($^oC$) [a] | $\Delta H$ (kcal·mol$^{-1}$)[b] | $E_a$ (kcal·mol$^{-1}$)[c] | $k_{37^oC}$ (min$^{-1}$)[d] | $m^{\neq}$ (kcal·mol$^{-1}$·M$^{-1}$)[e] |
|---|---|---|---|---|---|
| ***Naturally-occurring variants*** | | | | | |
| **WT** | 82.1 | 547±5 | 109±5 | $6.4\cdot10^{-10}$ | 0.76±0.10 |
| **P11L** | 73.8 | 399±5 | 95±3 | $1.1\cdot10^{-7}$ | 0.64±0.04 |
| **I340M** | 84.1 | 530±24 | 115±4 | $1.7\cdot10^{-11}$ | 0.62±0.11 |
| **LM** | 76.6 | 375±13 | 101±5 | $2.2\cdot10^{-8}$ | 0.94±0.07 |
| **LM-H83R** | 58.2 | 235±16 | 55±3 | $2.9\cdot10^{-3}$ | 1.19±0.02 |
| **LM-G170R** | 75.5 | 459±17 | 100±2 | $3.7\cdot10^{-8}$ | 0.96±0.05 |
| **LM-R197Q** | 77.9 | 401±6 | 110±6 | $3.7\cdot10^{-9}$ | 0.95±0.05 |
| **LM-I244T** | 75.8 | 402±9 | 103±2 | $1.1\cdot10^{-8}$ | 0.92±0.15 |
| **LM-A295T** | 77.5 | 500±10 | 105±8 | $1.2\cdot10^{-8}$ | 0.96±0.15 |
| **LM-A368T** | 76.5 | 439±6 | 102±7 | $2.1\cdot10^{-8}$ | 0.79±0.07 |
| ***Consensus variants*** | | | | | |
| **Q23R** | 84.5 | 475±10 | 112±5 | $1.1*10^{-10}$ | 0.82±0.15 |
| **S48H** | 85.5 | 520±24 | 104±3 | $2.33*10^{-10}$ | 0.66±0.06 |
| **S50H** | 67.1 | 357±17 | 67±2 | $6.82*10^{-5}$ | 0.42±0.14 |
| **D52E** | 84.4 | 544±34 | 104±6 | $1.30*10^{-10}$ | 0.76±0.06 |
| **V113A** | 83.2 | 502±13 | 101±1 | $1.98*10^{-9}$ | 0.75±0.01 |
| **HM** | 88.0 | 599±45 | 116±9 | $1.08*10^{-11}$ | 0.60±0.07 |
| **HE** | 81.5 | 486±13 | 101±1 | $7.20*10^{-10}$ | 0.85±0.06 |
| **EM** | 86.9 | 573±28 | 121±8 | $6.38*10^{-12}$ | 0.48±0.01 |
| **HEM** | 89.5 | 315±41 | 118±10 | $4.28*10^{-12}$ | 0.43±0.11 |
| **HEAM** | 91.0 | 613±20 | 122±5 | $9.5*10^{-14}$ | 0.46±0.13 |
| **RHEM** | 92.0 | 609±5 | 125±5 | $1.7*10^{-13}$ | 1.05±0.12 |
| **RHEAM** | 93.7 | 638±31 | 125±6 | $1.1*10^{-13}$ | 0.99±0.12 |

[a] Determined at 3K/min scan rate; values are from best-fits to a two-state kinetic model and the errors associated at ≤ 0.03$^o$C

[b] mean±s.d. of at least three independent experiments at different scan rates; Expressed per mole of dimer.

[c] mean±s.d. from the four consistency test proposed by [34].

[d] extrapolated from Arrhenius plots.

[e] mean±s.d. from DSC experiments at 3-5 urea concentrations and 3 different scan rates.

**Table 2. Structural changes associated to unfolding and solvation barriers components compared to the total denaturation ΔASA for the AGT variants.** Data are mean±s.d. for each group: all variants (N=22), naturally-occurring variants

(N=10) and consensus variants (N=12). In parenthesis, the corresponding range of values are shown.

| Contribution | % of total ΔASA | | |
|---|---|---|---|
| | All variants | Natural variants | Consensus variants |
| Unfolding+Solvation barriers | 13.1±1.8 (8.3-16.6) | 12.9±1.6 (9.2-15.4) | 13.3±2.0 (8.3-16.6) |
| Unfolding | 3.6±1.9 (0.4-7.3) | 4.5±1.5 (2.2-7.3) | 2.8±1.9 (0.4-5.6) |
| Solvation barriers | 9.6±3.0 (1.9-15.9) | 8.2±2.8 (1.9-13.2) | 10.5±2.8 (7.7-15.9) |

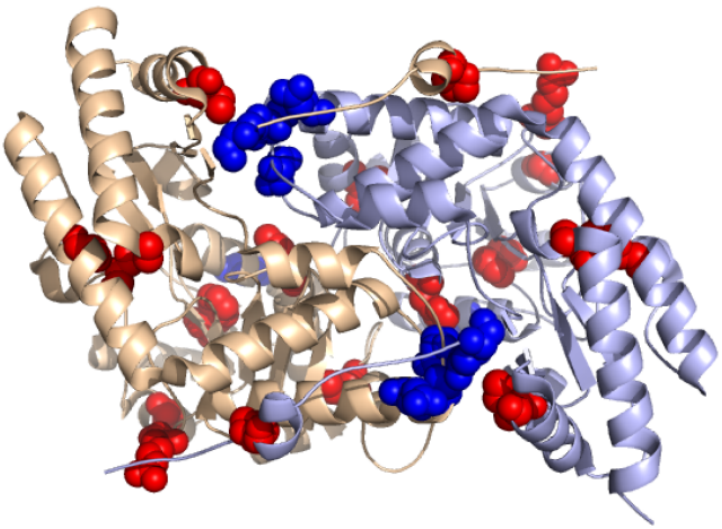

**Figure 1.** *Structural localization of mutations introduced in AGT.* Residues shown in blue are those involved in consensus mutations and those in red are changed in naturally-occurring variants.

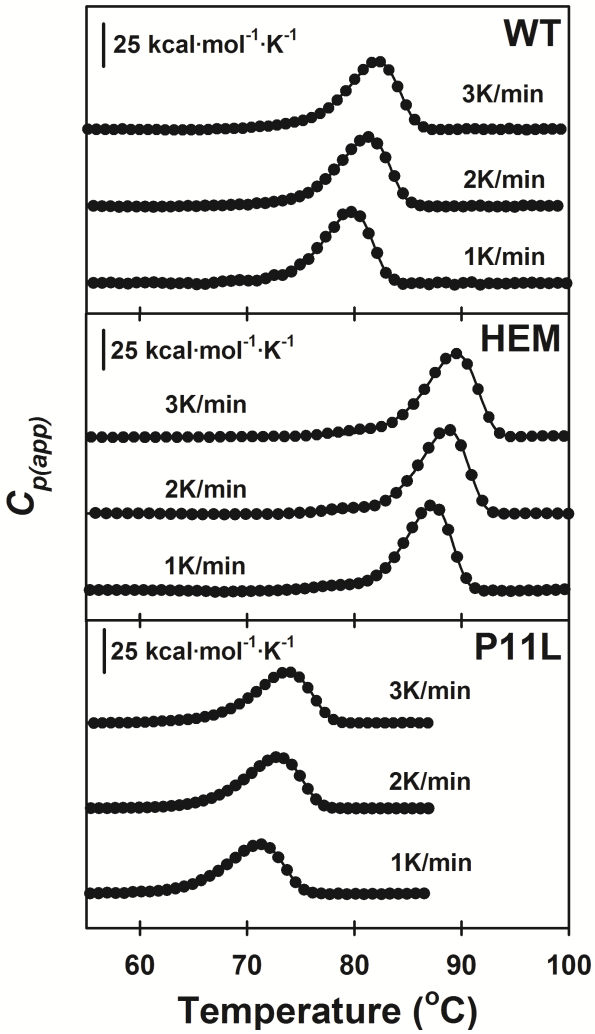


**Figure 2.** *Thermal denaturation of WT, P11L (a naturally-occurring destabilizing variant) and HEM (a consensus stabilizing variant).* Experiments were performed as described in [19,24] at the indicated scan rates. Lines are best-fits to a two-state irreversible denaturation model with first-order kinetics [16,34].

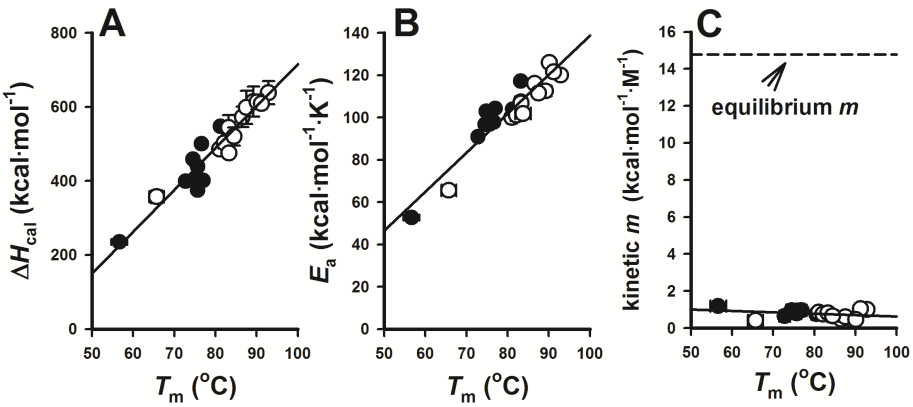


**Figure 3.** *Correlation of calorimetric enthalpies (A), activation energies (B) and kinetic m values ($m^{\ddagger}$ ) with temperature.* The x-ayis scale shows the average $T_m$ values from three different scan rates. The calorimetric enthalpies are the mean±s.d. from three different scan rates, the activation energies are from linearized Arrhenius plots (see Figure S1) and the kinetic *m* values are the mean±s.d. from three different scan rates using at least three urea concentrations. Open symbols are for consensus variants and closed symbols for naturally-occurring variants.

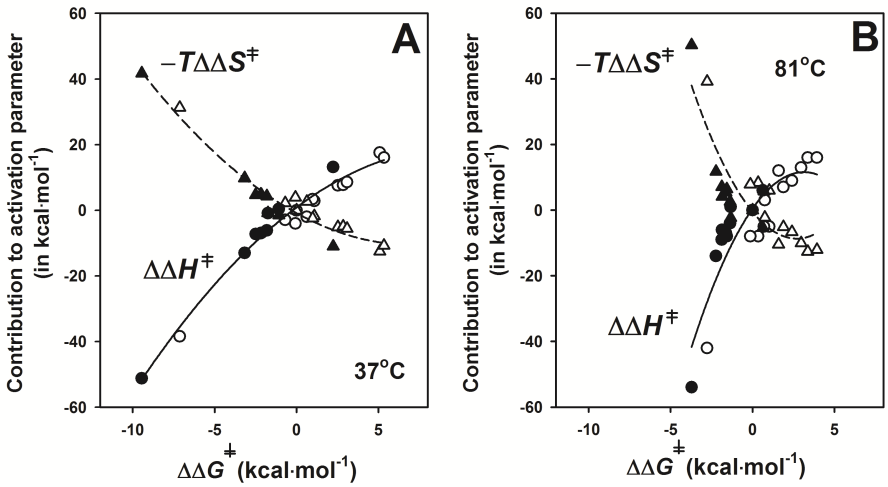


**Figure 4.** *Enthalpy/entropy compensations in the denaturation free energy barriers estimated at 37°C (panel A) or 81°C (panel B).* Open symbols are for consensus variants and closed symbols for naturally-occurring variants of AGT. Lines are only to guide the eye and have no theoretical meaning. For sake of clarity, propagated errors are not displayed.

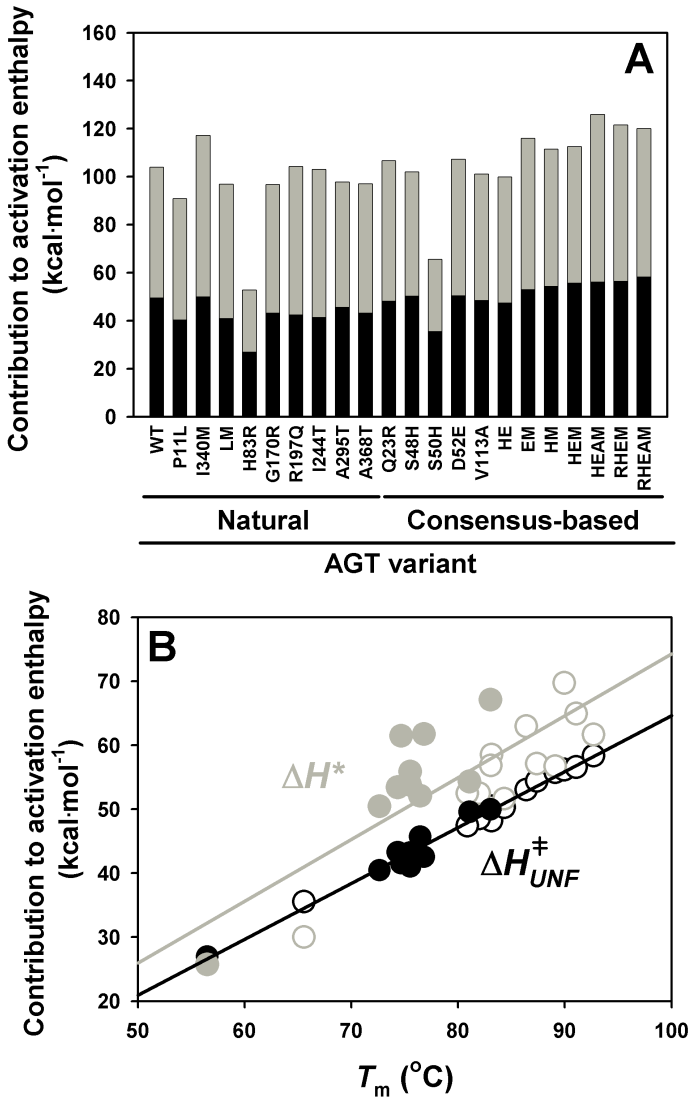


**Figure 5.** *Unfolding (black) and solvation barrier (grey) contributions to activation enthalpies.* A) contributions for individual variants; B) temperature dependence of these contributions. The values are the mean from using either the experimental $\Delta H_{cal}$ or the theoretical $\Delta H$ obtained at the $T_m$ of the AGT variant from [31]($\Delta H_{60^oC}$, 622 kcal·mol$^{-1}$ and $\Delta C_p$, 14.78 kcal·mol$^{-1}$·K$^{-1}$) from the change in $ASA_{polar}$ and $ASA_{apolar}$. The average kinetic *m* value for the 22 AGT variants (0.72±0.21 kcal·mol$^{-1}$·M$^{-1}$) is used in the calculations. The closed and open symbols in panel B correspond to naturally-occurring (closed) and consensus (open) variants. Data are calculated using two different sets of values and thus, standard errors are not displayed.

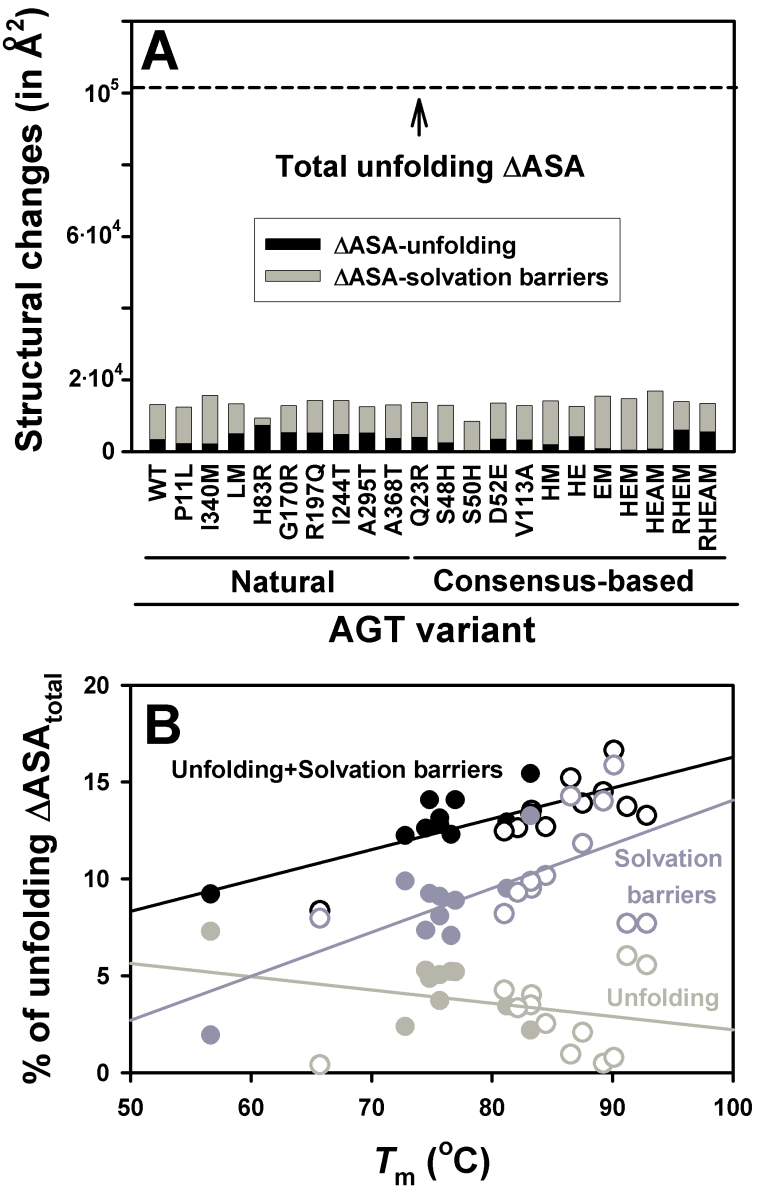


**Figure 6.** *Structural changes between the native and denaturation transition states associated to unfolding and solvation barriers.* A) Structural changes (as total ΔASA) for all AGT variants compared to the total unfolding ΔASA (the horizontal dashed line indicate the total unfolding ΔASA of AGT native state and a model for the unfolded state based on a –Gly-X-Gly- model). B) Dependence of the structural changes (as a % of the total unfolding ΔASA) on the stability ($T_m$ value) of the AGT variants. The closed and open symbols in panel B correspond to naturally-occurring (closed) and consensus (open) variants. Data are calculated from those displayed in Figure 5, and errors are not displayed.

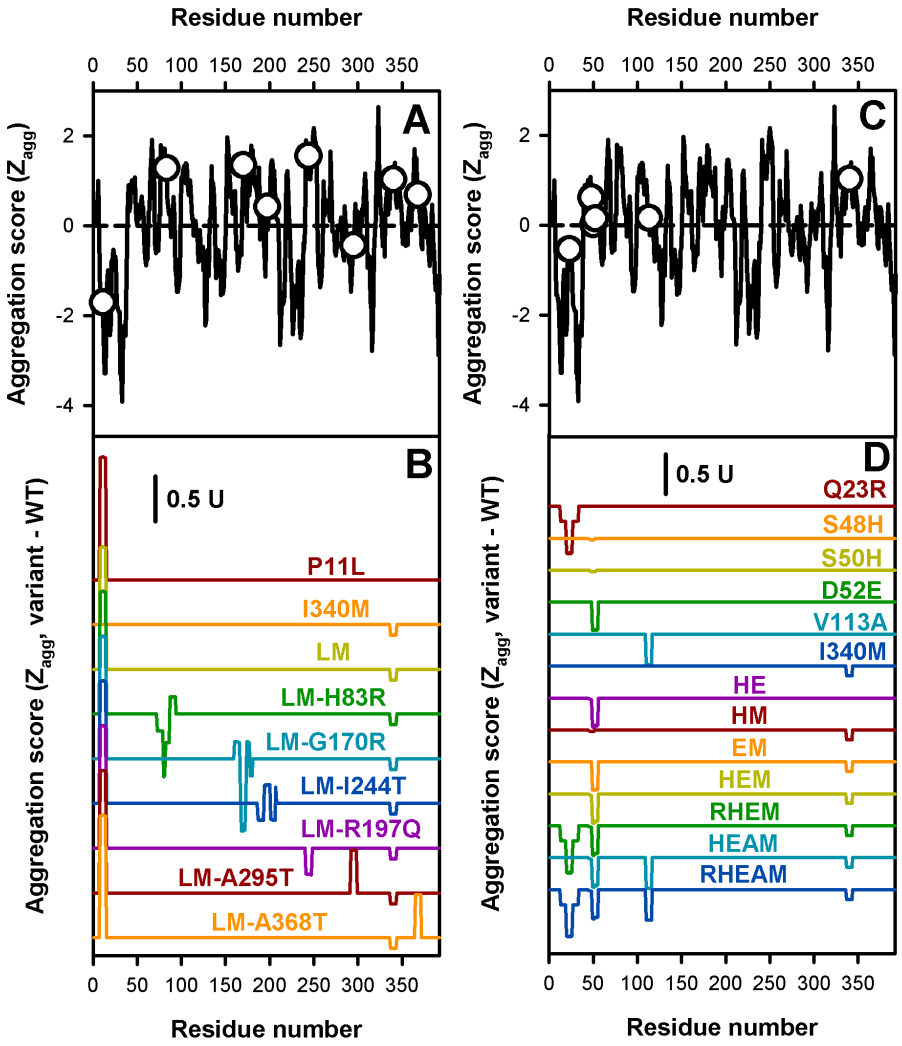


**Figure 7.** *Sequence-based prediction of aggregation propensities from unfolded states ($Z_{agg}$ score) using the Zyggregator algorithm.* The propensities per residue are shown in panels A (naturally-occurring variants) and B (consensus variants), and the position of mutant sites are indicated as circles. The difference between each variant and the WT sequence are shown in panel C and D for naturally-occurring and consensus variants, respectively.

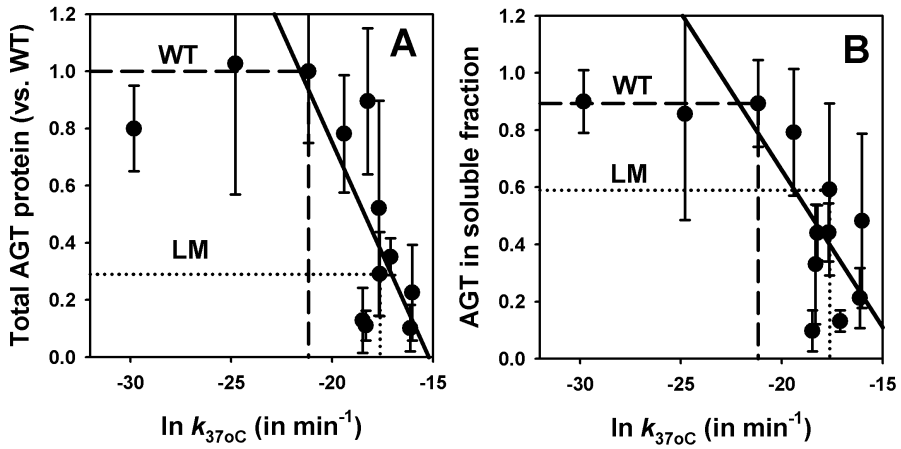


**Figure 8.** *Correlation between kinetic stability of holo AGT proteins and intracellular foldability (A, total protein levels; B, fraction of soluble protein) in eukaryotic cells.* The dashed and dotted lines show WT and LM values as a thresholds, while thick solid lines are linear fits of AGT variants with kinetic stabilities equal or lower than WT. The correlation coefficients are: $r^2$=0.56 (panel A) and 0.48 (panel B), and their statistical significance are $p<0.01$ (panel A) and <0.02 (panel B). Expression data are from [19,24].

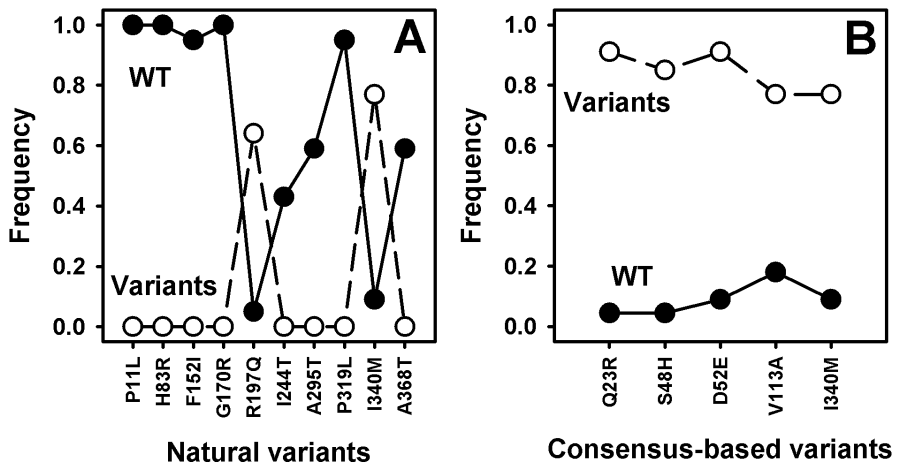


**Figure 9.** *Divergence in conservation of natural-occurring (panel A) and consensus-based (panel B) variants in AGT sequences from vertebrates.* Closed symbols indicate the frequency of the residue found in human AGT sequence while the open symbols refer to the corresponding variant. Details on sequences used for alignments has been reported elsewhere [24].