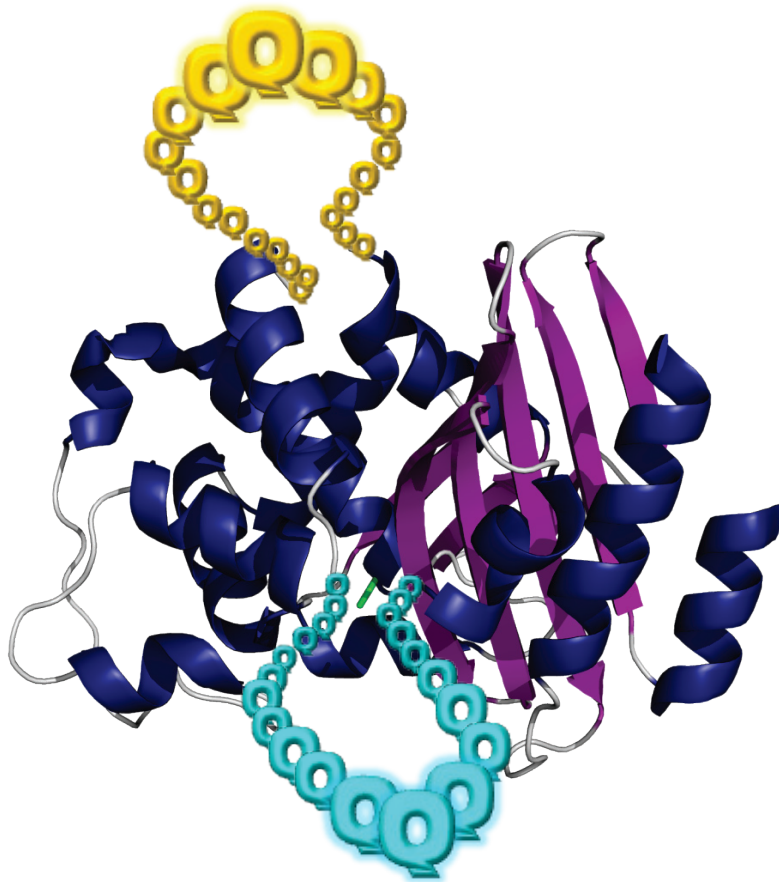




Université de Liège
Faculté des Sciences
Laboratoire d'Enzymologie et Repliement des Protéines
Prof Matagne André & Dr Dumoulin Mireille



Creation of model proteins to investigate the mechanism of aggregation of expanded-polyglutamine proteins. Insertion of polyglutamine tracts into the β -lactamase BlaP



Dissertation présentée par
SCARAFONE Natacha
en vue de l'obtention du grade de Docteur en Sciences
Année académique 2011-2012

Résumé

Les maladies à polyglutamine (polyQ) sont caractérisées par la formation d'agrégats intranucléaires de type amyloïde par des protéines contenant une expansion polyglutamine au-dessus d'une taille seuil. Ces agrégats insolubles et/ou certains de leurs précurseurs solubles pourraient jouer un rôle dans la pathogenèse de ces maladies. Le seul point commun connu entre les protéines impliquées est la présence d'un stretch polyQ étendu suggérant que leur agrégation est étroitement liée à l'expansion du polyQ au-dessus d'une taille seuil. Cependant, plusieurs études ont montré que les régions en-dehors des polyQ peuvent aussi influencer les propriétés d'agrégation des protéines à polyQ. Dans cette thèse, des protéines chimères ont été créées en insérant différentes tailles de séquences polyQ à deux positions (197-198 et 216-217) de la β -lactamase BlaP de *Bacillus licheniformis* 749/C. Les propriétés structurales et thermodynamiques de ces chimères ainsi que leurs propriétés d'agrégation en conditions natives et dénaturantes ont été étudiées en utilisant une série de techniques biophysiques dont la fluorescence, le dichroïsme circulaire, l'absorbance, la diffraction des rayons X par les fibres et la microscopie électronique à transmission.

Nous avons tout d'abord créé et caractérisé des chimères contenant 23, 30, 55 et 79Q en position 197-198 (chimères 197-198). Aucune de ces insertions n'affecte la structure de BlaP mais elles déstabilisent la protéine de 7,6-8,8 kJ/mol quelle que soit la longueur du polyQ. Similairement à ce qui est observé avec les protéines associées aux maladies, il existe un nombre seuil de glutamines au-dessus duquel les chimères de BlaP s'agrègent en fibres de type amyloïde. Ce seuil est compris entre 55 et 79Q en conditions natives et entre 30 et 55Q en conditions dénaturantes. La valeur de cette taille seuil dépend par conséquent de l'intégrité structurale de BlaP et donc des contraintes stériques et/ou conformationnelles exercées sur le stretch polyQ par BlaP.

Ensuite, nous avons comparé les propriétés de ces chimères avec celles de chimères contenant des séquences polyQ de mêmes longueurs en position 216-217 (chimères 216-217). La structure tertiaire de BlaP est altérée dans les chimères 216-217 et ces chimères sont moins stables que les chimères 197-198. Cependant, les expériences de dénaturation à l'équilibre par l'urée suggèrent que les chimères 216-217 forment une série d'états intermédiaires discrets entre les états natif et dénaturé ou que leur état dénaturé diffère significativement de celui des chimères 197-198. Finalement, les propriétés d'agrégation des chimères 216-217 diffèrent de celles des chimères 197-198. Premièrement, en conditions natives, le nombre seuil de glutamines au-dessus duquel les chimères 216-217 forment des fibres (compris entre 30 et 55Q) est plus faible que celui observé pour les chimères 197-198 (compris entre 55 et 79Q). Ce résultat suggère que les contraintes stériques et/ou conformationnelles exercées sur le stretch polyQ sont plus faibles lorsque celui-ci est inséré en position 216-217. Deuxièmement, les résultats obtenus en conditions natives et dénaturantes indiquent que BlaP pourrait assister l'agrégation en fibres des chimères 216-217 alors qu'il a un effet inhibiteur ou pas d'effet sur la formation de fibres par les chimères 197-198 contenant 55Q et 79Q, respectivement.

En conclusion, l'ensemble des résultats obtenus montrent que les propriétés d'agrégation des chimères de BlaP résultent d'un effet combiné complexe entre la propension du stretch polyQ à promouvoir la formation de fibres et l'effet modulateur de BlaP. Les chimères de BlaP sont donc de bons modèles pour étudier en détail comment les propriétés de la protéine hôte influencent les propriétés d'agrégation des protéines à expansion polyQ.

Summary

Polyglutamine (polyQ) diseases are characterized by the formation of intranuclear amyloid-like aggregates by proteins containing an expansion of a polyQ tract above a threshold length. These insoluble aggregates and/or some of their soluble precursors are thought to play a role in the pathogenesis of the diseases. The only known common point between the causative proteins is the expanded polyQ tract, suggesting that their aggregation critically depends on the expansion of the polyQ tract above a threshold length. Several studies have however shown that the non-polyQ regions can also influence the aggregation behavior of polyQ proteins. In this work, polyQ chimeras were created by inserting different polyQ lengths in two positions (197-198 and 216-217) of the β -lactamase BlaP from *Bacillus licheniformis* 749/C. The structural and thermodynamic properties of the polyQ chimeras as well as their aggregating properties under native and denaturing conditions were investigated using a range of biophysical techniques including fluorescence, circular dichroism, absorbance, x-ray fiber diffraction and transmission electron microscopy.

We have first created and characterized chimeras with 23, 30, 55 and 79Q inserted in position 197-198 (chimeras 197-198). None of these insertions modifies the structure of BlaP; they do, however, significantly destabilize the enzyme by 7.6-8.8 kJ/mol independently of the polyQ length. Similarly to the proteins associated with diseases, there is a threshold number of glutamines above which the BlaP chimeras aggregate into amyloid-like fibrils. It is comprised between 55 and 79Q and between 30 and 55Q under native and denaturing conditions, respectively. This threshold value therefore depends on the structural integrity of BlaP and thus on the steric and/or conformational constraints applied by BlaP to the polyQ tract.

We have then compared the properties of these chimeras with those of the chimeras containing polyQ of the same lengths in position 216-217 (chimeras 216-217). The tertiary structure of BlaP is slightly perturbed in the chimeras 216-217 and these chimeras are less stable than the chimeras 197-198. However, the urea-induced equilibrium unfolding experiments suggest that chimeras 216-217 populate a series of discrete intermediate states between the native and unfolded states or that their unfolded state significantly differs from that of chimeras 197-198. Finally, the aggregating properties of the polyQ chimeras 216-217 differ from those of chimeras 197-198. First, the threshold number of glutamines above which chimeras 216-217 readily form fibrils under native conditions (between 30 and 55Q) is lower than that observed for chimeras 197-198 (between 55 and 79Q) suggesting that the steric and/or conformational constraints imposed on the polyQ tract are lower when it is inserted in position 216-217. Secondly, the results obtained under both native and denaturing conditions indicate that the BlaP moiety could assist fibril formation by chimeras 216-217 while it has an inhibiting effect and no effect on fibril formation by the chimeras 197-198 with 55 and 79 glutamines, respectively.

Taken together, these results show that the aggregating properties of BlaP chimeras result from a very complex interplay between the propensity of the polyQ tract to mediate fibril formation and the modulating effect of the BlaP moiety. BlaP chimeras present therefore valuable models to investigate in details how the properties of the host protein influence the aggregation behavior of expanded polyQ proteins.

Table of contents

Abbreviations	1
Chapter 1. Introduction	3
1. Protein synthesis and folding	3
2. Misfolding diseases: the group of amyloid diseases	3
3. Amyloid fibrils	5
4. Polyglutamine (polyQ) diseases	8
4.1. What are polyQ diseases?.....	8
4.2. Common features of polyQ diseases	8
4.2.1. Existence of a pathological threshold number of glutamines.....	8
4.2.2. Anticipation phenomenon	8
4.2.3. Formation of intranuclear inclusions.....	9
4.2.4. Gain of toxic function.....	10
4.3. PolyQ disease-associated proteins.....	10
4.3.1. Expression and localization.....	10
4.3.2. Function(s).....	11
4.3.3. Size and structure	13
4.3.4. Role of glutamine repeats in proteins.....	13
4.3.5. The polyQ tract: the common critical determinant for polyQ diseases.....	14
4.4. The role of protein misfolding and aggregation in polyQ diseases.....	14
4.5. Aggregation mechanism of expanded-polyQ proteins	16
4.5.1. PolyQ peptides, the simplest models to investigate polyQ-mediated aggregation....	16
4.5.2. Influence of protein context.....	17
4.5.3. Aggregation of proteins associated with polyQ diseases	19
4.5.3.1. Ataxin 3	19
4.5.3.2. Huntingtin.....	21
4.5.3.3. Ataxin-1.....	26
4.5.4. Creation and characterization of model proteins	27
4.5.4.1. Myoglobin-derived polyQ proteins	28
4.5.4.2. GST-derived polyQ proteins	29
4.5.4.3. CRABPI-derived polyQ proteins	30
4.5.4.4. Thioredoxin-derived polyQ proteins	31
4.5.4.5. SpA-derived polyQ proteins.....	31
4.5.4.6. Apomyoglobin-derived polyQ proteins.....	32
4.5.5. Conclusions	33
Chapter 2. Objectives	37
1. Open questions	37
2. What is BlaP and why is it a potentially good scaffold to create model polyQ proteins?.....	38
3. How the location of the permissive sites have been determined?	39
4. Strategy and outline of the work	40
Chapter 3. Materials and methods	43
1. Molecular biology	43
1.1. Construction of the expression vectors of chimeric β -lactamases with polyQ insertions in position 197-198.....	43
1.2. Construction of the expression vectors of chimeric β -lactamases with polyQ insertions in position 216-217.....	44
2. Protein expression and purification.....	45

3. Size-exclusion chromatography	46
4. N-terminal sequencing and mass spectrometry	46
5. Quantification of the protein	46
6. Enzymatic activity measurements	46
7. Fluorescence and circular dichroism (CD) measurements	47
8. Fluorescence and circular dichroism spectra of native proteins.....	47
9. Urea-induced unfolding experiments	47
10. Heat-induced unfolding.....	49
11. Aggregation kinetics.....	49
12. Thioflavin T (ThT) fluorescence measurements	50
13. Transmission electron microscopy.....	50
14. X-ray fibre diffraction	50
15. Fibril activity	51
Chapter 4. Creation and characterization of BlaP chimeras with polyQ tracts in position 197-198	53
1. Creation of chimeras with polyQ inserts in position 197-198.....	53
2. Production and purification of BlaP ₁₉₇₋₁₉₈ and its polyQ chimeras	54
3. Effects of polyQ insertions in position 197-198 on the catalytic properties of BlaP ₁₉₇₋₁₉₈	57
4. The polyQ tract adopts a disordered structure and does not perturb the overall structure of BlaP ₁₉₇₋₁₉₈	58
5. The polyQ tract destabilizes BlaP ₁₉₇₋₁₉₈ and the extent of destabilization is largely independent of the polyQ length.....	60
6. The threshold length of the polyQ tract above which chimeras aggregate into amyloid-like fibrils depends on the structural integrity of BlaP ₁₉₇₋₁₉₈	66
6.1. Aggregation under denaturing conditions	66
6.2. Aggregation under native conditions.....	70
7. Amyloid-like fibrils formed by BlaP ₁₉₇₋₁₉₈ (Gln) ₇₉ are enzymatically inactive	74
8. Conclusions	76
Chapter 5. Creation and characterization of BlaP chimeras with polyQ tracts in position 216-217	77
1. Creation of chimeras with polyQ inserts in position 216-217.....	77
2. Production and purification of BlaP ₂₁₆₋₂₁₇ and its polyQ chimeras	78
3. Effects of polyQ insertions in position 216-217 on the catalytic properties of BlaP	81
4. Effects of polyQ insertions in position 216-217 on the overall structure of BlaP.....	83
5. Effects of polyQ insertions in position 216-217 on the stability of BlaP.....	86
6. Effects of polyQ insertions in position 216-217 on the aggregating properties of BlaP	92
6.1. Aggregation under native conditions.....	93
6.2. Aggregation under denaturing conditions	95
7. Conclusions	97
Chapter 6. Conclusions and perspectives	99
1. Conclusions	99
2. Perspectives.....	101
References	105
Annexes	115
Annexe 1	115
Annexe 2	116
Annexe 3	117

Abbreviations

a.a., amino acid
ApoMb, apomyoglobin
ASA, accessible surface areas
Atx, ataxin
BCA, bicinchoninic acid
BlaP, β -lactamase from *Bacillus licheniformis* 749/C
BlaP₁₉₇₋₁₉₈(Gln)_x, BlaP with X glutamines inserted in position 197-198
BlaP₂₁₆₋₂₁₇(Gln)_x, BlaP with X glutamines inserted in position 216-217
BSA, bovine serum albumin
bp, base pair
CACNA1 α , α subunit of voltage-dependent calcium channel
CAG, codon for glutamine
CD, circular dichroism
C_m, concentration for mid-unfolding
CNS, central nervous system
CRABPI, cellular retinoic acid binding protein I
 $\Delta G^{\circ}_{N-U}(\text{H}_2\text{O})$, difference in the free energy between the folded (N) and unfolded (U) states
DNA, deoxyribonucleic acid
dNTP, deoxyribonucleotide triphosphate
DRPLA, dentatorubral pallidoluysian atrophy
ESI-QTOF MS, electrospray ionization-quadrupole time-of-flight mass spectrometry
GST, Glutathione S-transferase
HD, Huntington's disease
Htt, huntingtin
LB, Luria-Bertani
m, dependence of the free energy on denaturant concentration
Mb, myoglobin
MBP, maltose binding protein
N, native state
Ni-PDC, nickel pentadentate chelator
PBS, 50 mM phosphate buffer containing 150 mM NaCl
PCR, polymerase chain reaction
PolyQ, polyglutamine
PPI, prolyl peptidyl isomerase
Q, glutamine
SCA, spinocerebellar ataxia
SBMA, spinobulbar muscular atrophy
SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC, size exclusion chromatography
SpA, B domain of the protein A from *Staphylococcus aureus*

Abbreviations

TB, terrific broth

TBP, TATA-box binding protein

TEM, transmission electron microscopy

ThT, thioflavin T

T_m^{app} , apparent temperature of mid-transition

Trp, tryptophan residue

U, unfolded state

UIM, ubiquitin interacting motif

UPS, ubiquitin proteasome system

Chapter 1. Introduction

1. Protein synthesis and folding

Proteins are the major functional components of the cell and they are involved in numerous cellular processes such as transport, regulation of transcription and protein synthesis. They may act alone or in conjunction with other proteins, nucleic acids, small molecules or ions. Proteins are synthesized on the ribosomes as linear chains of amino acids linked together in a specific sequence; each protein is a particular combination of the 20 different amino acids. In order to accomplish their biological role(s), the majority of proteins have to acquire a globular three-dimensional structure (or native structure) upon a process called protein folding. On the other hand, some proteins lack such a native globular structure, at least when they are not associated with their binding partners, and others contain both disordered regions and folded globular domains (Campioni *et al.*, 2010).

In the crowded environment of cells, the protein folding process is assisted by a large number of auxiliary factors such as molecular chaperones and folding catalysts (Gething and Sambrook, 1992). These factors help the polypeptide chains to correctly and efficiently fold but they do not however determine the final native conformation of the protein. This latter is entirely determined by the amino acid sequence of proteins (Anfinsen, 1973). The chaperones protect the nascent polypeptide chain from non-native interactions and from intermolecular interactions. The folding catalysts accelerate the slow steps in the folding process. The most important are the prolyl peptidyl isomerases (PPI) which catalyse the isomerization of Xaa-Pro bonds (Schiene and Fischer, 2000).

Despite these auxiliary factors, it is not exceptional that misfolded proteins are still produced and nature has therefore evolved a sophisticated quality control system to differentiate between correctly folded and misfolded proteins (Ellgaard and Helenius, 2001). Proteins that are identified as incorrectly folded are predetermined to be degraded by proteasomes.

2. Misfolding diseases: the group of amyloid diseases

Despite all these control systems, the probability that some proteins fail to adopt or to remain in their functional conformation is however not negligible. Given the crucial roles of proteins in cells, this failure generally leads to serious diseases which are known as protein conformational diseases or misfolding diseases (Chiti and Dobson, 2006).

The most common and studied group of these diseases comprises more than thirty disorders associated with the conversion of a specific peptide or protein from its soluble state into extracellular amyloid fibrils or intracellular amyloid-like fibrils (Chiti and Dobson, 2006). These disorders are referred to as amyloid diseases. Amyloid-like fibrils are fibrillar aggregates which are deposited intracellularly as inclusions but which typically possess all the characteristics of amyloid fibrils (see below). Table 1 gives an overview of some of these conformational diseases.

Table 1: Human diseases associated with formation of extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics. Table adapted from (Chiti and Dobson, 2006).

DISEASES	AGGREGATING PROTEIN OR PEPTIDE
Neurodegenerative diseases	
Alzheimer's disease	Amyloid β peptide
Spongiform encephalopathies	Prion protein or fragments thereof
Parkinson's disease	α -Synuclein
Dementia with Lewy bodies	α -Synuclein
Frontotemporal dementia with Parkinsonism	Tau
Amyotrophic lateral sclerosis	Superoxide dismutase 1
Huntington's disease	Huntingtin with polyQ expansion
Spinocerebellar ataxias	Ataxins or TATA box-bindings with polyQ expansion
Spinal and bulbar muscular atrophy	Androgen receptor with polyQ expansion
Non-neuropathic systemic amyloidoses	
AL amyloidosis	Immunoglobulin light chains or fragments
AA amyloidosis	Fragments of serum amyloid A protein
Senile systemic amyloidosis	Wild-type transthyretin
Hemodialysis-related amyloidosis	β 2-microglobulin
ApoAI, II or IV amyloidosis	N-terminal fragments of apolipoprotein AI, II or IV
Lysozyme amyloidosis	Mutants of lysozyme
Fibrinogen amyloidosis	Variants of fibrinogen α -chain
Icelandic hereditary cerebral amyloid angiopathy	Mutant of cystatin C
Non-neuropathic localized diseases	
Type II diabetes	Amylin, also called islet amyloid polypeptide (IAPP)
Injection-localized amyloidosis	Insulin
Cataract	γ -Crystallins
Pulmonary alveolar proteinosis	Lung surfactant protein C
Inclusion-body myositis	Amyloid β peptide
Aortic medial amyloidosis	Medin
Pituitary prolactinoma	Prolactin

Amyloid diseases can be systemic (affecting many different organs) or organ-specific (Table 1). When the organ affected is the brain, this leads to neurodegenerative diseases such as Alzheimer's and Huntington's diseases (Table 1). Some amyloid diseases are inherited, due to destabilizing mutations in the precursor protein; others are caused by overabundant or abnormal protein production, or *infectious* agents such as prions. It is noteworthy however, that most of the aggregation diseases are not associated with either genetic mutations or *infectious* agents but are caused by sporadic events in particular at old age. The amyloid diseases are often hard to diagnose and are in most cases incurable, frequently leading to premature death. They often manifest late in life and inflict enormous psycho-sociological and economical burdens in western societies. It has been estimated, for example, that 10 % of the population over 60 and 50% of the population over 85 will develop Alzheimer's disease (Zurdo, 2005). Extensive research is therefore being carried out to tackle the prevention and treatment of these diseases.

3. Amyloid fibrils

Amyloid fibrils are highly organized aggregates. In transmission electron microscopy, they appear as unbranched structure of indefinite length (typically more than 1 μm) (Figure 1). They generally consist of several protofilaments (typically 2-6) twisted together (rope-like fibrils) (Serpell *et al.*, 2000, Sunde and Blake, 1997) or associated laterally (ribbon-like fibrils) (Bauer *et al.*, 1995, Pedersen *et al.*, 2006, Saiki *et al.*, 2005). Each protofilament is about 2-5 nm in diameter.

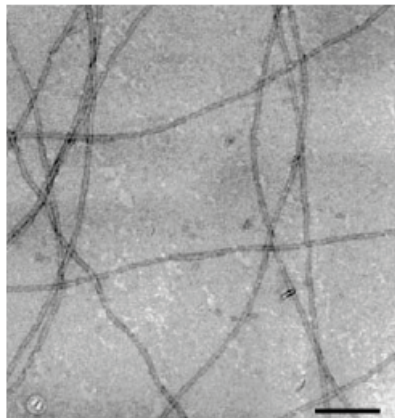


Figure 1: Image of human lysozyme amyloid fibrils obtained after negative staining with uranyl acetate by transmission electron microscopy. Scale bar, 400 nm (Stefani and Dobson, 2003).

Amyloid fibrils are structures rich in β -sheets. Their X-ray diffraction pattern is typical of a cross- β structure in which the β -sheets are parallel to the axis of the fibril and their β -strands run perpendicularly to this same axis (Figure 2A, B and C). This X-ray fiber diffraction pattern is characterized by an equatorial reflection at 9-12 Å and a meridional reflection at around 4.7 Å which correspond to the face-to-face separation of two β -sheets and the space between two adjacent β -strands, respectively (Figure 2A and B) (Sunde and Blake, 1997).

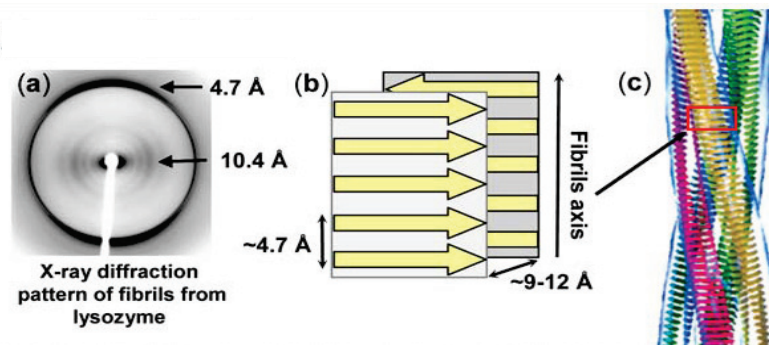


Figure 2: (A) Example of X-ray diffraction pattern obtained with amyloid fibrils grown from human lysozyme. (B) Schematic representation of a cross- β structure. (C) Cartoon model of an amyloid fibril. From (Dumoulin and Dobson, 2004).

The cross- β of an amyloid fibril confers to this latter the ability to bind specific dyes such as Congo red and thioflavin T (ThT) (Nilsson, 2004). The binding of the dye congo red results in green birefringence under cross-polarized light. The ThT dye fluorescence is extensively used to follow fibril formation *in vitro*.

There is no sequence homology between the proteins or peptides associated with amyloidoses and in their native state, these proteins and peptides do not share any structural homology. Nevertheless, in their amyloid state, they all adopt highly organized fibrillar structures with the characteristics described above (Stefani and Dobson, 2003). Moreover, a number of studies have shown that disease-unrelated proteins can also form amyloid fibrils *in vitro* when they are placed in particular conditions (Stefani and Dobson, 2003, Uversky and Fink, 2004). These observations have led to the idea that the ability to form amyloid fibrils is an intrinsic property of the polypeptide main chain and that therefore any protein can form fibrils *in vitro* when incubated under appropriate conditions. It is therefore assumed that the formation of the core structure of the fibrils implies hydrogen bonds between the carbonyl and amide group of the main chain (Dobson, 2004a).

Although the formation of amyloid fibrils is a generic property of the polypeptide main chain, the details of the structure (as the lengths of the β -strands and whether they are arranged in parallel or antiparallel within each sheet, the fraction of the residues of a polypeptide chain included in the core structure) are determined by the interactions of the various side-chains

with each other and with the solvent (Chiti and Dobson, 2006). The amino acid sequence of the protein may influence the stability of the amyloid fibrils and the propensity of the protein or peptide to aggregate into amyloid fibrils (Chiti and Dobson, 2006, Dobson, 2004b).

The aggregation into amyloid fibrils occurs generally via a nucleation-elongation mechanism which generally comprises a lag phase followed by an exponential growth phase (Chiti and Dobson, 2006). The lag phase corresponds to time required for nuclei to form. Once a nucleus is formed, fibrils growth is thought to proceed rapidly by further association of monomers or oligomers with the nucleus. In this type of aggregation mechanism, the addition of preformed fibrils which can act as seed shortens or suppresses the lag phase (Figure 3).

The aggregation process is generally complex and a series of prefibrillar species including the nucleus and a series of soluble aggregates is usually formed on the pathway of fibril formation (Figure 4) (Dobson, 2003). The earliest species visible by electron or atomic force microscopy generally resemble small bead-like structures and are often described as amorphous aggregates or micelles (Figure 4A). They then transform into species with more distinct morphologies, referred to as protofibrils or protofilaments (Figure 4B). These structures are commonly short, thin, sometimes curly, fibrillar species that assemble into mature fibrils (Figure 4C).

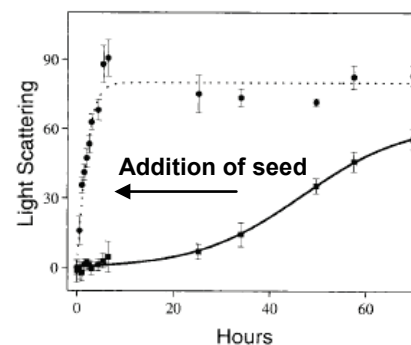
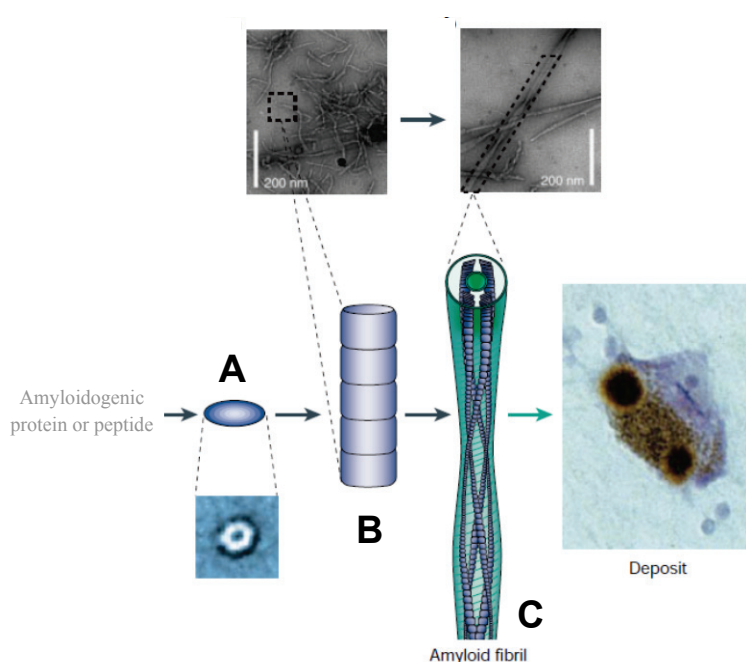


Figure 3: Aggregation time-courses of the K₂Q₃₇K₂ peptide in absence of added preformed fibrils (square) and following addition of 5% of preformed fibrils from K₂Q₃₇K₂ peptides (circles) (Chen *et al.*, 2001).

It is becoming evident that mature fibrils are rather biologically inert species and that the prefibrillar species are the most cytotoxic species (Glabe and Kaye, 2006, Walsh and Selkoe, 2004).

Figure 4: The different prefibrillar species formed during fibril formation. The earliest species generally resemble bead-like structures (A), they then transform into structures called protofibrils or protofilaments (B) which assemble into mature fibrils (C). Adapted from (Dobson, 2003).

4. Polyglutamine (polyQ) diseases

4.1. What are polyQ diseases?

Polyglutamine (polyQ) diseases are inherited progressive and unremitting neurodegenerative disorders caused by an unstable expansion of CAG repeats in the coding region of unrelated genes (Orr and Zoghbi, 2007). CAG is coding for glutamine and the expansion of these CAG repeats is therefore translated into an expanded polyQ stretch in the corresponding protein. At least, nine polyQ diseases are known so far, including Huntington's disease (HD), (the most prevalent one), several spinocerebellar ataxias (SCA1, 2, 3, 6, 7 and 17), dentatorubral pallidoluysian atrophy (DRPLA) and spinobulbar muscular atrophy (SBMA) (Table 2). These diseases are autosomal-dominant inherited disorders, except SBMA which is an X-linked recessive disease. PolyQ diseases are characterized by a neuronal cell loss and depending on the disease, different neuronal subtypes within the central nervous system (CNS) are affected (Table 2) (Hands and Wyttenbach, 2010). Each disease is characterized by specific symptoms, including neurological and psychiatric symptoms (Table 2).

4.2. Common features of polyQ diseases

PolyQ diseases share several common characteristics suggesting that they have a common pathophysiological mechanism. These common characteristics are described below.

4.2.1. Existence of a pathological threshold number of glutamines

Each of the nine polyQ disorders is associated with one specific protein (Table 2). In healthy people, a polyQ tract is already present in the sequence of these proteins and the number of glutamine in this tract is polymorphic in the human population (Table 2). The disease only occurs when the polyQ tract is expanded due to mutations above a threshold length; this pathological threshold length slightly varies from one disease to another but for most of them, it is comprised between 35 and 45 glutamines (Table 2).

4.2.2. Anticipation phenomenon

With moderate sized expansion (i.e. slightly above the pathological threshold length), polyQ diseases typically manifest at mid-life. However, the expanded CAG repeat (coding for the expanded polyQ tract) is unstable and tends to expand in the successive generations, further leading to earlier age of onset and also more severe pathology (Bauer and Nukina, 2009); this

phenomenon is called the “anticipation phenomenon”. In the case of Huntington’s disease (HD), expansions ranging from 40 to 55 glutamines generally lead to HD cases with adult onset whereas expansions of 70 and more glutamines invariably cause juvenile forms of the disease.

Table 2: Polyglutamine (polyQ) disorders

Disease	Protein	Normal repeat length	Expanded repeat length	Location of cellular toxicity in the CNS	Main clinical features
HD	Huntingtin (Htt)	6-34	36-121	Caudate nuclei, striatum, cerebral cortex	Psychiatric, cognitive and motor abnormalities
SBMA	Androgen receptor (AR)	9-36	38-62	Spinal anterior horn, facial nucleus, hypoglossal nucleus, skeletal muscle	Proximal muscle and bulbar muscle weakness, atrophy and fasciculation, endocrine abnormalities, hand tremor, muscle cramps, neuropathy
DRPLA	Atrophin-1	7-34	49-88	Globus pallidus, subthalamic nucleus, dentate nucleus, cerebral, cerebellar white matter	Ataxia, epilepsy, myoclonus, choreoathetosis, dementia
SCA1	Ataxin-1	6-44 ⁽¹⁾	39-82	Cerebellar purkinje cells, brainstem, spinal cord, globus pallidus	Ataxia, progressive motor deterioration, dysarthria, oculomotor palsy, amyotrophy and sensory disturbance
SCA2	Ataxin-2	15-24	32-200	Cerebellar purkinje cells and granular cells, brainstem, substantia nigra, spinal cord, thalamus	Ataxia, Parkinsonian rigidity, bradykinesia, amyotrophy, sensory disturbance, mental deterioration
SCA3	Ataxin-3	13-36	45-84	Cerebellum, substantia nigra, brainstem, dentate nucleus, globus pallidus, spinal cord, thalamus	Ataxia, dystonia, Parkinsonism, neuropathy, amyotrophy, ophthalmoplegia, nystagmus, eyelid retraction, faciolingual fasciculation, dysphagia, weight loss
SCA6	Ataxin-6 (CACNA1A)	4-19	10-33	Cerebellar purkinje cells, dentate nucleus, inferior olive	Ataxia, dysarthria, oculomotor disorders, incontinence, peripheral neuropathy
SCA7	Ataxin-7	4-35	37-306	Cerebellar purkinje cells, dentate nucleus, inferior olive, spinal cord, retina, brainstem	Ataxia, retinal degeneration, dysphagia, dysarthria
SCA17	TATA-binding protein (TBP)	25-42	47-63	Cerebral cortex, cerebellum	Ataxia, dementia, epilepsy

HD, Huntington’s disease; **SBMA**, spinobulbar muscular atrophy; **DRPLA**, Dentatorubral-pallidoluyisian atrophy; **SCA**, spinocerebellar ataxia; **CACNA1A**, α subunit of voltage-dependent calcium channel; **TBP**, TATA-box binding protein. ⁽¹⁾Normal SCA1 alleles ≥ 21 CAG are interrupted with 1-2 CAT, whereas disease-causing alleles are pure (CAG)_n. Information presented in this table comes from (Hands and Wyttenbach, 2010) and (Padiath *et al.*, 2005).

4.2.3. Formation of intranuclear inclusions

Another characteristic common to all polyQ diseases is the formation by the causative proteins containing an expanded polyQ tract of neuronal intranuclear inclusions (NIIs) containing amyloid-like fibrils (Figure 5A, B and C) (Diaz-Hernandez *et al.*, 2004, Difiglia *et al.*, 1997, Gourfinkel-An *et al.*, 1998, Gutekunst *et al.*, 1999, Holmberg *et al.*, 1998, Paulson *et al.*, 1997b, Skinner *et al.*, 1997, Cummings *et al.*, 1998). In some polyQ diseases such as HD and SCA3, the formation of cytoplasmic inclusions was also observed (Difiglia *et al.*, 1997, Gutekunst *et al.*, 1999, Seidel *et al.*, 2010).

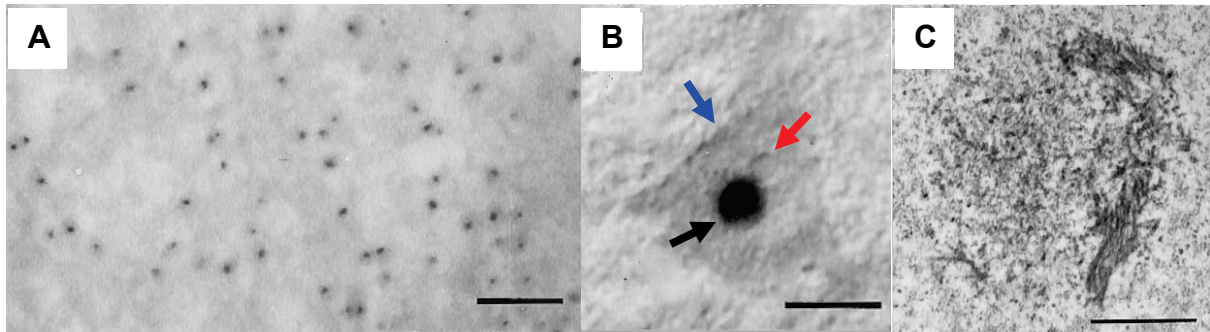


Figure 5: Huntingtin immunoreactivity in post-mortem HD brain. (A) Cortex with labeled inclusions in the nucleus (NIIs). Scale bar is 50 μm . (B) Cortical pyramidal neurons of the cortex. Blue, black and red arrows respectively indicate the nucleus, the NII and the nucleolus. Scale bar is 10 μm . (C) Serial section of NII containing fibrils organized in random and parallel arrays. Scale bar is 1 μm . From (Difiglia *et al.*, 1997).

4.2.4. Gain of toxic function

In some polyQ diseases, the expansion of the polyQ tract causes a partial loss of function of the causative protein which could account, at least in part, for some of the clinical features specific to each disease. For instance, in SBMA [i.e. the disease associated with polyQ expansion in the androgen receptor (AR)], the patients present mild signs of partial androgen insensitivity. There is however significant evidence that polyQ diseases result principally from a toxic gain of function of the proteins due to the expansion of the polyQ tract. Loss-of-function mutations in the AR gene cause the syndrome of testicular feminization which does not feature any motor neuron disease such as SBMA (Pinsky *et al.*, 1992). In addition, multiple knock-out mouse models of polyglutamine proteins have been created and none of them mimics polyQ diseases phenotypes (Duyao *et al.*, 1995, Matilla *et al.*, 1998, Nasir *et al.*, 1995, Zeitlin *et al.*, 1995).

As shown later in the introduction, the new toxic properties conferred to the proteins by the expansion of the polyQ tract are thought to be related to the aggregation phenomenon.

4.3. PolyQ disease-associated proteins

4.3.1. Expression and localization

The proteins involved in polyQ diseases are widely expressed throughout the brain and in peripheral tissues (Ross, 1997). Despite the wide expression of these proteins, polyQ disease pathology is typically restricted to certain neuronal subtypes in the CNS. This suggests that certain specific cellular conditions exist in vulnerable neurons that may cause the selective cytotoxicity of the proteins.

The causative proteins are targeted to different compartments in the cell and differ in their predominant cellular localization; some are primarily nuclear or cytoplasmic, others are present in both nucleus and cytoplasm and one is associated with the membrane (Table 3) (Hands and Wyttenbach, 2010).

4.3.2. Function(s)

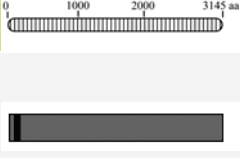
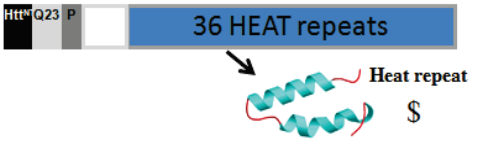

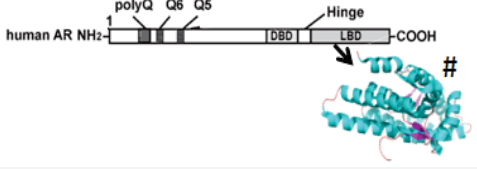


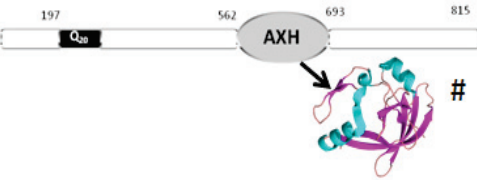


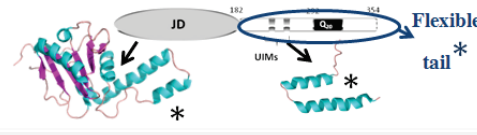


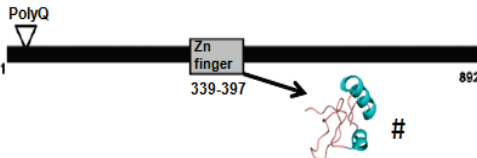
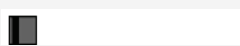
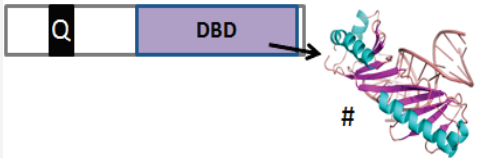
The function of only four of these proteins is rather well established (AR, CACNA1A, TBP and ataxin-3); for the other five proteins, only hypothetical functions have been proposed (Table 3). Although most proteins have a role in transcriptional regulation, there is no function common to all polyQ disease-related proteins.

Table 3: Functions and predominant cellular locations of polyQ disease-associated proteins (modified from (Hands and Wyttenbach, 2010))

Disease	Protein	Known functions Potential functions	Predominant cellular localization
HD	Htt	Regulation of transcription, scaffolding protein, vesicle trafficking, axonal transport, postsynaptic and calcium signalling and bioenergetic homeostasis	Cytoplasmic
SBMA	AR	Nuclear hormone receptor, modulates gene transcription	Cytoplasmic and nuclear
DRPLA	Atrophin-1	Transcriptional co-repressor, assembling signalling complexes at the synapse	Cytoplasmic and nuclear
SCA1	Ataxin-1	RNA processing and transcriptional repression, modulates protein folding	Nuclear
SCA2	Ataxin-2	RNA metabolism, the regulation of apoptosis and actin-polymerisation	Cytoplasmic
SCA3	Ataxin-3	Degradation of ubiquitinated proteins	Cytoplasmic
SCA6	CACNA1A	Subunit of a voltage-dependent calcium channel	Transmembrane
SCA7	Ataxin-7	Regulation of transcription	Nuclear
SCA17	TBP	Regulation of transcription	Nuclear

HD, Huntington's disease; **SBMA**, spinobulbar muscular atrophy; **DRPLA**, Dentatorubral-pallidoluysian atrophy; **SCA**, spinocerebellar ataxia; **Htt**, huntingtin; **AR**, androgen receptor; **CACNA1A**, α -subunit of a voltage-dependent calcium channel; **TBP**, TATA-box binding protein.

Table 4: Size, location of the polyQ tract and structural data available for the polyQ disease-associated proteins

Protein	Size ⁽¹⁾ (a.a.)	Localization of the polyQ tract (■) ⁽²⁾	Structural data ⁽³⁾	References ⁽⁴⁾
Htt	3144 (23Q)			(a)
AR	920 (23Q)			(b)
Atrophin-1	1191 (20Q)		-----	
Ataxin-1	816 (12Q)			(c)
Ataxin-2	1313 (23Q)		-----	
Ataxin-3	370 (19Q)			(d)
CACNA1A	2511 (13Q)		-----	
Ataxin-7	892 (10Q)			(e)
TBP	339 (38Q)			(f)

⁽¹⁾ Number of amino acids in the polyQ disease-associated proteins when the polyQ tract is composed of a number of glutamines equal to that indicated in brackets. From (Robertson *et al.*, 2011a).

⁽²⁾ From (Saunders and Bottomley, 2009).

⁽³⁾ Not to scale. # and * indicate the structures determined by X-ray diffraction and NMR, respectively. \$ indicates predicted secondary structures.

⁽⁴⁾ (a): (Li *et al.*, 2006); (b): (Pereira De Jesus-Tran *et al.*, 2006); (c): (Chen *et al.*, 2004); (d): (Nicastro *et al.*, 2005), PDB ID 2KLZ; (e): PDB ID 2KKR; (f): (Nikolov *et al.*, 1996).

4.3.3. Size and structure

With the exception of ataxin-3 and TBP, the polyQ disease-related proteins have sizes ranging from 800 to >3000 residues (Table 4) (Robertson *et al.*, 2011a) and thus, have high molecular weights. The nine proteins show no sequence homology apart from the polyQ tract which is located at a different position in each protein. Domains within protein have been identified using the Pfam database (Robertson *et al.*, 2011b) and most of the causative proteins are made of several domains. The structures of only some of these domains have been determined by X-ray diffraction or NMR (Table 4). The polyQ tract in the majority of the proteins is located in structurally undefined regions comprised between identified functional domains or N/C-terminal to these domains. In the majority of the disease-related proteins, the polyQ tract is therefore thought to be located in an unstructured and/or flexible region of the protein (Robertson and Bottomley, 2010). Although the amino acid sequences flanking the polyQ tract share no similarity, there is an apparent preference for low sequence complexity (Robertson and Bottomley, 2010).

4.3.4. Role of glutamine repeats in proteins

Homopeptide repeats such as homoglutamine repeats are not an uncommon feature in eukaryotic proteins (Faux *et al.*, 2005). However, it is currently not clear whether polyQ tracts have a specific function in proteins or if they have simply developed over the time-course of evolution at sites where their presence and their length polymorphism are tolerated. The analysis of sequences from polyQ proteins in various organisms shows that the presence of a polyQ tract, its size and its location can vary for the same protein from different species (Schaefer *et al.*, 2012, Wetzel, 2012).

For example, Htt from human possess an N-terminal polyQ tract while its orthologs from *Drosophila* genus possess multiple polyQ sequences that are not located in the N-terminus (Schaefer *et al.*, 2012). In addition, the Htt orthologs from some lower organisms such as *Cyana* contain no polyQ tract.

As another example, while the sequences surrounding the polyQ tract in the TATA-box binding protein from different species (such as *Xenopus*, fish, chicken, rabbit, human,...) are highly conserved, the length of the polyQ tract is highly variable and ranges from 4 to 38Q (Wetzel, 2012).

Taken together, these observations suggest that polyQ tracts might not be essential and that they could act, at least in some cases, as a spacer between structural and/or functional units. However, it was also proposed that polyQ tracts could be involved in protein interactions, since polyQ-containing proteins in human are generally enriched among proteins that are involved in complexes and establish significantly more interactions than proteins without polyQ tract (Schaefer *et al.*, 2012). Moreover, statistical studies have also shown that in several species, proteins containing a polyQ tract are more represented among proteins having functions related to transcriptional regulation and nuclear localization (Harrison, 2006, Alba and Guigo, 2004).

4.3.5. The polyQ tract: the common critical determinant for polyQ diseases

In conclusion, the only known common point between these nine proteins seems to be the polyQ tract suggesting that the expansion of this latter is the critical determinant for polyQ diseases. The central role of polyQ expansions in polyQ diseases is further supported by the study of Ordway *et al.* (Ordway *et al.*, 1997). In this study, a long polyQ tract was inserted into an unrelated murine protein (hypoxanthine phosphoribosyltransferase) and mice expressing this mutant protein develop a phenotype similar to that observed for polyQ diseases such as neurological symptoms, formation of inclusions and premature death.

4.4. The role of protein misfolding and aggregation in polyQ diseases

Dysfunction of various cellular processes such as transcriptional dysregulation, proteasome impairment, unfolded protein response and mitochondrial abnormalities are observed in polyQ diseases (Bauer and Nukina, 2009). It is however difficult to determine precisely which event occurs prior to the others. A large body of evidence indicates however that protein misfolding and aggregation could play an important role in polyQ disease pathogenesis.

The protein quality control machinery is activated in all polyQ diseases. Inclusions in polyQ diseases are ubiquitinated and co-localize with proteasomes indicating that expanded polyQ proteins are targeted for degradation through the ubiquitin proteasome system (UPS) (Cummings *et al.*, 1998, Difiglia *et al.*, 1997, Paulson *et al.*, 1997b, Schmidt *et al.*, 2002, Zander *et al.*, 2001). Several heat shock protein (Hsp) are also found to co-localize with the inclusions (Cummings *et al.*, 1998, Schmidt *et al.*, 2002, Zander *et al.*, 2001). Moreover, several members of the quality control machinery such as Hsp and quality control ubiquitin

ligases were shown to reduce aggregation and/or toxicity of expanded polyQ proteins in different cellular and animal models (Bauer and Nukina, 2009, Williams and Paulson, 2008).

It was proposed that inclusions induce cell death by a recruitment-based mechanism of toxicity. The inclusions could indeed sequester not only components of the quality control machinery but also other important cellular proteins such as transcription factors; when sequestered, the proteins are no longer available to accomplish their function and this could lead to cellular dysfunctions. For example, in addition to proteasome subunits and molecular chaperones already mentioned above, transcription factors such as CREB-binding protein (CBP) have been found in nuclear inclusions formed by several polyglutamine-expanded proteins including Htt, AR, ataxin-1 and atrophin-1 in animal disease models or human brains (Mccampbell *et al.*, 2000, Nucifora *et al.*, 2001, Stenoien *et al.*, 2002).

Although inclusions are usually found in the affected regions of diseased-brains rather than in unaffected (Difiglia *et al.*, 1997, Gourfinkel-An *et al.*, 1998, Paulson *et al.*, 1997b, Skinner *et al.*, 1997), several cellular and mouse models showed that their presence does however not necessarily correlate with cell death (Hodgson *et al.*, 1999, Lunkes and Mandel, 1998, Lunkes *et al.*, 1999, Saudou *et al.*, 1998, Reddy *et al.*, 1998). It was therefore proposed that the formation of inclusions is rather a protective cellular process to sequester misfolded species than the toxic culprit (Arrasate *et al.*, 2004). On the other hand, many studies support the view that as in alzheimer's diseases, the putative toxic species are not the final amyloid fibrils but rather the other oligomeric species formed upstream in the process of fibrillization (Olshina *et al.*, 2010, Takahashi *et al.*, 2008). For instance, in cells expressing model polyQ fusion proteins, soluble oligomeric species were detected by fluorescence resonance energy transfer experiment and their appearance correlates with toxicity (Takahashi *et al.*, 2008). Furthermore, oligomers appearing prior symptom onset and inclusion formation were isolated in mouse model of SBMA (Li *et al.*, 2007). Finally, toxicity of polyQ proteins was also attributed to the presence of misfolded monomeric conformer formed prior aggregation (Nagai *et al.*, 2007).

The investigation of the mechanism by which expanded polyQ tracts trigger the aggregation of protein has therefore been the subject of intensive research; the results of those are summarized in the next section.

4.5. Aggregation mechanism of expanded-polyQ proteins

4.5.1. PolyQ peptides, the simplest models to investigate polyQ-mediated aggregation

Since the only apparent common point between the nine polyQ disease-associated proteins is the presence of the expanded polyQ tract, many studies have been carried out with simple polyQ peptides in order to better understand the aggregation mechanism of expanded polyQ proteins (Chen *et al.*, 2001, Chen *et al.*, 2002, Kar *et al.*, 2011, Perutz *et al.*, 1994).

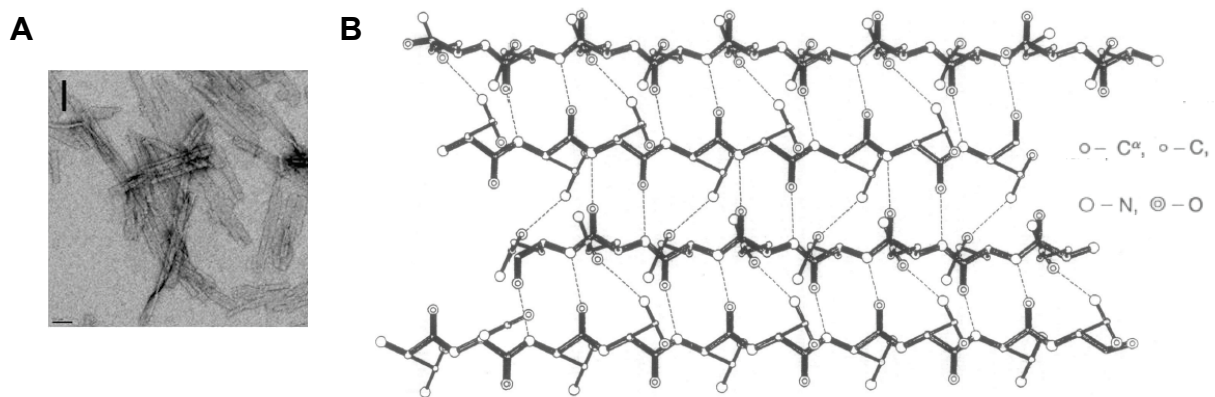


Figure 6: (A) Transmission electron micrograph images obtained for the aggregates formed by $K_2Q_{18}K_2$. From (Kar *et al.*, 2011). (B) Model of homopolymeric glutamine peptide forming “polar zipper”. From (Perutz *et al.*, 1994).

Simple polyQ peptides have an intrinsic tendency to aggregate into amyloid-like fibrils (Figure 6A). The sidechain homogeneity of the consecutive polar glutamines has lead Max Perutz to built an atomic model in which he showed that polyQ peptides could form a structure called “polar zipper” in which antiparallel β -strands are stabilized by hydrogen bonds between both main-chain and side-chain amides (Figure 6B) (Perutz *et al.*, 1994). Max Perutz observed by circular dichroism that a polyQ peptide containing 15 glutamines ($D_2Q_{15}K_2$) adopts a β -structure prior to aggregate, suggesting that the monomers form hydrogen-bonded hairpin. On the basis of a preliminary NMR study showing that 10 glutamines inserted in or replacing the inhibitory loop of the chymotrypsin inhibitor 2 has a random coil structure, Perutz postulated that in proteins, shorter polyQ tracts have a random coil conformation, whereas longer repeats form stable β -hairpin that self-associate to form tightly packed β -sheet.

Chen *et al.* have shown that the rate of aggregation of polyQ peptides increases with the number of glutamines (Figure 7) (Chen *et al.*, 2001). This relationship between the aggregation rate and polyQ length has been correlated to the anticipation phenomenon

observed in polyQ diseases (i.e. the longer the polyQ the more severe and premature the disease) to support the idea that aggregation is an early key event in polyQ diseases

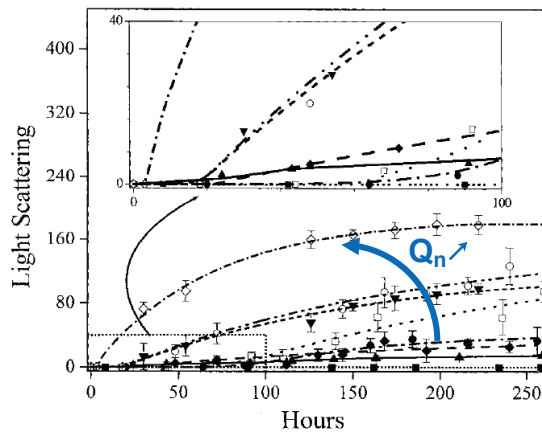


Figure 7: kinetics of aggregation of polyQ peptides $K_2Q_nK_2$ with $n= 15$ (■·····), 20 (▲—), 25(●—), 28 (◆— —), 32 (□· · · ·), 37 (▼---), 41 (○—) and 44 (◇---) (Chen *et al.*, 2001).

Finally, detailed kinetic analysis revealed that the aggregation of polyQ peptides $K_2Q_nK_2$ into amyloid-like fibrils follows a nucleation-polymerization mechanism which does not involve the formation of oligomeric or protofibrillar stable intermediates (Bhattacharyya *et al.*, 2005, Chen *et al.*, 2002).

4.5.2. Influence of protein context

Although the aggregation of polyQ proteins critically depends on the expansion of the polyQ stretch above the pathological threshold, there is evidence that simple polyQ peptides are too simple model to study their aggregation properties. First, it was observed that the minimal aggregation threshold of polyQ peptides is significantly lower ($\geq 15Q$) than that observed for the proteins associated with diseases: this clearly indicates that the presence of surrounding sequences has an influence on the aggregating properties of polyQ sequences (Perutz *et al.*, 1994, Sharma *et al.*, 1999). The effects of surrounding sequences was also evidenced by Bhattacharrya and co-workers who showed that the addition of a polyproline sequence at the C-terminal end of polyQ peptides decreases both their rate of *in vitro* aggregation and the stability of the resulting aggregates without changing their aggregation mechanism (Bhattacharyya *et al.*, 2006). Such polyproline sequence C-terminal to the polyQ tract is present in huntingtin (Htt), the protein associated with Huntington's disease and was shown to influence the aggregation and toxicity of the huntingtin exon 1 in yeast model (Dehay and Bertolotti, 2006, Duennwald *et al.*, 2006).

Other studies have reported the importance of sequences flanking the polyQ tract in modulating the *in vivo* aggregation propensity of polyQ proteins (De Chiara *et al.*, 2005b, Nozaki *et al.*, 2001). Nozaki *et al.* showed that GFP-fusion proteins containing a polyQ length above the pathological threshold with at both sides a fixed number (8 or 9) of flanking amino acid from Ataxin-2, Htt, atrophin-1 or ataxin-3 have indeed different aggregation propensity in a cellular model (Nozaki *et al.*, 2001). Moreover, the authors show that mutating four hydrophobic residues in the upstream polyQ flanking sequence of ataxin-2 to charged amino acids decreased the rate of aggregation, further highlighting the impact of sequence specificity upon aggregation.

Interestingly, the deletion of the AXH domain of the ataxin-1 or its replacement by the homologous domain from the transcription factor HBP1 reduces the *in vivo* aggregation propensity of ataxin-1 although this domain is distant from the polyQ tract by ~350 residues (De Chiara *et al.*, 2005b).

Taken together, these observations indicate that the polyQ expansion is not the only factor determining the aggregation behavior of polyQ proteins and that it is essential to investigate the mechanism of aggregation of the polyQ disease-related proteins. It was proposed that the expansion of the polyQ tract would destabilize and/or cause structural perturbations of the proteins, facilitating and/or leading to the formation of partially unfolded species. Such species generally expose at least part of their main chain and hydrophobic residues to solvent and are therefore prone to intermolecular interactions leading to fibril formation. Such a mechanism has been described for other proteins associated with amyloidoses, including transthyretin (Johnson *et al.*, 2005) and lysozyme (Dumoulin *et al.*, 2003, Dumoulin *et al.*, 2006). Effects of polyQ expansion on the structure, stability and dynamic of the surrounding domains in proteins must therefore be investigated.

Unfortunately, the majority of proteins involved in polyQ diseases are difficult to produce and to handle mainly due to their large size and insoluble character. Thus, *in vitro* studies were carried out with very few polyQ-disease proteins. As a consequence, the knowledge of the aggregation mechanism of polyQ disease-associated proteins concerns essentially ataxin-3 which is one of the smallest polyQ disease-associated proteins, the huntingtin exon 1 and to a lesser extent, ataxin-1. Since the common critical determinant for polyQ diseases is the polyQ tract, an alternative strategy in order to better understand the mechanism of aggregation of polyQ-containing proteins consists in designing and characterizing a series of model polyQ

proteins based on a series of protein scaffolds. This approach is also justified biologically by the previously mentioned study of Ordway *et al.* in which the expression of non disease-related protein containing a long polyQ tract induces a phenotype close to that of polyQ diseases in mouse (Ordway *et al.*, 1997). The next pages summarize the results obtained for both disease-associated and model polyQ proteins.

4.5.3. Aggregation of proteins associated with polyQ diseases

4.5.3.1. Ataxin 3

Ataxin-3 (Atx-3) is a 42 kDa protein associated with SCA3; it is the only full-length causative protein that has been extensively characterized *in vitro*. This ubiquitously expressed protein is predominantly cytoplasmic but is also found in the nucleus (Paulson *et al.*, 1997a).

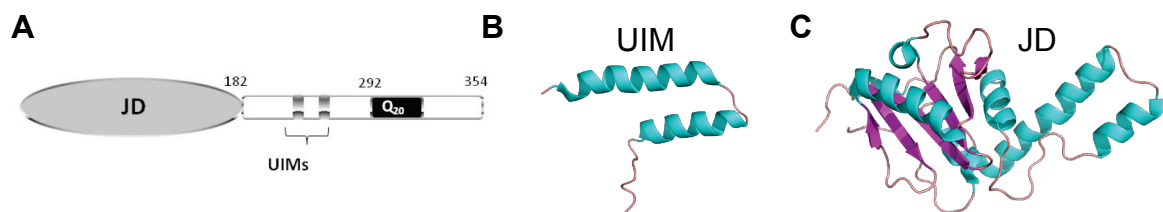


Figure 8: (A) Domain architecture of ataxin-3 with 20 glutamines (Q20). JD, Josephin domain. UIM, ubiquitin interacting motifs [Modified from (Papaleo and Invernizzi, 2011)]. Structures of the tandem UIM domain (B) [pdb ID: 2KLZ, (Song *et al.*, 2010)] and the JD domain (C) [pdb ID: 1YZB, (Nicastro *et al.*, 2005)] of ataxin-3 obtained by NMR.

It is a de-ubiquitinating enzyme which consists in a folded globular N-terminal Josephin domain (JD) which is responsible for the ubiquitin hydrolase activity and a C-terminal region including two ubiquitin interacting motifs (UIM) and the polyQ tract (Masino *et al.*, 2003, Nicastro *et al.*, 2009, Nicastro *et al.*, 2005) (Figure 8A). The Josephin domain is a highly conserved domain across a number of species including mammals, nematodes, insects and plants (Albrecht *et al.*, 2003). Structures of the Josephin domain and the tandem UIM were obtained by NMR and are shown in Figure 8B and C. The manifestation of SCA3 is observed when the polyQ tract in ataxin-3 has a length of at least 45 glutamines (Table 2).

Although several groups have described the production of Atx-3 with up to 64Q in *E. coli* as a soluble protein (Chow *et al.*, 2004, Ellisdon *et al.*, 2006, Natalello *et al.*, 2011), others have produced Atx-3 with up to 78Q in fusion with GST or MBP (Bevivino and Loll, 2001, Masino *et al.*, 2003).

Structural characterization of non-expanded ataxin-3 (with 18 and 26Q) by NMR experiments and limited proteolysis has shown that the region C-terminal to the globular Josephin domain, which comprises the polyQ tract and the two helical UIM, is a flexible tail (Masino *et al.*, 2003). In addition, the selective ^{15}N -labelling of glutamines have allowed a direct insight by NMR into the structure of the polyQ tract. The data obtained suggested that the non-expanded polyQ tract is flexible and exposed to the solvent and argued against a β -conformation stabilized by interaction across glutamine side-chains. Bevivino *et al.* found using CD measurements that the increase of the polyQ length in Atx-3 from non-pathogenic (27Q) to pathogenic length (78Q) was accompanied by loss of α -helical content and a gain of random coil (Bevivino and Loll, 2001). On the opposite, also based on the far-UV CD spectra, Chow *et al.* reported no major differences in secondary structure content between non-expanded (15 and 28Q) and expanded (50Q) Atx-3 (Chow *et al.*, 2004). These results are difficult to compare because (i) in one case (Bevivino *et al.*), Atx-3 was fused to a carrier protein (MBP) and in the other case (Chow *et al.*), it was produced alone in a soluble form and (ii) they used different pathological polyQ lengths. Thus, the effects of polyQ expansion on the structure of the ataxin-3 are still unclear.

The high propensity of ataxin-3 to aggregate upon unfolding has prevented a quantitative determination of the effects of polyQ expansion on protein stability (Chow *et al.*, 2004). The qualitative results reported so far however suggest that ataxin-3 with an expanded polyQ tract (50Q) is not destabilized compared to ataxin-3 with normal polyQ length (i.e. 15 and 28Q).

The aggregation properties of ataxin-3 with polyQ tracts of different lengths were thoroughly studied (Ellisdon *et al.*, 2007, Ellisdon *et al.*, 2006, Natalello *et al.*, 2011, Saunders *et al.*, 2011). Expanded ataxin-3 (with 64Q) aggregates *in vitro* according to a two-stage pathway (Figure 9, expanded) (Ellisdon *et al.*, 2007, Ellisdon *et al.*, 2006). The first step of aggregation is mediated by intermolecular interactions between the Josephin domain, involved a structural change of this domain and leads to the formation of β -rich SDS-soluble protofibrils. The second step is mediated by polyQ interactions and leads to the formation of SDS-insoluble large fibrils whose core seems stabilized by hydrogen bonds between both side-chain and main-chain amides (Natalello *et al.*, 2011). Surprisingly, non-expanded ataxin-3 (with 15Q or with a QHQ motif) also aggregates *in vitro* and forms SDS-soluble protofibrils whose formation is mediated by the common Josephin domain (Figure 9, non expanded). However, the second aggregation step which is characterized by the formation of large SDS-insoluble

fibrils is not observed; this suggests that the presence of an expanded polyQ tract ($\geq 45Q$) is required for the ataxin-3 to form SDS-insoluble fibrils.

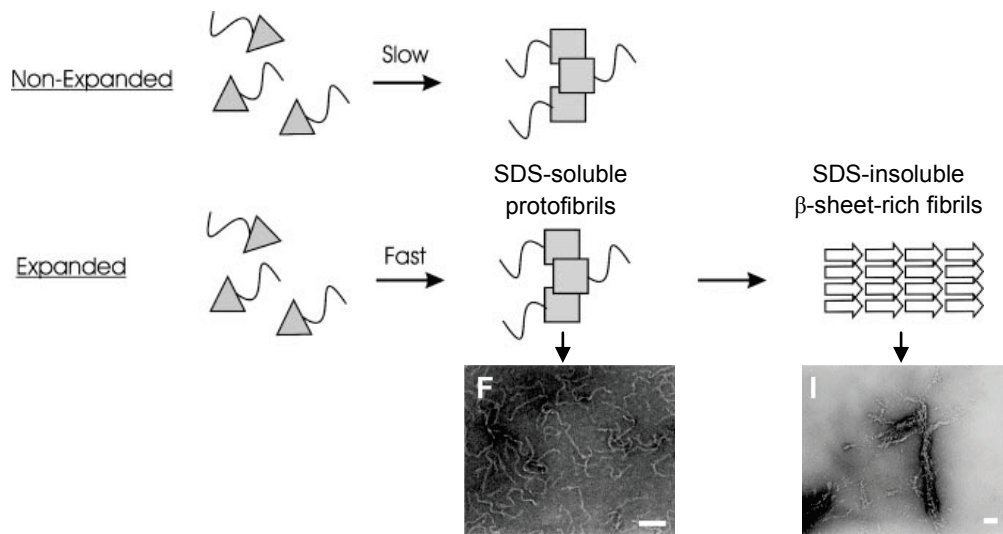


Figure 9: Aggregation mechanism of ataxin-3 with normal and expanded polyQ length [adapted from (Ellisdon *et al.*, 2006)]. The Josephin domain and the ubiquitin-interacting motifs domain are represented as a triangle with the C-terminal polyQ tract as a tail.

Interestingly, although the formation of SDS-soluble protofibrils occurs with all ataxin-3 variants (i.e. whatever the length of the polyQ tract), it is faster when the protein contains an expanded polyQ tract. Thus, although the polyQ region is not directly involved in the first step of aggregation, its expansion does indirectly modulate the aggregation propensity of the non-polyQ region through a mechanism that is yet to be fully understood.

It was unexpected that ataxin-3 with a non-pathological polyQ length has an intrinsic tendency to aggregate *in vitro* since it does not aggregate *in vivo*. Thus, it has been proposed that the interaction of ataxin-3 with binding partners may stabilize it and/or limit its conformational flexibility so that it reduces its aggregation propensity (Masino *et al.*, 2004). In support of this hypothesis, a recent study has shown that stabilization of the Josephin domain by mutations decreases the rate of *in vitro* formation of the SDS-soluble protofibrils by ataxin-3 (Saunders *et al.*, 2011).

4.5.3.2. Huntingtin

Huntingtin (Htt) is a large ubiquitously expressed protein (~340 kDa, 3144 aa). It is a primarily cytosolic protein which can however shuttle between the cytoplasm and the nucleus (Hands and Wyttenbach, 2010). Its structure and exact function are unknown. However, it was proposed that:

- (i) Htt is involved in transcription regulation,
- (ii) as a scaffolding protein, Htt mediates many protein-protein interactions and plays a role in vesicle interactions and fast axonal transport (Difiglia *et al.*, 1995, Gauthier *et al.*, 2004, Gunawardena *et al.*, 2003).
- (iii) Htt could regulate calcium signaling and bioenergetic homeostasis (Lee *et al.*, 2007).

Htt is associated with Huntington's disease (HD) when it contains a polyQ tract with more than 35 glutamines (Table 2). No *in vitro* investigation have been carried out with full-length Htt, probably because (i) of its large size, (ii) it was shown that the expression of its exon 1 (Htt exon 1) with an expanded polyQ tract is sufficient to induce HD phenotype in mouse model (Mangiarini *et al.*, 1996) and (iii) there is growing evidence that proteolytic cleavages of Htt releasing fragments containing the exon 1 is required for its aggregation and toxicity in HD disease (Cooper *et al.*, 1998, Graham *et al.*, 2006, Martindale *et al.*, 1998). Thus, all the studies have aimed to characterize the structural and aggregation properties of Htt exon-1 rather than those of the full-length protein.

Htt exon 1 is composed of a N-terminal segment of 17 amino acids followed by the polymorphic polyQ tract, a homoproline sequence (11 prolines) and a proline/glutamine-rich sequence (Figure 10).

MATLEKLMKAFESLKSF (Q)_N P₁₁ QLPQPPPQAQPLLQPQPPPPPPPPPPGPAVAEEPPLKR

Figure 10: Amino acid sequence of the exon 1 of huntingtin.

To increase its solubility, Htt exon 1 has been generally produced and characterized in fusion through its N-terminus to a carrier protein, such as glutathione S-transferase (GST) (Hollenbach *et al.*, 1999, Legleiter *et al.*, 2010, Scherzinger *et al.*, 1997, Scherzinger *et al.*, 1999), thioredoxin (Bennett *et al.*, 2002, Peters-Libeu *et al.*, 2012) or maltose-binding protein (MBP) (Kim *et al.*, 2009, Peters-Libeu *et al.*, 2012, Poirier *et al.*, 2002).

Two major studies have been performed to characterize the structure of the Htt exon 1. First, Bennett *et al.* have characterized the structural properties of Htt exon 1 fused to thioredoxin and containing polyQ lengths below and above the pathological threshold of HD (Bennett *et al.*, 2002). The data they obtained by CD and NMR suggest that Htt exon 1 with the different polyQ lengths are random coil. They do however not exclude that the polyproline region C-terminal to the polyQ tract can form a polyproline helix. In the second study, Kim *et al.* have

crystallized a fragment of Htt exon 1 containing 17Q (a non-pathological polyQ length) and fused to the C-terminus of MBP and have determined its structure by X-ray diffraction (Figure 11) (Kim *et al.*, 2009). The X-ray structure of this fragment of Htt exon 1 consists in an amino-terminal α -helix for the first 17 residues and a polyproline helix in the polyproline region which propagates in the first residues of the proline/glutamine-rich region (Figure 11). The 17Q in this exon adopt multiple conformations including α -helix, random coil and extended loop, indicating that its structure is influenced by the conformation of the neighboring protein regions.

These two studies principally diverge in the structure adopted by the 17 N-terminal residues of Htt exon 1 which is α -helical in one or mainly disordered in the other. These differences could come from the method used to investigate the structural properties of Htt exon 1 but also from the different fusion partners used to increase its solubility in these studies.

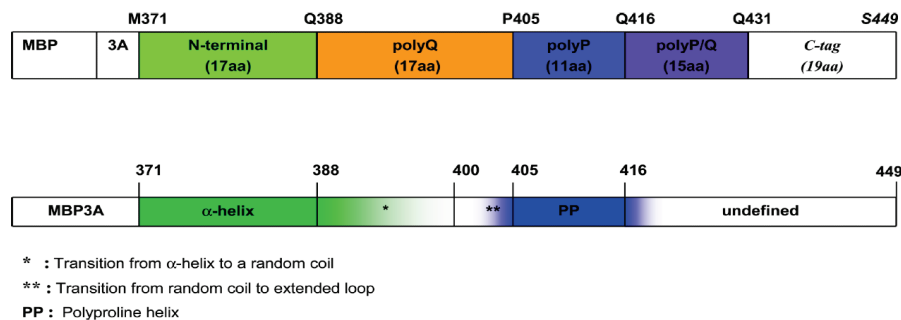


Figure 11: Schematic diagram of the secondary structure elements of Htt exon 1 fragment with 17Q. The first bar represents the sequence of the fusion protein MBP-Htt exon 1 fragment containing 17Q which consists of MBP protein (not shown to scale), 3 Alanines linker (3A), 17 aa long N-terminal region of Htt (green), polyQ region with 17 glutamine residues (poly17Q, orange), polyproline region with 11 proline residues (poly11P, blue), 15 aa long mixed polyglutamine and polyproline region (polyP/Q, purple), and C-terminal tag of 19 aa (white). The second bar is a summary of structural information obtained from analysis of Htt17Q-EX1 structures resolved in the seven crystals. The main secondary structure elements of Htt17Q-EX1 are shown: α -helix (bold green), transition from α -helix to random coil region (shaded green), random coil (uncoloured), transition to extended loop region (shaded blue), polyproline helix (bold blue), transition from polyproline helix to unstructured region (shaded blue). From (Kim *et al.*, 2009).

The structural features of the polyQ tract in monomeric Htt exon 1 were also investigated using polyQ specific monoclonal antibodies such as MW1 and 3B5H10. These polyQ specific antibodies were shown to bind preferentially to expanded polyQ tracts (Ko *et al.*, 2001, Peters-Libeu *et al.*, 2005), further supporting that expanded polyQ tract could adopt a specific aggregation-prone conformation as proposed by Perutz.

Based on affinities and stoichiometry measurements, Bennett *et al.*, have however proposed that the antibody MW1 antibody owes its preferential binding to long polyQ tracts to a ‘linear lattice’ effect of the polyQ sequences (Bennett *et al.*, 2002). However, the recent determination of the structure of the anti-polyQ antibody 3B5H10 indicates that the paratope

of this antibody forms a groove suitable for two β -rich polyQ strands (Peters-Libeu *et al.*, 2012). Using small-angle X-ray scattering, it was shown that 3B5H10 recognizes a compact two-stranded conformation of the polyQ in monomeric Htt that emerges when the polyQ stretch expands (Peters-Libeu *et al.*, 2012). These results suggest that the expanded polyQ tract can adopt different conformations in monomeric Htt.

Proteins made of GST fused to Htt exon 1 containing 20, 30 and 51 glutamines remain soluble while those fused to 83 and 122Q readily form high molecular weight SDS insoluble aggregates (Scherzinger *et al.*, 1997). The critical length of the polyQ stretch in the fusion proteins leading to aggregation is therefore greater than 51Q. However, removal of the GST moiety from the fusion proteins results in the formation of insoluble amyloid-like fibrils by Htt exon 1 when it contains 51 and more glutamines (Scherzinger *et al.*, 1997). Further characterization of more polyQ lengths between 20 and 51Q (27, 32, 37, 39, 40, 42 and 45Q) show that only Htt exon 1 with 37 or more glutamines readily aggregates into insoluble fibrils after removal of the GST moiety (Scherzinger *et al.*, 1999). These results suggest that the polyQ threshold length for amyloid-like fibril formation by Htt exon-1 alone is comprised between 32 and 37Q and is similar to the pathological threshold observed in HD (> 35Q). Above the threshold polyQ length, the rate of amyloid-like fibril formation of Htt exon 1 increases with the polyQ length (Legleiter *et al.*, 2010).

The aggregation into amyloid-like fibrils of htt exon 1 was shown to occur, both *in vitro* and *in vivo*, through the formation of intermediate oligomeric species such as spherical oligomers or protofibrils (Legleiter *et al.*, 2010, Poirier *et al.*, 2002). These findings indicate that the mechanism of fibril formation of Htt exon 1 deviates from that of simple polyQ peptides which do not form any stable intermediate species (Bhattacharyya *et al.*, 2005, Chen *et al.*, 2002). Thus, they suggest that flanking sequences modulate the aggregation properties of polyQ sequences.

As reported above, the GST-Htt exon 1 fusion protein with 51Q does not self-aggregate unless the GST moiety is cleaved off (Scherzinger *et al.*, 1997). Interestingly, GST-Htt exon 1 fusion protein with 51Q in which most of the proline-rich region C-terminal to the polyQ tracts (Figure 10) is lacking forms amyloid like fibrils without prior proteolytic cleavage of the GST moiety (Hollenbach *et al.*, 1999). These results have first indicated that the proline-rich region C-terminal to the polyQ tract has a suppressive effect on Htt exon 1 aggregation and have triggered a range of experiments aiming at understanding how sequences flanking the polyQ tract in Htt exon 1 influence its aggregation behavior. These experiments have been

principally performed with synthetic polyQ peptides corresponding to fragment of Htt exon 1. In these peptides, the polyQ sequence was flanked by either:

- the 17 N-terminal residues of Htt exon 1 at its N-terminus (green sequence in Figure 10 and referred to as htt-NT17) (Thakur *et al.*, 2009),
- a homoproline sequence at its C-terminus (10 prolines maximum), referred to as polyP sequence (Bhattacharyya *et al.*, 2006, Darnell *et al.*, 2007) or
- both the htt-NT17 and the polyP sequences (Thakur *et al.*, 2009).

Studies on these synthetic polyQ peptides have indicated that:

- the presence of a polyproline region directly C-terminal to the polyQ sequence in Htt exon 1 could have a suppressive effect on its aggregation by disfavoring aggregation-competent conformations of the polyQ tract due to the formation of a constraining polyproline helix in the polyP sequence (Bhattacharyya *et al.*, 2006, Darnell *et al.*, 2007) This is in good agreement with the X-ray data discussed above (Kim *et al.*, 2009).
- the Htt-NT17 sequence fused to polyQ length below and above the pathological threshold for HD has no stable secondary structure, which is in agreement with the study of Bennett *et al.* on Htt exon 1 (Bennett *et al.*, 2002). They have also shown that the Htt-NT17 segment has an intrinsic tendency to collapse in an aggregation-resistant compact coil structure but an attached expanded polyQ sequence induces in this segment a more extended state. When the Htt-NT17 segments are extended, they can interact and initiate the aggregation process of the synthetic polyQ peptides with expanded polyQ (Thakur *et al.*, 2009). The Htt-NT17 segments enhance the aggregation rate of the synthetic polyQ peptides and this enhancing effect is dominant over the suppressive effect mediated by the polyP.

Based on the mechanism of aggregation of synthetic polyQ peptides, it was proposed that Htt exon 1 carrying an expanded polyQ tract ($\geq 35Q$) could form fibrils via a two-step mechanism, similar to that described for the expanded ataxin-3 but with some variations (Figure 12). In the model, the aggregation is initiated by intermolecular interaction between the Htt-NT17 segments leading to the formation of a small amount of stable spherical oligomers. The cores of these oligomers comprise the Htt-NT17 segments but not the polyQ tracts. Then, a hypothetical nucleation event occurring within these oligomers initiates the formation of amyloid fibrils whose core is essentially made by the polyQ tract (Figure 12). This model is

consistent with the fact that spherical oligomers are observed on the aggregation pathway of Htt exon 1 both *in vitro* and *in vivo* (see above).

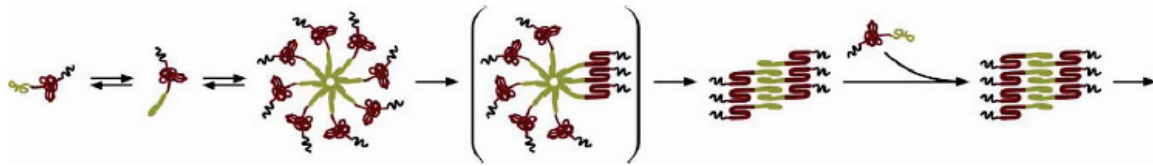


Figure 12: Representation of the aggregation mechanism of Htt exon 1 initiated by the NT segments. This drawing is schematic and is not meant to imply any details of aggregates [adapted from (Thakur *et al.*, 2009)]. The HT-NT17 segment is shown in green and the polyQ tract is in red and the polyproline sequence is in black.

4.5.3.3. Ataxin-1

Ataxin-1 is a 816-residue protein which is primarily located in the nucleus. This is an RNA-binding protein which is thought to act as a transcriptional repressor (Irwin *et al.*, 2005, Tsai *et al.*, 2004, Yue *et al.*, 2001). Ataxin-1 comprises an AXH domain (spanning residues 562-693) which is quite distant from the polyQ tract (starting at position 197) (Table 4). The X-ray structure of the C-terminal AXH domain of ataxin-1 has been obtained (Chen *et al.*, 2004). This is a highly dynamic chameleon-like domain which contains an OB fold (α/β), a structural motif found in many oligonucleotide-binding proteins.

Expansion of the polyQ tract above 38 glutamines in ataxin-1 leads to SCA1. Although the aggregation mechanism of ataxin-1 has not been extensively investigated *in vitro*, there is evidence that it could also aggregate via a two-stage mechanism. First, the AXH domain can spontaneously aggregate when isolated (De Chiara *et al.*, 2005a, De Chiara *et al.*, 2005b). Secondly, its deletion or replacement with a homologous non-aggregation-prone domain is associated with a decrease in the rate of formation of nuclear inclusions in a cell model (De Chiara *et al.*, 2005a, De Chiara *et al.*, 2005b). Thus, similar to the Josephin domain of ataxin-3 and to the N-terminus of Htt exon 1, the AXH domain has a rate-modulatory role on ataxin-1 aggregation.

The studies on the three polyQ disease-related proteins or protein fragments clearly indicate that the aggregation properties of polyQ-containing proteins results from a very complex interplay between the polyQ tract and non-polyQ regions. There is experimental evidence that the ataxin-3 and Htt exon-1 aggregates via a complex multi-domain mechanism initiated by intermolecular interactions within non polyQ regions. The polyQ tract only involves in later stages for the formation of the core of mature fibrils. Such multi-domain mechanism of

aggregation was also suspected for ataxin-1. It remains however to find if such a mechanism is relevant to the other 6 proteins associated with diseases.

4.5.4. Creation and characterization of model proteins

A series of model polyQ proteins made of a well-characterized soluble protein (scaffold) fused to a polyQ tract or in which glutamine repeats are inserted have been designed and characterized (Table 5). The structure, stability and folding of these protein scaffolds have generally been characterized in detail, providing a solid foundation to interpret any changes caused by the polyQ tract. In some of the model proteins, the glutamines are flanked by several residues from the proteins involved in polyQ diseases. Five protein scaffolds have been widely used so far: glutathione S-transferase (GST) (Bulone *et al.*, 2006, Masino *et al.*, 2002), myoglobin (Tanaka *et al.*, 2003, Tanaka *et al.*, 2001) and its apoform (Tobelman and Murphy, 2011), thioredoxin (Nagai *et al.*, 2007, Nagai *et al.*, 2000), the B domain of staphylococcal protein A (SpA) (Robertson *et al.*, 2011b, Robertson *et al.*, 2008) and cellular retinoic acid binding protein I (CRABPI) (Ignatova and Gierasch, 2006, Ignatova *et al.*, 2007) (Table 5). Their size varies from 7 kDa (SpA) to 25 kDa (GST). Their structure is well established; they are either all- α , α/β protein or β -barrel protein (Table 5). The main results of the characterization of these models proteins are summarized in the next paragraphs.

Table 5: Model polyQ proteins

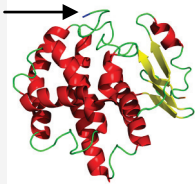
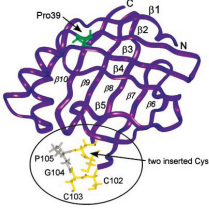
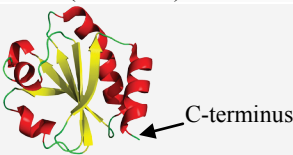

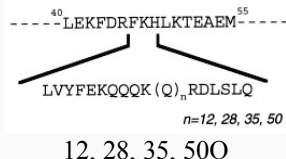
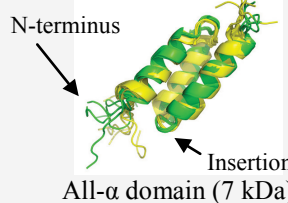
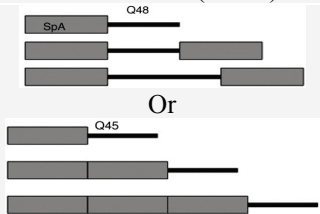
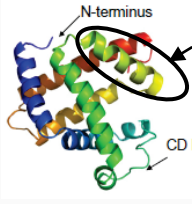
FUSION			
Host protein	Structure	PolyQ length	Ref. ⁽¹⁾
GST (<i>Schistosoma Japonicum</i>)	 α/β protein (25.5 kDa)	<u>C-terminal</u> 22 and 41Q	(a)
CRABPI (Mouse)	 β -barrel (15.5 kDa)	Htt exon 1 fused to both N- and C-terminus <u>C-terminal</u> 20, 33, 40, 53, 64Q <u>N-terminal</u> 20, 53Q	(b)
Thioredoxin (<i>E. coli</i>)	 α/β protein (11.7 kDa)	<u>C-terminal</u> 19, 35, 45, 62, 81Q	(c)

Table 5: Model polyQ proteins (continued)

INSERTION			
Host protein	Structure	PolyQ length	Ref. ⁽¹⁾
Myoglobin (Sperm whale)	<p>CD loop</p>  <p>All-α protein (17 kDa)</p>	 <p>12, 28, 35, 50Q</p>	(d)
	INSERTION AND FUSION		
Host protein	Structure	PolyQ length	Ref. ⁽¹⁾
B domain of staphylococcal protein A (<i>S. aureus</i>)	<p>N-terminus</p>  <p>All-α domain (7 kDa)</p>	<p><u>Insertion</u> 9, 25Q</p> <p><u>N-terminal fusion</u> 5, 9, 13, 20, 24, 31, 35, 40, 45, 48, 52Q</p>	(e)
	 <p>Or</p> <p><u>Interdomain</u> 48, 64Q</p> <p>Or</p> <p><u>C-terminal to 1, 2 or 3 domain(s)</u> 45Q</p>	(f)	
Apomyoglobin	 <p>F -Helix is unfolded in apo form</p> <p>Myoglobin structure All-α protein (~17 kDa)</p>	<p><u>Insertion</u> 8, 16, 20, 24, 28, 34, 38, 46Q</p> <p><u>N-terminal fusion</u> 8, 16, 32Q</p>	(g)

⁽¹⁾ (a): (Bulone *et al.*, 2006, Masino *et al.*, 2002); (b): (Ignatova and Gierasch, 2006, Ignatova *et al.*, 2007, Thompson *et al.*, 1995); (c): (Katti *et al.*, 1990, Nagai *et al.*, 2007, Nagai *et al.*, 2000); (d): (Tanaka *et al.*, 2003, Tanaka *et al.*, 2001); (e): (Robertson *et al.*, 2008); (f): (Robertson *et al.*, 2011b); (g): (Tobelmann and Murphy, 2011)

4.5.4.1. Myoglobin-derived polyQ proteins

Sperm whale myoglobin (Mb) was the first scaffold used to investigate the effects of the insertion of polyQ tracts on the structural, thermodynamic and aggregating properties of proteins not associated with polyQ diseases. This scaffold was chosen due to its high stability and high solubility (Tanaka *et al.*, 2001). It is a 153-residue protein (~17 kDa) consisting of eight α -helices that occupy 80% of the total residues and that does not contain any β -sheet structures (Table 5). PolyQ tracts of 12, 28, 35 and 50Q were inserted into the solvent-exposed loop between the C and D helices (Table 5, C-D loop); this loop is the most flexible region of the protein. The polyQ sequence was flanked by 10 and 8 residues from Atx-3 at its N and C-terminus, respectively. The chimeric proteins were produced as soluble proteins in *E*

coli; Mb variants with up to 50Q could be produced in sufficient amounts to be characterized using a range of techniques.

The polyQ tract (≥ 28 Q) forms an intramolecular β -sheet without altering the content in secondary structure of myoglobin. It however perturbs the tertiary structure at the surface of myoglobin and destabilizes the protein in a polyQ length dependent manner (Tanaka *et al.*, 2001). Upon incubation at 37°C, chimera with 50Q (Mb-50Q) readily form fibrillar aggregates with amyloid-like properties (i.e. they bind Congo red, exhibit the characteristic X-ray pattern of cross- β structure) while chimeras with 35Q (Mb-35Q) and 28Q (Mb-28Q) require longer incubation time to form such aggregates. Myoglobin alone and with 12Q (Mb-12Q) did not form any type of aggregates. Significant differences in the early stages of fibrillization of Mb-35Q and Mb-50Q were however observed essentially by SAXS measurements (Tanaka *et al.*, 2003). Mb-50Q forms large non-fibrillar aggregates (80 to ~90 monomers, referred to as quasi-aggregates) in the early stage of fibrillization while Mb-35Q does not. The 50Q tract seems, at least in part, exposed to the solvent in these quasi-aggregates while it is buried in the fibrils. Since the tertiary structure at the surface of myoglobin is partially unfolded in Mb-50Q, it was therefore proposed that the formation of the quasi-aggregate by Mb-50Q is essentially due to the unfolding of the host protein surface, which exposes some hydrophobic residues to the surface. Thus, the aggregation of Mb50Q could occur via a multi-domain process. In Mb-35Q, the myoglobin moiety is less unfolded than in Mb-50Q and thus, hydrophobic residues not or less exposed prevent the formation of quasi aggregates by the Mb moiety.

4.5.4.2. GST-derived polyQ proteins

Glutathione S-transferase is a well characterized 25.5 kDa protein of known structure; it is an α/β protein (Table 5). PolyQ tract of 22 and 41 residues were fused to GST (GST-Q22 and GST-Q41) without altering its secondary structure; they exhibit a random coil conformation as probed by CD and NMR (Masino *et al.*, 2002). DLS measurements have shown that GST alone, GST-Q22 and GST-Q41 exist essentially as a dimer (Bulone *et al.*, 2006). The proteins remain soluble at room temperature and the temperature has to be increased to initiate aggregation. The thermal-induced aggregation of GST, GST-Q22 and GST-Q41 occurs through the conversion of dimers into soluble aggregates and then, the conversion of soluble aggregates to insoluble large aggregates. This process is associated with a structural transition characterized by a significant enrichment in β -content. The two-step aggregation is common

to both the fusion proteins (GST-Q22 and GST-Q44) and GST alone, suggesting that GST which is per se able to undergo irreversible aggregation plays a key role in determining the aggregation pathway of the fusion protein. However, only GST-Q41 finally forms fibrils while GST alone or with 22Q forms amorphous aggregates, indicating that a minimal polyQ length is required to promote the formation of fibrils. This length is in the pathological range observed with the proteins associated with diseases. Thus, the aggregation of GST-Q41 could be defined as a multi-domain process.

4.5.4.3. CRABPI-derived polyQ proteins

CRABPI (cellular retinoic acid binding protein I) is a 136 amino-acid long (~15.5 kDa) member of the intracellular lipid-binding protein family. This murine protein has a β -barrel structure composed of 10 antiparallel β -strands with a short helix-turn-helix between strand 1 and 2 (Table 5) (Thompson *et al.*, 1995). A tetra-Cys sequence was inserted to allow the expressed protein to be labeled in the cell with a low molecular weight membrane permeable fluorescent dye, FIAsh (Table 5). Htt exon I was fused to the C-terminus (with 20, 33, 40, 53 or 64Q) or to the N-terminus (with 20 or 53Q) of CRABP I (Ignatova and Gierasch, 2006). The *in vivo* aggregation of the resulting fusion proteins was followed in *E. coli* expressing system and only those containing 53 and 64Q form SDS-insoluble fibrils in a polyQ length-dependent way (i.e. the propensity to aggregate of the protein with 64Q is higher than that of the protein with 53Q). Chimeras with 20Q and 33Q essentially remain soluble while those with 40Q form SDS-soluble aggregates. The location of the insertion (i.e. N- or C-terminal) did not alter the influence of Htt exon1 on the aggregation propensity of the fusion protein.

Chimeras with 20 and 53Q were produced in *E. coli*, purified and characterized. While the chimeras with 20Q at the N- or C-terminal could be purified from the soluble fractions of cell lysate, the chimera with 53Q were isolated from the insoluble fraction after solubilization with 8 M urea. The fusion of Htt exon I with 20Q does not have any effect on the structure and stability of CRABPI. On the opposite, the fusion of Htt exon 1 containing 53Q significantly destabilizes CRABPI and perturbs both its secondary and tertiary structure. Similar structural perturbations and destabilization were observed whether Htt exon 1 was fused to the N- or C- of CRABPI.

The in-cell aggregation of the chimera made of Htt exon I with 53Q fused to the C-terminus of CRABPI proceeds via a multistep process (Ignatova *et al.*, 2007). The first stage consists in the formation of detergent-soluble spherical oligomers whose core is composed of CRABPI;

in the second stage, detergent-resistant fibrils with a core dominated by the polyQ tract are formed. The polyQ tract in the Htt exon I plays an indirect role in the formation of the early oligomers by destabilizing and structurally perturbing the CRABP I moiety, enhancing its ability to form an aggregation-prone species. This study shows for the first time that the multistep aggregation process described previously *in vitro* for Htt exon 1 and ataxin-3, can also occur *in vivo*.

4.5.4.4. Thioredoxin-derived polyQ proteins

E. coli thioredoxin is a compact 11.7 kDa protein which consists in two conformational domains, $\beta\alpha\beta\alpha\beta$ and $\beta\beta\alpha$ connected by a single-turn of α -helix and a 3/10 helix (Katti *et al.*, 1990). The β -sheets form the core of the molecule and the helices, the external surface (Table 5). It is a highly soluble protein which can be expressed at high concentration in *E. coli* and can be easily purified.

Nagai *et al.* have fused polyQ tracts with 19, 35, 62 and 81Q to the C-terminus of thioredoxin (Nagai *et al.*, 2000). The fusion proteins obtained have an enhanced α -helical content compared to thioredoxin alone and this α -helical content increases with polyQ length suggesting that the polyQ tract adopts a α -helical dominant structure. Only thioredoxin fused to 62 and 81Q aggregates *in vitro* and the aggregation rate increases with polyQ length. The secondary structural changes of thioredoxin-Q62 during the aggregation process were examined and suggest that, at least, the polyQ tract undergoes a conformational transition from an α - to a β -structure in the monomeric state, which precedes the assembly into amyloid-like fibrils (Nagai *et al.*, 2007).

4.5.4.5. SpA-derived polyQ proteins

The B domain of *Staphylococcus aureus* protein A (SpA) is a 7 kDa domain which is composed of a bundle of 3 α -helices (Table 5). In their first study, Robertson *et al.* have fused 5 to 52Q to the C-terminus of SpA domain and have shown that the polyQ tract which adopts a random coil structure does not alter the stability, structure and dynamics of SpA, whatever the length of the polyQ tract is (Robertson *et al.*, 2008). When the polyQ length is ≥ 35 Q, it however promotes the aggregation into fibrils of the SpA-Q_N fusion proteins with an aggregation rate increasing as the polyQ length increases. The authors suggest that it is the intrinsic propensity of the polyQ domain to promote fibril formation that triggers the aggregation of the SpA-Q_N fusion proteins. In order to explore the influence of repeat

location, polyQ sequences were also inserted within a flexible loop between the helices 2 and 3 of SpA (Table 5). No chimera with more than 25Q could be expressed. The insertions of 9Q and 25Q destabilize in a polyQ length-dependent manner SpA; they also affect the tertiary structure of SpA but not its secondary structure. No aggregation was observed with these polyQ lengths.

In order to understand how domain architecture of the protein and the location of the polyQ tract can affect the aggregation of polyQ protein, Robertson *et al.* have also designed a series of proteins where the polyQ tract was fused to several consecutive SpA domains or inserted between two SpA domains (Table 5) (Robertson *et al.*, 2011b). They showed that when a pathological polyQ length (48Q) is present between two SpA domains, it aggregates more slowly than when it is fused to the C-terminus of one SpA domain. These results suggest that the presence of the polyQ tract between two domains may restrict its conformational freedom and may also increase the steric constraints imposed on the polyQ tract. They also demonstrate that increasing the number of consecutive SpA domains fused to a pathological polyQ tract (45Q) decreases the aggregation rate of the polyQ protein. This suggests that the increased molecular weight and/or total electrostatic charge/solubility caused by additional domains is an inhibitory factor of aggregation.

4.5.4.6. Apomyoglobin-derived polyQ proteins

Tobelman *et al.* have generated a small library of apomyoglobin-derived polyQ proteins with a polyQ tract of various lengths and positions (Tobelman and Murphy, 2011). The apo form of sperm whale myoglobin has approximately the same all- α structure than the holo form of the protein with the exception of one α -helix which is disordered in the apo form (Table 5). The group of Murphy has essentially chosen to work with the apoMb because its native fold is less structured and stable than that of the holo form. PolyQ tracts were either fused at the N-terminus or inserted into the CD loop. Mutant proteins with 8, 16 and 32Q fused to the N-terminus of apomyoglobin largely retain the tertiary structure of the wild-type protein but exhibit a modest increase in α -helical content, suggesting that the glutamines may adopt some α -helical structure. The fusion proteins are marginally more resistant to chemical unfolding than the ApoMb. All fusion proteins with 8, 16 and 32Q form ordered fibrillar aggregates in small polyQ-dependent proportion and their formation seems to arise from specific polyQ mediated intermolecular interaction.

Significantly different and complex effects were described when the polyQ tract (8, 16, 38 and 46Q) was inserted in the CD loop:

(i) the insertion perturbs the tertiary structure of ApoMb in a polyQ length-independent manner,

(ii) the insertion of 8Q does not change the secondary structure content of ApoMb, while a tract containing 16-28Q induces a loss of α -helical structure, an increase of β -sheet and a large increase in disordered structures. As the polyQ length further increases (34-46Q), the loss of α -helical structure remain similar but there is a larger increase in β -sheet. This suggests that the polyQ tract along with the flanking residues adopt some β -sheet structure as the polyQ length increases.

(iii) the insertion destabilizes ApoMb and the extent of destabilization is independent of the length of the polyQ tract.

(iv) chimeras with 8, 16 and 38Q readily form soluble chain-like oligomers and the formation of these oligomers seems to be, however, induced by the unfolding of apomyoglobin and to occur independently from an involvement of specific polyQ-polyQ interactions. Indeed, a chimera with a glycine-serine octet repeat (GS₈) instead of a polyQ stretch also form such aggregates.

Taken together, the results obtained in this study indicate that the location of the polyQ tract has a dominant influence on the nature of the effects on the structure and aggregation of the protein, whereas the length of the polyQ tract modulates the extent of these effects.

4.5.5. Conclusions

Two major scenarios, which are not necessarily mutually exclusive, have been initially put forward to explain how expanded polyQ tracts promote the aggregation of proteins.

Firstly, because the only known common point between the nine proteins associated with polyQ diseases is the expanded polyQ tract, it was initially thought that their aggregation behaviour was mainly dictated by this latter. It has been suggested that long polyQ repeats have a high intrinsic propensity to form stable “polar zippers” which consist of antiparallel β -strands stabilized by hydrogen bonds between both main-chain and side-chain amides (Perutz *et al.*, 1994). The formation of such structures was thought to trigger aggregation into amyloid

fibrils. However, a range of stable conformations have been observed for the polyQ tract in the context of a protein, as summarized in Table 6. The variability between these studies suggests that the conformation adopted by the polyQ tract is sensitive to the protein context. However, the fact that the predominant conformation of the expanded polyQ tract (i.e. the polyQ length inducing fibril formation) is not a β -sheet does not exclude that such a conformation are rarely and transiently populated before aggregation. The use of anti-polyQ antibodies further support this hypothesis and it was shown for Htt exon 1 that although NMR and CD experiments suggest that the polyQ tract does not adopt stable conformation involving extensive α or β -structure, the polyQ tract could also adopt a β -structure in monomeric Htt exon 1.

In the second scenario, expanded polyQ tracts have been suggested to destabilize the proteins and thereby facilitate the formation of partially unfolded species (Ignatova *et al.*, 2007, Tanaka *et al.*, 2001, Thakur *et al.*, 2009). Such species generally expose at least part of their main chain and hydrophobic residues to solvent and are therefore prone to intermolecular interactions leading to fibril formation. Such a mechanism has been described for other proteins associated with amyloidosis, including transthyretin (Johnson *et al.*, 2005) and human lysozyme (Dumoulin *et al.*, 2006, Dumoulin *et al.*, 2003). This hypothesis has been mainly tested with model polyQ proteins due to the problems encountered to produce and handle polyQ disease-associated proteins. The results obtained with these model polyQ-proteins indicate that the effects of different polyQ lengths on the stability/structure of the host protein depend on the location of this latter and that destabilization and/or structural perturbations are not necessarily observed (Table 6). This suggests that destabilization/structural perturbations may contribute to pathogenesis in some polyQ diseases, but are not essential for aggregation.

The results described above are summarized in Table 6 which clearly show that the situation is much more complicated than anticipated and the main conclusions that emerge from this table are that:

- in general, there is an apparent threshold length above which polyQ proteins readily form fibrils *in vitro*. Above this apparent threshold length, the rate of polyQ-dependent fibril formation (or the propensity to form these fibrils for the *in vivo* aggregation of CRABPI fused to Htt exon 1) is increased as the polyQ length increases. As the dependence of the

aggregation rate of polyQ peptides on the polyQ length, this can also be correlated to the anticipation phenomenon observed in polyQ diseases.

- the aggregation behavior of polyQ-containing proteins are not only influenced by the polyQ length but is also modulated by the flanking domains/sequences. The presence of non-polyQ flanking sequences or domains could reduce the aggregation propensity of polyQ-containing proteins by sterically hindering polyQ intermolecular interaction and/or by restricting the polyQ conformational change required for fibril formation. Such influence was proposed for the polyproline region C-terminal to the polyQ tract in Htt exon 1 but also in model polyQ proteins, to explain the influence of domain architecture on polyQ-mediated aggregation.

Table 6: Summary of studies on polyQ disease-associated proteins and model polyQ proteins

Protein	Location	Effects on the structure of the protein or that of the flanking sequences		PolyQ structure	Effects on stability	Threshold length for fibril formation ⁽¹⁾	Effects of polyQ length on aggregation rate or propensity
		3 ^{ary}	2 ^{ary}				
Htt exon 1	/	+		Random coil/ β	/	+ (32-37Q), MDA	+
Atx-3	/	(Not clear)		/	-	+ ⁽²⁾ , MDA	/
Myoglobin	Insertion	+	-	β ($\geq 28Q$)	+	+ (12-28Q), MDA for 50Q	+
GST ⁽³⁾	Fusion	/	-	Random coil	/	+(22Q-41Q), MDA	/
CRABPI ⁽⁴⁾	Fusion	+	+	/	+	+ (40-53Q), MDA	+
Thioredoxin	Fusion	/		α $\alpha \rightarrow \beta$ ($\geq 62Q$)	/	+ (35-62Q)	+
SpA	Fusion	-	-	Random coil	-	+ (20-35Q)	+
	Insertion	+	-	/	+	/	/
Apomyoglobin	Fusion	-	-	α	+ ⁽⁵⁾	-	/
	Insertion	+	+	β ($\geq 16Q$)	+ ⁽⁶⁾	/ ⁽⁷⁾	/

MDA, multidomain aggregation, +, there is an effect. -, no effect observed. /, not studied.

⁽¹⁾ The range of polyQ lengths in which the threshold length to readily form fibrils is comprised is indicated in brackets.

⁽²⁾ Only the expanded ataxin-3 form SDS-insoluble fibrils

⁽³⁾ Thermal-induced aggregation

⁽⁴⁾ Fusion to Htt exon 1 and in-cell aggregation

⁽⁵⁾ PolyQ tracts increase resistance to chemical unfolding

⁽⁶⁾ PolyQ tracts increase sensitivity to chemical unfolding

⁽⁷⁾ oligomerization produced through structural perturbation and not associated with the polyQ tract

In addition, flanking domains could also assist and/or enhance the rate of polyQ-dependent fibril formation and can determine the molecular species formed during the aggregation. Indeed, the formation of several intermediate oligomeric species is observed on the aggregation pathway of different polyQ-containing proteins while such species have not been identified for polyQ peptide aggregating systems. A multi-domain aggregation pathway was

described/proposed for three of the polyQ disease-associated proteins and for several model polyQ-proteins (Table 6). In these multi-domain aggregation pathways, the formation of the early oligomeric species is initiated by intermolecular interaction between non-polyQ regions and not the polyQ tracts. Although the polyQ tract is not apparently involved in the formation of the early oligomeric species, it does or could, in some cases, increase the aggregation propensity of the non polyQ regions by destabilizing and/or inducing structural perturbation in the non-polyQ domains.

However, the initial aggregation of non-polyQ regions is not a necessary step that precedes polyQ intermolecular interactions in polyQ-containing protein as suggested by the study on the fusion protein SpA-polyQ. It is therefore possible that the formation of amyloid-like fibrils by some of the polyQ disease-related proteins not characterized so far also aggregate through such simple mechanism essentially involving the polyQ tract.

In conclusion, the studies on polyQ disease-related proteins as well as those on model polyQ proteins highlight the existence of a very complex interplay between the intrinsic properties of the polyQ tract to mediate aggregation and the modulating effects of the host proteins.

Chapter 2. Objectives

1. Open questions

As presented in the introduction, a number of studies have highlighted the existence of a complex interplay between the intrinsic properties of the polyQ tract to trigger protein aggregation and the modulating effects of the host protein. To gain a better insight into the general principles governing this complex phenomenon, it is necessary to further investigate in detail which properties of the host protein influence the ability of polyQ stretches to mediate aggregation. The characterization of model polyQ proteins has unambiguously pointed to the length and the location of the polyQ tract as important factors (Ignatova and Gierasch, 2006, Robertson *et al.*, 2011b, Robertson *et al.*, 2008, Tanaka *et al.*, 2001, Tobelmann and Murphy, 2011). A number of questions are, however, not yet fully addressed; these include the role of the sequence, size, structure, stability, dynamics and topology of the host protein. There is therefore a clear need for the generation of new model polyQ proteins of various sizes, topologies and structures and the extensive characterization of their structural, thermodynamic, dynamic and aggregating properties.

In proteins associated with polyQ diseases, the polyQ tract is separated from the C- or N-terminal ends by at least 50 residues, except for the proteins associated with Huntington's disease (HD, 17 residues) and spinocerebellar ataxia type 6 (SCA6, 30 residues) (Robertson *et al.*, 2011b). Many model proteins characterized so far, however, display a polyQ tract fused to their N- or C-terminus (Klein *et al.*, 2007, Masino *et al.*, 2002, Robertson *et al.*, 2008, Tobelmann and Murphy, 2011), probably because polypeptide chains with an inserted polyQ sequence are difficult to express (Robertson *et al.*, 2008, Tobelmann and Murphy, 2011). It is therefore essential to investigate more model proteins containing an inserted polyQ stretch since the presence of flanking sequences at both extremities of the polyQ tract is likely to impose constraints that are more physiologically relevant than when the polyQ tract has a free extremity. With this aim in mind, the objectives of my work were to insert polyQ sequences of various lengths into another well-characterized protein, the β -lactamase from *Bacillus licheniformis* 749/C, BlaP and to characterize the properties of the chimeric proteins obtained.

2. What is BlaP and why is it a potentially good scaffold to create model polyQ proteins?

The β -lactamase BlaP is an enzyme produced by the bacteria *Bacillus licheniformis* 749/C and secreted in the external environment in order to escape the lethal action of β -lactam antibiotics. This enzyme irreversibly hydrolyzes with a high efficiency the amide bond of the β -lactam ring converting the antibiotic into an inactive product. In the broad family of β -lactamases, serine and zinc enzymes are distinguished. The active-site serine β -lactamases are distributed in three different classes: A, C and D according to their amino acid sequences. All active-site zinc β -lactamases are in the class B and possess one or two zn^{2+} ions (Matagne *et al.*, 1998). BlaP is a serine enzyme that belongs to the class A β -lactamases.

We thought that BlaP could be a good protein scaffold to insert polyQ sequences because:

- its three-dimensional structure is known:

Like all class A β -lactamases, this 264-residue enzyme is organized in two structural domains (Figure 13): the all- α domain and the α/β domain. The active site is located into the groove at the interface of the two domains (Knox and Moews, 1991).

- Detailed informations concerning its catalytic properties are available (Matagne *et al.*, 1990) and its enzymatic activity is a good reporter of its structural integrity.
- Most importantly, this protein contains several permissive insertion sites.

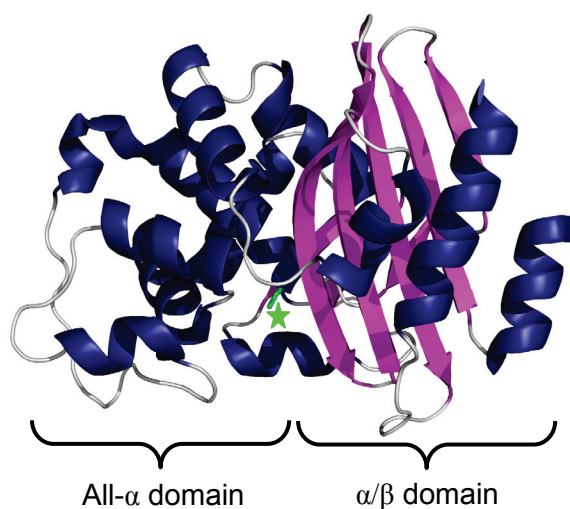


Figure 13: X-ray structure of the β -lactamase BlaP from *Bacillus licheniformis* 749/C produced using PyMOL (DeLano Scientific LLC, South San Francisco, CA, USA) using the pdb entry 4BLM (Knox and Moews, 1991). The active-site serine (Ser70) is indicated by a green star.

3. How the location of the permissive sites have been determined?

The possibility to insert peptides in different positions into the sequence of class A β -lactamases was initially carried out with the enzyme TEM-1 using a pentapeptide scanning mutagenesis strategy (Hallet *et al.*, 1997, Ruth *et al.*, 2008). This method consists in the random insertion of a transposon (Tn4430) into the TEM-1 gene and its subsequent *in vitro* deletion. This deletion leaves 15 bp (base pairs) into the gene and thus 5 random amino acids are added to the amino acid sequence of TEM-1. A pentapeptide has been inserted in such way at 24 different locations of the protein (Figure 14). The different insertion sites have been classified in three types depending on the level of resistance that the hybrid protein confers to *E. coli* cells toward ampicillin, a β -lactam antibiotic:

- 6 permissive sites (P) which confer a high level of resistance. These sites are located into two protruding loops which are distant from the catalytic site: the loops between the α -helices 8 and 9 and between the strand β 5 and the C-terminal helix.
- 8 non-permissive sites (NP) that do not confer any resistance. They are clustered around the catalytic site.
- 10 semi-permissive (SP) sites which confer an intermediate level of resistance; they are localized in the other regions of the enzyme.

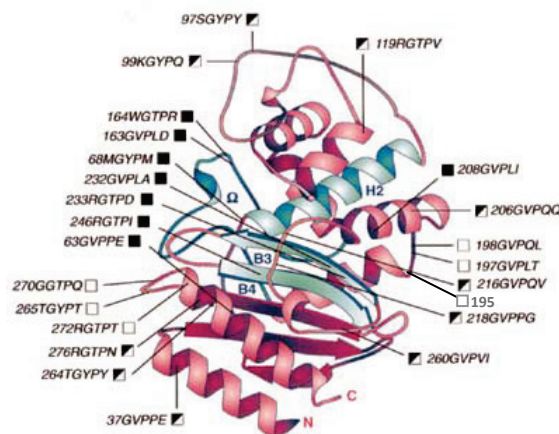


Figure 14: Structure of TEM-1 β -lactamase with the 24 pentapeptide insertions generated by pentapeptide scanning mutagenesis. Position number of the insertion and sequence of the pentapeptides inserted are indicated. Open, half-filled and filled boxes correspond to permissive, semi-permissive and non-permissive insertion sites. Images adapted from (Hallet *et al.*, 1997).

Ruth *et al.* have further tested the tolerance of TEM-1 to the insertion of longer (up to ~36 amino acids) and structured peptides in eight positions tested by pentapeptide scanning mutagenesis [37(SP), 195 (P), 198(P), 206(SP), 216(SP), 218(SP), 232(NP) and 260 (SP)] (Ruth *et al.*, 2008). In this study, *E. coli* resistance towards ampicillin conferred by the hybrid proteins was also determined and the presence of soluble chimeras in *E. coli* cells was detected by western blot. The results obtained indicated that the best chance to produce an

active soluble hybrid protein was when the larger peptides were inserted in positions 195, 198, 216 or 260 of TEM-1. The level of ampicillin resistance and the amount of soluble hybrid protein detected were however dependent on the peptide inserted and the insertion site used, respectively.

TEM-1 is known to be highly sensitive to proteases (Wu *et al.*, 1991) and it was shown that the non-permissivity of some insertions in this protein was due to a further increased sensitivity of the hybrid proteins to proteases (Chevigne *et al.*, 2007).

In contrast, BlaP is a natural protease-insensitive class A β -lactamase that has been isolated from *Bacillus licheniformis*, a gram-positive bacterium that secretes various and numerous proteases. BlaP appeared therefore to be a better scaffold than TEM-1 to insert exogenous polypeptides. Moreover, TEM-1 and BlaP have a very similar structure (Figure 13 and 14) and ~40% and ~20% of their amino acids are identical and homologous, respectively. The permissive insertion sites identified in TEM-1 can therefore served as a basis for engineering hybrid protein with BlaP.

On this basis, Dr Filée from our institute has developed an *E. coli* expression vector (pNYESBlaP) allowing the insertion of a heterologous polypeptide sequence into the permissive loop between the α -helices 8 and 9 (in position 197-198) of BlaP (Figure 15A and B). Several chimeric β -lactamases with inserts of various lengths and structures have been recombinantly produced in *E. coli* in significant amounts [(Vandevenne *et al.*, 2007, Vandevenne *et al.*, 2008) and unpublished results].

4. Strategy and outline of the work

This thesis focuses on the creation and characterization of BlaP chimeras containing polyQ inserts at different locations to investigate the mechanism by which polyQ tracts trigger the aggregation of proteins. The strategy is to compare the effects of the insertion of polyQ tracts of lengths shorter and longer than the pathological threshold length observed in polyQ diseases (i.e. 35-45 residues). The effects of the insertion on the catalytic, structural, thermodynamic and aggregating properties of the polyQ chimeras have been investigated using a range of biophysical techniques including fluorescence, circular dichroism (CD), transmission electron microscopy (TEM) and X-ray diffraction and will be compared to that of the β -lactamase without polyQ tract. The final goal is to identify a correlation, if any,

between the aggregating properties of the proteins and the effects of the inserts on the structure and thermodynamic properties of BlaP.

The choice of the insertion sites have been made in the light of the results presented above. The first site tested was therefore that in position 197-198 located into the loop between the α -helices 8 and 9 as it is the most extensively investigated (Figure 15A and B) (Vandevenne *et al.*, 2007, Vandevenne *et al.*, 2008, Chevigne *et al.*, 2007). The second insertion site was chosen among the two other sites that were shown to be the most promising for the insertion of long polypeptide in TEM-1 and that are not in the same loop than the site 197-198. They are located in position 216 (loop between α -helices 9 and 10) and in position 260 (in the N-terminal end of the strands B5).

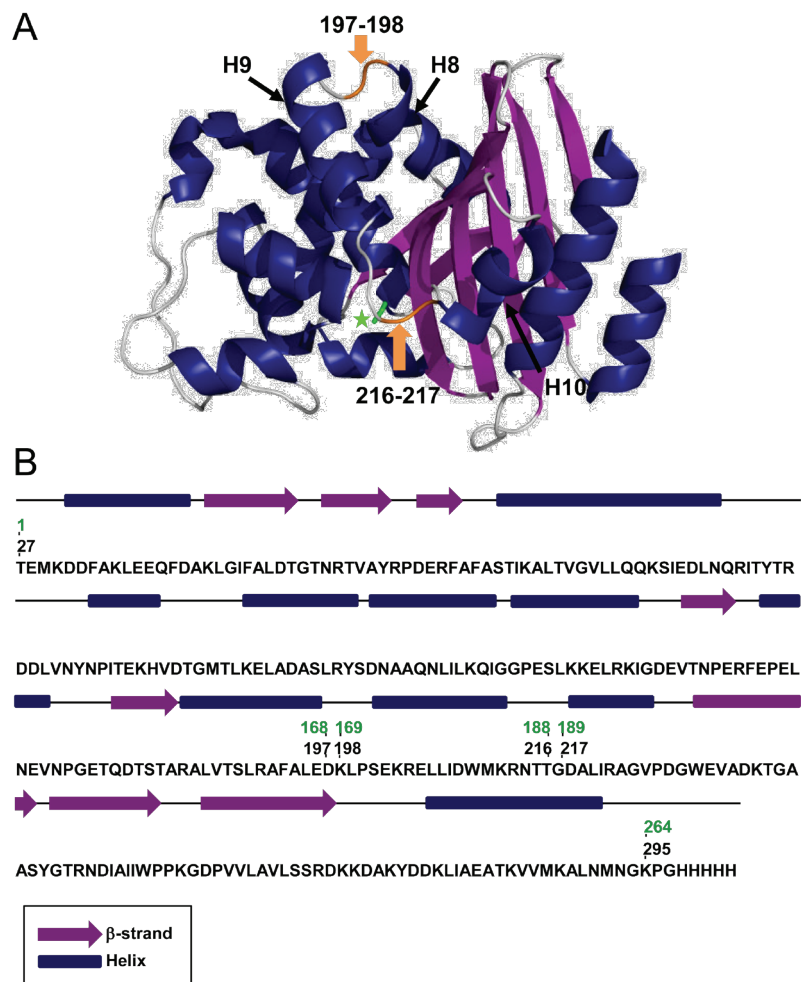


Figure 15: (A) X-ray structure of the β -lactamase BlaP from *Bacillus licheniformis* 749/C produced using PyMOL (DeLano Scientific LLC, South San Francisco, CA, USA) using the pdb entry 4BLM (Knox and Moews, 1991). The active-site serine (Ser70) is indicated by a green star and insertion sites 197-198 and 216-217 are shown by an orange arrow. (B) Topology of BlaP. Residue numbers in black correspond to the numbering scheme of class A β -lactamases (Ambler *et al.*, 1991) and residue numbers shown in green correspond to the sequential numbering of the mature form of BlaP.

The site 216 was selected since it is located into a loop whereas the site 260 is in an element of secondary structure. Moreover, hybrid proteins created by insertion of a peptide of 18 amino acids in position 216 and 260 of TEM-1 were produced and purified by Ruth *et al* (Ruth *et al.*, 2008). The determination of their molecular weight by mass spectrometry indicated that the insertion in position 260 increases the susceptibility of TEM-1 to proteolysis (the last α -helix was deleted). Although BlaP is more resistant to proteases than TEM-1, we preferred to avoid the risk of proteolysis and we therefore decided to insert polyQ tracts in position 216-217 of BlaP (Figure 15A and B).

The next chapter introduces the materials and methods that are relevant to the following sections of the thesis. In the fourth and fifth chapters, we report and discuss the results obtained for chimeras containing polyQ tracts in position 197-198 and 216-217, respectively. Finally, all the results obtained are summarized and discussed further in the sixth chapter.

Chapter 3. Materials and methods

1. Molecular biology

1.1. Construction of the expression vectors of chimeric β -lactamases with polyQ insertions in position 197-198

A library of $(CAG)_n$ double-strand DNA fragments was constructed by an overlapping PCR strategy (Ordway and Detloff, 1996), using the following oligonucleotides: 5'-(CAG)₁₃-3' and 5'-(CTG)₁₃-3'. A PCR using Pfu DNA polymerase (Promega, Madison, WI, USA) was performed as follows: 3 min at 95°C, 30 sec at 94°C, 1 min at 55°C and 30 sec at 68°C; the last three steps were repeated 35 times. The library was purified from a 2% agarose gel and the extremities of each double-strand DNA fragment were blunt-ended with Pfu DNA polymerase and dNTPs at 68°C over a period of 30 min. The polyCAG DNA fragments were then dephosphorylated for 30 min at 37°C using calf intestinal alkaline phosphatase (Roche, Mannheim, Germany). Finally, the polyCAG double-strand DNA library was inserted into the *Sma*I restriction site, which was previously introduced into the gene of BlaP (between codons for residues 197 and 198) carried by the constitutive expression vector pNY ESBLAP (Vandevenne *et al.*, 2007) (Figure 16). Note that the insertion of the *Sma*I restriction site introduces a dipeptide proline-glycine into the amino acid sequence of BlaP; the resulting protein will be referred to as BlaP₁₉₇₋₁₉₈.

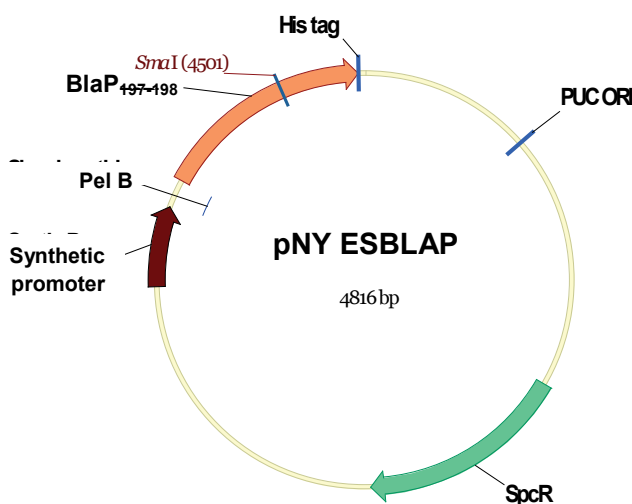


Figure 16: Expression vector pNY ESBLAP encoding for BlaP₁₉₇₋₁₉₈. PUC ORI: *E. coli* origin of replication from plasmid pUC; SpcR: spectinomycin resistance gene; Synthetic promoter: promoter constructed by Dr. P. Filée (University of Liège, Belgium); Pel B: signal peptide for periplasmic secretion; BlaP₁₉₇₋₁₉₈: β -lactamase BlaP gene interrupted by a *Sma*I restriction site between codons for residues 197 and 198 and followed by the nucleotide sequence coding for a dipeptide glycine-proline and a (His)₅ tag.

In this vector, the gene of BlaP₁₉₇₋₁₉₈ is followed by the nucleotide sequence coding for an additional C-terminal dipeptide glycine-proline and a (His)₅ tag. The resulting library of expression vectors was used to transform *E. coli* DH5 α cells (Invitrogen, Paisley, UK). Note

that the β -lactamase activity which enables *E. coli* to become resistant to β -lactam antibiotics can be used as a reporter to efficiently select clones producing soluble chimeras in which BlaP is correctly folded. Consequently, the transformed cells were plated on LB (Luria-Bertani) medium containing ampicillin ($10 \mu\text{g}\cdot\text{mL}^{-1}$, Sigma) in addition to spectinomycin ($75 \mu\text{g}\cdot\text{mL}^{-1}$, Sigma) for which the plasmid contains a gene of resistance. The presence of an insert within the BlaP gene was checked by colony-PCR on more than 50 randomly selected transformants. Then, plasmids from about 10 colonies carrying an insertion were amplified by cell culture and extracted. As a second control for the presence of an insert, purified plasmids were digested with the restriction enzyme SmaI whose recognition site is not regenerated when an insert is present. Finally, the sequence of the insert was determined by the Sanger method [realized at the CIP by Iris Thamm or at the GIGA GenoTranscriptomics technology platform (Liège, Belgium)].

1.2. Construction of the expression vectors of chimeric β -lactamases with polyQ insertions in position 216-217

Synthetic genes of BlaP₂₁₆₋₂₁₇, BlaP₂₁₆₋₂₁₇(Gln)₃₀, BlaP₂₁₆₋₂₁₇(Gln)₅₅ and BlaP₂₁₆₋₂₁₇(Gln)₇₉ flanked at both sides by the nucleotide sequences allowing their cloning into the expression vector pNY (in 5': gagctcaggcc; in 3': tgagaattc) were ordered from Genscript USA Inc and received in the cloning vector pUC57. These 5' and 3' sequences contain the recognition site for the restriction enzymes *SacI* and *EcoRI*, respectively (Figure 17). These vectors have been referred to as pUC57 BlaP₂₁₆₋₂₁₇(Gln)_n with n= 0, 30, 55 and 79). The synthetic genes contain in 3' the nucleotide sequence coding for an additional C-terminal dipeptide glycine-proline (Gly-Pro) followed by a (His)₅ tag since these amino acids were present in the sequence of BlaP₁₉₇₋₁₉₈ and its chimeras. First, the expression vector pNY ESBLAP and the plasmids pUC57 BlaP₂₁₆₋₂₁₇(Gln)_n were digested by *SacI* and *EcoRI* in order to obtain the free expression vector pNY and to excise the synthetic genes. These synthetic genes were then dephosphorylated and cloned into the expression vector pNY.

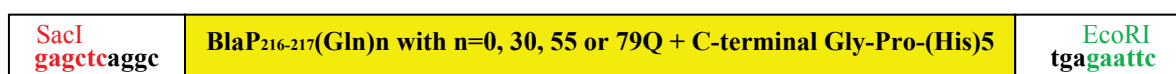


Figure 17: Schematic representation of the synthetic genes.

2. Protein expression and purification

The proteins were expressed in *E. coli* JM109 strains (Promega). A 100 mL LB (Luria-Bertani) medium preculture, containing $75 \mu\text{g}\cdot\text{mL}^{-1}$ of spectinomycin and $10 \mu\text{g}\cdot\text{mL}^{-1}$ of ampicillin, was inoculated with the transformed cells and incubated at 37°C for approximately 7 hours. Two liters of modified TB (Terrific Broth) medium containing no glycerol, but supplemented with $75 \mu\text{g}\cdot\text{mL}^{-1}$ of spectinomycin, $10 \mu\text{g}\cdot\text{mL}^{-1}$ of ampicillin, $4.2 \mu\text{M}$ of biotin and $2.2 \mu\text{M}$ riboflavin, were inoculated with 5 mL of preculture and incubated overnight at 37°C . Periplasmic proteins were then extracted by osmotic shock as described by Vandevenne *et al.* (Vandevenne *et al.*, 2007). The proteins of interest, which were expressed with a C-terminal (His)₆ tag, were purified in a single step by metal chelate affinity chromatography, using a 20 mL Ni-PDC column (Affiland, Liège, Belgium). The periplasmic proteins were loaded onto the Ni-PDC column equilibrated with 4 volumes of PBS (50 mM sodium phosphate, 150 mM NaCl), pH 7.5. After loading, the column was washed successively with 60 mL of: (i) PBS, pH 7.5; (ii) PBS, pH 7.5, containing 2 M NaCl; (iii) PBS, pH 7.5, containing 10 mM imidazole. Proteins were eluted with a linear imidazole gradient (0-300 mM) in PBS. If the enzymatically active fractions (probed with nitrocefin as the substrate) contain more than 95% of the protein of interest [as assessed by 15% (w/v) SDS-PAGE], they were collected and pooled. They were then dialyzed four times against 15 L of milliQ water, lyophilized and stored at 4°C , or dialyzed two times against 15 L of PBS and stored at -20°C . If contaminant proteins were still present in significant amount after purification by affinity chromatography, a second step of purification was performed by anion-exchange chromatography using a 5 ml HiTrap Q HP column (GE healthcare, Uppsala, Sweden). For this second step of purification, the fractions eluted from the Ni-PDC column and containing the protein of interest were pooled and dialyzed against 15 L of 20 mM Tris-HCl pH 8. They were then loaded onto a HiTrap Q HP column equilibrated with 5 volumes of 20 mM Tris-HCl pH 8. The proteins were then eluted using a NaCl linear gradient (0-1 M) and enzymatically active fractions containing more than 95% of the protein of interest [as assessed by 15% (w/v) SDS-PAGE] were pooled and dialyzed two times against 15 L of PBS pH 7.5.

3. Size-exclusion chromatography

A Superdex 200 GL 10/300 column (GE Healthcare) was equilibrated with PBS buffer, pH 7.5. Solutions of proteins in PBS, pH 7.5, were injected and eluted at a flow rate of $0.5 \text{ mL} \cdot \text{min}^{-1}$. Elution of protein was monitored by absorbance measurements at 280 nm. The column was calibrated using 7 protein standards: hen egg white lysozyme (Belovo, Bastogne, Belgium), chicken egg white albumin (Sigma, A2512), phosphorylase B from rabbit muscle (Sigma, P6635), β -galactosidase from *E. coli* (Sigma, 48275), aldolase from rabbit muscle (Sigma, A2714), catalase from bovine liver (Sigma, C40) and thyroglobulin from bovine thyroid (Sigma, T9145). Blue dextran 2000 (GE healthcare) was used to determine the void volume in the column.

4. N-terminal sequencing and mass spectrometry

N-terminal sequencing was performed using the Edman degradation procedure according to Han's protocol (Han *et al.*, 1977). The molecular masses were determined using electrospray ionisation quadrupole time-of-flight mass spectrometry (ESI-QTOF MS) at the GIGA proteomics platform (Liège, Belgium). The percentage of each species observed within the same sample injection run was estimated from the centroid value of peak.

5. Quantification of the protein

The molar extinction coefficient ($33000 \text{ M}^{-1} \cdot \text{cm}^{-1}$ at 280 nm) of BlaP₁₉₇₋₁₉₈ was determined experimentally using the bicinchoninic acid (BCA) assay from Pierce (Rockford, IL, USA). This value was used for the determination of the concentrations of BlaP₁₉₇₋₁₉₈, BlaP₂₁₆₋₂₁₇ and their polyQ chimeras.

6. Enzymatic activity measurements

Determination of catalytic parameters. Usually, complete time-courses of the hydrolysis of cephalothin (Sigma, C4520) were recorded at 260 nm, using a UVIKON 860 spectrophotometer (Kontron Instrument, Zürich, Switzerland). The catalytic parameters (k_{cat} and K_{m}) were determined using the integrated form of the Henri-Michaelis equation (De Meester *et al.*, 1987). In cases where K_{m} value was too high, initial rate of hydrolysis were measured and only the ratio $k_{\text{cat}}/K_{\text{m}}$ was determined. All reactions were performed at 30°C and in 50 mM sodium phosphate buffer, pH 7. The protein samples were diluted with 50 mM

sodium phosphate buffer pH 7 containing 0.1 mg·ml⁻¹ of bovine serum albumin (BSA). Standard deviations were computed on the basis of the results of experiments performed at various substrate and enzyme concentrations.

Activity measurements for the table of purification. Initial rates of cephalothin or benzylpenicillin (Rhône Polenc, Paris, France) hydrolysis were measured to rapidly determine the specific activity. The hydrolysis of benzylpenicillin was monitored at 235 nm.

7. Fluorescence and circular dichroism (CD) measurements

Fluorescence data were acquired using either a Cary Eclipse spectrofluorimeter equipped with a Peltier-controlled holder (Varian, Mulgrave, Australia) or a LS50B spectrofluorimeter (Perkin-Elmer, Norwalk, CT, USA) and a 1 cm pathlength cell. CD measurements were performed on a Jasco J-810 spectropolarimeter (Jasco, Tokyo, Japan) equipped with a Peltier-controlled holder and using 1 mm and 1 cm pathlength cells for far-UV and near-UV measurements, respectively.

8. Fluorescence and circular dichroism spectra of native proteins

All spectra were recorded at 25°C in PBS, pH 7.5, using protein concentrations of 4.6 μM for fluorescence and far-UV CD measurements, and 20 μM for near-UV CD measurements. Five fluorescence emission spectra were recorded in the 300-400 nm range ($\lambda_{exc}=295$ nm, $slit_{exc}=2.5$ nm, $slit_{em}=5$ nm), at a rate of 600 nm·min⁻¹ using the Cary Eclipse spectrofluorimeter, and averaged. Twenty CD spectra were acquired at a rate of 50 nm·min⁻¹, both in the far-UV (200-250 nm) and near-UV (250-350 nm) regions, and averaged. The bandwidth and the response time were 1 nm and 0.5 sec, respectively. All protein spectra (fluorescence and CD) were corrected for the buffer contribution.

9. Urea-induced unfolding experiments

Protein samples (4.6 μM) in PBS, pH 7.5, at various urea concentrations [VWR BDH Prolabo, 0 to 5.5 M (for BlaP₂₁₆₋₂₁₇(Gln)₇₉, 0 to 4.5 M) by 0.1 M increments] were unfolded to equilibrium by incubation for *ca.* 16 h at 25°C. Urea concentrations were determined from the refractive index measurements (Pace and Scholtz, 1997) using a R5000 refractometer from Atago (Tokyo, Japan):

$$[\text{urea}]_{\text{exp}} = (117.66 \Delta\eta) + (29.753 \Delta\eta^2) + (185.56 \Delta\eta^3) \quad (1)$$

where $\Delta\eta$ is the difference between the refractive index of PBS buffer and samples containing urea.

Five fluorescence spectra of each sample ($\lambda_{exc}=295$ nm, $\lambda_{em}=300-400$ nm) were recorded using either the LS50B (scan rate= 280 nm \cdot min $^{-1}$) or the Cary Eclipse (scan rate= 600 nm \cdot min $^{-1}$) spectrofluorimeter and averaged. Far-UV CD signal at 222 nm (bandwidth= 1 nm, response= 4 sec) was also monitored and for each sample, 30 data were acquired at the rate of one datum every 6 s, averaged. The background of the solutions (PBS buffer + denaturant) was subtracted from the fluorescence spectra and CD signals.

Unfolding curves were established by plotting the fluorescence intensity at 323 nm (for BlaP₁₉₇₋₁₉₈ and its chimeras) or at 335 nm (for BlaP₂₁₆₋₂₁₇ and its chimeras) and the far-UV CD signals at 222 nm as function of the urea concentration.

Moreover, other protein samples were denatured for 3 hours at 25°C in 5.5 M urea (or in 4.5 M for BlaP₂₁₆₋₂₁₇(Gln)₇₉) (under these conditions, all the investigated proteins have been shown to be completely unfolded, data not shown) and renatured (for *ca.* 16 h) by dilution to different urea concentrations (from 5.5 to 0.55 M or from 4.5 to 0.45 M). The reversibility of the unfolding transitions was demonstrated by comparing the fluorescence and CD signals recorded with these samples to those obtained for the samples unfolded at similar urea concentrations.

Data analysis

The thermodynamic parameters were computed on the assumption of a two-state model for the unfolding reaction ($N \rightleftharpoons U$). On this basis, the transition curves were analyzed according to the following equation (Dumoulin *et al.*, 2002, Pace and Scholtz, 1997):

$$y_{obs} = \{(y_N + p \times [D]) + (y_U + q \times [D]) \times \exp [a]\} / (1 + \exp [a]) \quad (2)$$

where $a = -(\Delta G^{\circ}_{N-U}(H_2O) - m \times [D])/RT$ and y_{obs} , y_N and y_U are the value of measured variable at a given denaturant concentration and the values of this parameters for the native (N) and denatured (U) states, respectively. $\Delta G^{\circ}_{N-U}(H_2O)$ corresponds to the difference in the free energy between the folded and unfolded states in absence of denaturant (also defined as the conformational stability of the protein). m is the measure dependence of the free energy on the

denaturant concentration, and $[D]$ is the denaturant concentration. p and q are the slopes of the pre- and post-transition baselines. R is the gas constant and T is the absolute temperature.

10. Heat-induced unfolding

Heat-induced unfolding was monitored by the changes in the intrinsic fluorescence intensity [slits_{exc/em}=5 nm, λ_{exc} =295 nm, λ_{em} =323 nm for BlaP₁₉₇₋₁₉₈ and its chimeras or λ_{em} =350 nm for BlaP₂₁₆₋₂₁₇ and its chimeras] using the Cary Eclipse spectrofluorimeter, and in far-UV CD signal at 222 nm (bandwidth=1 nm, response=4 sec). The protein concentration was 4.6 μ M in PBS, pH 7.5, and mineral oil was added on top of the samples to limit solvent evaporation. The temperature was increased from 25 to 85-90°C at a rate of 0.5°C·min⁻¹; the fluorescence and the CD data were acquired every 0.5°C. The temperature in the cell was measured with a PT200 thermocouple (IMPO Electronic, Olgod, Denmark). The reversibility of the heat-induced unfolding was assessed by monitoring the changes in the fluorescence and CD signals upon cooling the sample down to 25°C at a rate of 0.5°C·min⁻¹.

Data analysis

The thermal denaturation was computed on the assumption of an apparent two-state model for the unfolding reaction ($N \rightleftharpoons U$). On this basis, the transition curves were analyzed according to the following equation (El Hajjaji *et al.*, 2009):

$$y_{obs} = \{(y_N + p \times T) + (y_U + q \times T) \times \exp [a]\} / \{1 + \exp [a]\} \quad (3)$$

where $a = -(\Delta H_m (1-T/T_m))/RT$ and where y_{obs} is the observed CD signal at 222 nm (or the fluorescence at 323 nm) at a given temperature, and y_N and y_U represent the values of y for the native and denatured states, respectively. p and q are the slopes of the pre- and post-transition baselines, respectively. R is the gas constant, and T is the absolute temperature. T_m is the temperature of mid-transition and ΔH_m is the enthalpy value at T_m . Since the transitions are not fully reversible, only the (apparent) T_m value is considered.

11. Aggregation kinetics

A series of tubes containing 100 μ L of 110 μ M protein in PBS, pH 7.5, containing 0.2% sodium azide, were incubated either (i) at 37°C in the absence of urea or (ii) at 25°C in the presence of concentration in urea for mid-unfolding or (iii) at 25°C in the presence of 3.5 M

urea. Airtight tubes (Multiply Safecup, Sarstedt, Nümbrecht, Germany) were used to limit evaporation. At selected times, one tube was removed and centrifuged for 50 min at 12000 rpm. The supernatant was used to determine the quantity of remaining soluble protein by absorbance measurements at 280 nm and the protein integrity was demonstrated by SDS-PAGE analysis. Unless otherwise stated, the aggregation time-courses were repeated three times with proteins originating from different production and purification batches. For one of the aggregation time-courses of BlaP₁₉₇₋₁₉₈ and its chimeras, aliquots of samples at the initial (T_0) and end (T_f) time-points were taken (in triplicate) before centrifugation for thioflavin T (ThT) fluorescence measurements; they were kept frozen until analysis. For all proteins, samples at T_f were also analyzed by transmission electron microscopy (TEM).

12. Thioflavin T (ThT) fluorescence measurements

To 5 μ L protein sample was added 1.5 mL of 10 mM sodium phosphate buffer, 150 mM NaCl, 50 μ M ThT (Sigma, T3516), pH 7. Ten fluorescence emission spectra were recorded (using the Cary Eclipse spectrofluorimeter) at 25°C with stirring in the 450-600 nm range ($\lambda_{exc}=440$ nm, $slit_{exc/em}=5$ nm) at a rate of 1200 nm \cdot min⁻¹, averaged, and corrected for the background fluorescence of the ThT solution alone.

13. Transmission electron microscopy

6 μ l of appropriately diluted samples were left for 4 min on carbon-coated 400-mesh copper grids, before being stained for 1 min with 2% uranyl acetate (w/v). The grids were then washed once with 2% uranyl acetate and finally, three times with milliQ water. Images were recorded on a Philips CEM100 transmission electron microscope operating at 100 kV.

14. X-ray fibre diffraction

Fibril suspensions were centrifuged and the pellet subjected to two washing cycles in milliQ water; between each cycle, the suspension was centrifuged and the supernatant discarded in order to remove any trace of soluble protein, buffer, or urea. The fibrils were then aligned using a modification of the stretchframe method as previously described (Mossuto *et al.*, 2010). X-ray diffraction data were collected at room temperature for *ca.* 10 min on a Bruker AXS FR591 diffractometer with Cu K α radiation with a wavelength of 1.5418 Å and

equipped with a MARDTB 345 mm image plate detector. The sample-to-detector distance was 300 mm.

15. Fibril activity

Samples containing 50 μL of BlaP₁₉₇₋₁₉₈(Gln)₇₉ (110 μM) were incubated during 192 hours in PBS, pH 7.5, at either 37°C in the absence of urea or 25°C in the presence of 1.85 or 3.5 M urea. Two aliquots of 5 μL were taken from these samples and used to measure initial rates of cephalothin (100 μM) hydrolysis (see section 6). These samples were then centrifuged for 30 min at 14000 rpm to separate the aggregates and the supernatant containing the remaining soluble proteins. Aggregates were subjected to two cycles of washing in 50 mM sodium phosphate, pH 7; between each cycle, the suspension was centrifuged and the supernatant discarded in order to remove any trace of soluble protein. The fibrils were finally resuspended in a volume of phosphate buffer equivalent to the volume of the samples before centrifugation. Activity measurements were carried out with both the solution containing remaining soluble proteins and the purified aggregates in phosphate buffer.

Chapter 4. Creation and characterization of BlaP chimeras with polyQ tracts in position 197-198

1. Creation of chimeras with polyQ inserts in position 197-198

A library of polyCAG double-strand DNA fragments of different lengths was constructed by an overlapping PCR strategy (Figure 18) (Ordway and Detloff, 1996), using the following oligonucleotides: 5'-(CAG)₁₃-3' and 5'-(CTG)₁₃-3'; CAG being one of the two codons for the glutamine. In this unconventional PCR, oligonucleotides serve as both primer and template. The temperature of the hybridization step (i.e. 55°C) was chosen to be significantly lower than the optimal hybridization temperature of the two oligonucleotides (i.e. 70°C). This setting allows, at each cycle, a partial overlap of the primers and those pairing with a 5' overhang provide a template for elongation by DNA polymerase (Figure 18). During this PCR, the number of different overlapping possibilities and the average length of the oligonucleotides increase at each cycle, generating a library of polyCAG double-strand DNA fragments of different lengths (Figure 18).

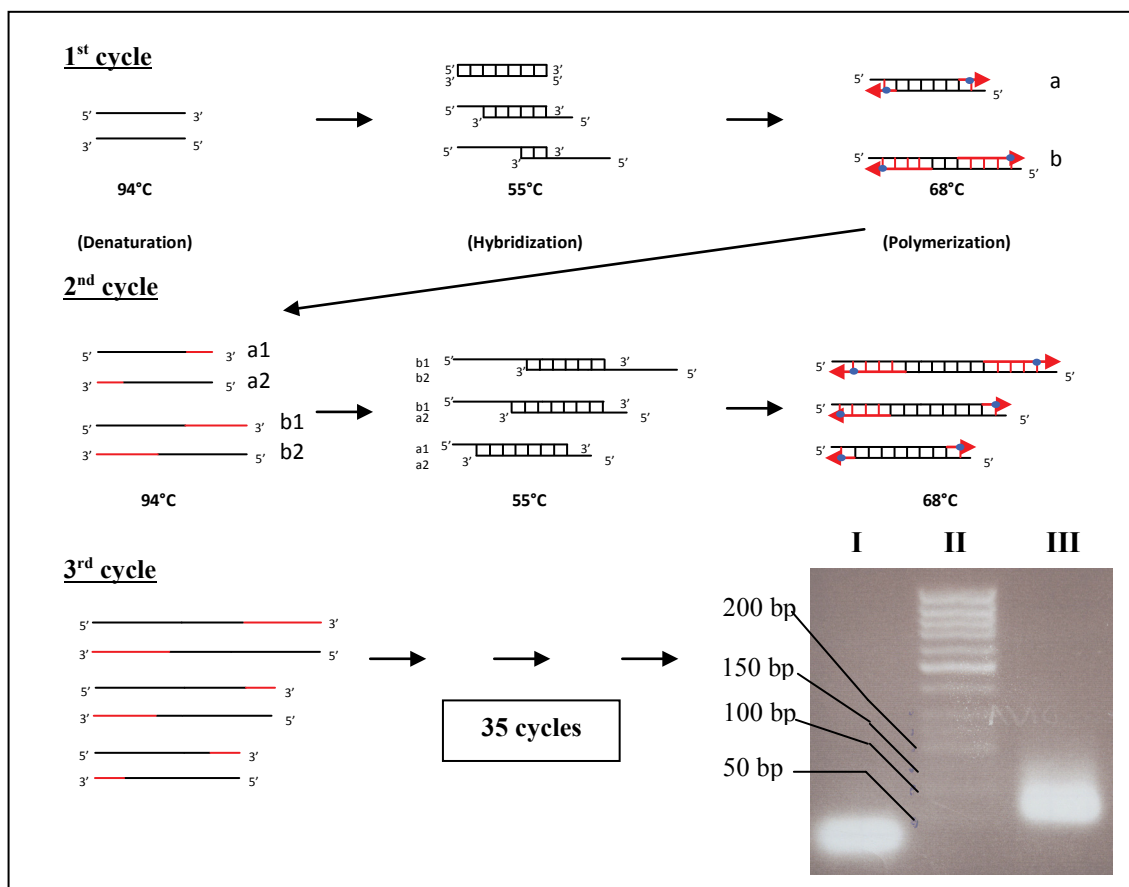


Figure 18: Overlapping PCR strategy to create the library of polyCAG double-strand DNA fragments of various lengths. The polyCAG library was generated by unconventional PCR and analyzed by 2% (w/v) agarose gel electrophoresis. **I**, oligonucleotides used for the PCR, 3'-(CAG)₁₃-5' and 3'-(CTG)₁₃-5'; **II**, DNA ladder; **III**, Double-strand polyCAG library.

By this method, we have created a library of variable-sized CAG repeats from 50 bp to more than 200 bp (see the agarose gel electrophoresis in Figure 18).

The polyCAG double-strand DNA fragments were then cloned into the gene of BlaP₁₉₇₋₁₉₈ carried by the expression vector pNY ESBlap (Materials and Methods, section 1.1). They were ligated into a SmaI restriction site previously introduced into the gene of BlaP at the location of interest, i.e. between codons for residues 197 and 198. The introduction of this restriction site leads to the addition of a proline-glycine (PG) dipeptide to the amino acid sequence of BlaP; these two amino acids will therefore flank the polyQ tract.

At the end, the expression vectors of different chimeric β -lactamases with 23, 30, 55 and 79 glutamines in position 197-198 were successfully constructed. These chimeras are referred to as BlaP₁₉₇₋₁₉₈(Gln)₂₃, BlaP₁₉₇₋₁₉₈(Gln)₃₀, BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉.

2. Production and purification of BlaP₁₉₇₋₁₉₈ and its polyQ chimeras

Proteins were expressed in the periplasm of *E. coli* cells (JM109). A 2 L culture was incubated one night at 37°C and the periplasmic proteins were then extracted by an osmotic shock. The recombinant proteins containing a C-terminal (His)₅ tag were purified by a single step affinity chromatography on a Ni-PDC column (Materials and Methods, section 2.). The enzymatically active fractions were then analyzed by SDS-PAGE and those containing more than 95 % of protein of interest were pooled and dialyzed 4 times against water before being lyophilized and stored at 4°C or 2 times against PBS (NaH₂PO₄/Na₂HPO₄ 50 mM, 150 mM NaCl), pH 7.5, before being frozen at -20°C. Typical yields of production and purification are presented for one production of BlaP₁₉₇₋₁₉₈, BlaP₁₉₇₋₁₉₈(Gln)₂₃, BlaP₁₉₇₋₁₉₈(Gln)₃₀, BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (Table 7). In general, the production/purification yields of BlaP₁₉₇₋₁₉₈ and its polyQ chimeras were similar, i.e. about 10-20 mg per liter of culture.

Analysis by SDS-PAGE of purified BlaP₁₉₇₋₁₉₈ and its polyQ-containing chimeras is shown in Figure 19. Although BlaP₁₉₇₋₁₉₈, BlaP₁₉₇₋₁₉₈(Gln)₂₃ and BlaP₁₉₇₋₁₉₈(Gln)₃₀ migrated at positions expected, according to the molecular mass markers, chimeras with either 55 or 79 glutamines show lower mobility and thus, greater apparent molecular masses than expected. Such anomalous behaviour, which was also observed for chimeras composed of myoglobin and polyQ sequences (Tobelman *et al.*, 2008), suggests that long polyQ sequences interfere with SDS binding.

Table 7: Purification of BlaP₁₉₇₋₁₉₈ and its polyQ chimeras (extracted from 2 L culture)

Protein (substrate)	Step	Total protein (mg)	¹ Specific activity ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	Total activity ($\mu\text{mol}\cdot\text{min}^{-1}$)	² Purification factor	³ Yield (%)
BlaP₁₉₇₋₁₉₈ (benzylpenicillin)	Periplasmic extract	106	1172	124818		
	Purified protein	19	5696	106629	4.86	85
BlaP₁₉₇₋₁₉₈(Gln)₂₃ (cephalothin)	Periplasmic extract	213	17	3648		
	Purified protein	27	107	2853	6.26	78
BlaP₁₉₇₋₁₉₈(Gln)₃₀ (cephalothin)	Periplasmic extract	147	33	4846		
	Purified protein	30	122	3625	3.7	75
BlaP₁₉₇₋₁₉₈(Gln)₅₅ (benzylpenicillin)	Periplasmic extract	110	2172	238986		
	Purified protein	30	7766	232990	3.57	97
BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cephalothin)	Periplasmic extract	207	29	6003		
	Purified protein	33.5	124	4154	4.28	69

¹The specific β -lactamase activity was measured using a 1mM benzylpenicillin solution for BlaP₁₉₇₋₁₉₈ and BlaP₁₉₇₋₁₉₈(Gln)₅₅ and using a 100 μM cephalothin solution for the other proteins (Materials and Methods, section 6). The use of two different substrates explains the large difference in the values obtained for the specific activity.

²The purification factor is equal to: (specific activity measured after purification)/(specific activity measured before purification).

³The yield is equal to: (total activity measured before purification)/(total activity measured after purification).

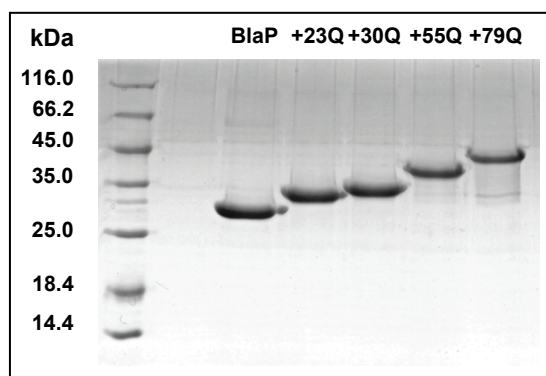


Figure 19: Purified BlaP₁₉₇₋₁₉₈ and polyQ-containing chimeras separated on a 15% (w/v) acrylamide gel and stained with coomassie blue. The first lane on the left shows protein markers with molecular masses as indicated, whereas the other lanes show BlaP₁₉₇₋₁₉₈ and the various chimeric proteins as indicated. Expected molecular masses are 30368, 33315, 34212, 37416 and 40491 Da, respectively, for BlaP₁₉₇₋₁₉₈ and chimeras with 23, 30, 55 and 79 glutamines.

The integrity of the proteins was checked by mass spectrometry (ESI-QTOF-MS). This analysis revealed the presence of three distinct peaks in the spectrum of each protein. Although in all cases, the main peak (estimated at > 70% of the population) corresponds to the theoretical mass deduced from the complete amino acid sequence, the two other peaks correspond to higher molecular mass species, i.e. +182 Da (10-20%) and +200 Da (< 10%). N-terminal sequencing indicated that the main peak corresponds to the enzyme with a well processed N-terminal sequence (T-E-M-K), whereas the peak at ~+200 Da corresponds to the

full-length protein plus the last two residues of the signal peptide (Q-A-T-E-M-K). The same N-terminal misprocessing was also observed for another BlaP chimera containing chitin-binding domain (ChBD) of the human macrophage chitotriosidase (Vandevenne *et al.*, 2007). The nature of the +182 Da adduct is not clear. Since the proportion of each molecular species is similar between various protein preparations, it is assumed that they do not interfere with the main conclusions of this work.

Size exclusion chromatography (SEC) analysis of protein preparations at high concentrations (50 μM and 120 μM) revealed that all proteins, particularly BlaP, form dimers although in small amount (at 120 μM , 5.6% for BlaP and < 1.5% for the four chimeras, Table 8 and Figure 20).

Table 8: Percentages of the different species observed by SEC

	[protein] (μM)	M (%)	D (%)	O (%)
BlaP₁₉₇₋₁₉₈	50	95.1	4.9	0
	120	94.4	5.6	0
BlaP₁₉₇₋₁₉₈(Gln)₂₃	50	99.4	0.6	0
	120	99.3	0.7	0
BlaP₁₉₇₋₁₉₈(Gln)₃₀	50	99.6	0.4	0
	120	99.5	0.5	0
BlaP₁₉₇₋₁₉₈(Gln)₅₅	50	98.9	1.1	0
	120	98.6	1.4	0
BlaP₁₉₇₋₁₉₈(Gln)₇₉	10	97.3	0.9	1.8
	50	96.4	1.1	2.5
	120	95.3	1.2	3.5

This analysis was carried out using a Superdex 200 GL 10/300 column using PBS buffer, pH 7.5. **M**, **D**, **O** stand for monomeric, dimeric and high molecular weight oligomeric species, respectively. The percentages given in the table correspond to: (peak area/total area in evaluated peaks)*100.

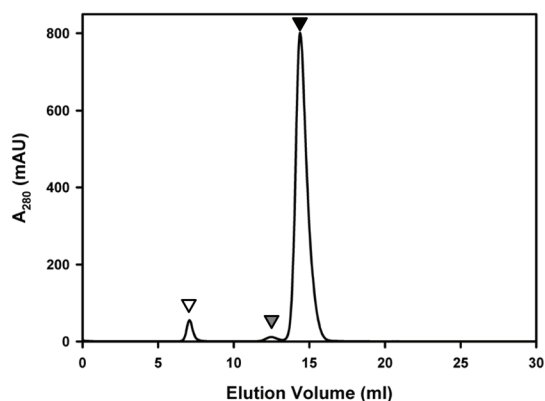


Figure 20 SEC analysis of BlaP₁₉₇₋₁₉₈(Gln)₇₉ at 120 μM in PBS pH 7.5; mAU, milli-absorbance unit. Monomeric, dimeric and high molecular weight species are indicated by black, grey and white arrows, respectively. The high molecular weight oligomeric species are eluted in the void volume of the column.

Minute quantities of high molecular weight species were also observed in BlaP₁₉₇₋₁₉₈(Gln)₇₉ solution (3.5% at 120 μ M and 2.5% at 50 μ M) (Table 8 and Figure 20). Interestingly, fractions containing these high molecular weight oligomers show some β -lactamase activity against nitrocefin (visual test) suggesting that within the oligomers, BlaP maintains its folded state although this hypothesis should be confirmed by other techniques such as circular dichroism. Based on their volume of elution from the column used (in the void volume) and assuming that these oligomeric species are spherical, their apparent molecular mass can be estimated to be higher than 600 kDa and thus formed by 15 or more monomers.

Decreasing the concentration of BlaP₁₉₇₋₁₉₈(Gln)₇₉ samples from 120 μ M to 10 μ M reduces the percentage of these oligomeric species to less than 2% (Table 8). Following the observation that in all cases, more than 95% of the protein population is monomeric at concentrations as high as 50 μ M, we assume that the presence of dimers and oligomers did not interfere with the determination of the catalytic and thermodynamic parameters, which were carried out with protein concentrations \leq 4.6 μ M.

3. Effects of polyQ insertions in position 197-198 on the catalytic properties of BlaP₁₉₇₋₁₉₈

The effects of polyQ insertions on the β -lactamase activity were investigated using cephalothin as the substrate. The complete hydrolysis of the cephalothin was followed by the decrease in the absorbance at 260 nm. The values of the catalytic parameters of the four chimeras are not significantly different from those of BlaP₁₉₇₋₁₉₈ (Table 9), suggesting that the structure in the environment of the active site was not affected by the insertion of polyQ tracts.

Table 9: Kinetic parameter values for BlaP₁₉₇₋₁₉₈ and its corresponding polyQ chimeras

	k_{cat} (s^{-1})	K_m (μ M)	k_{cat}/K_m ($mM^{-1}\cdot s^{-1}$)
BlaP₁₉₇₋₁₉₈	90 \pm 10	25 \pm 2.5	3600 \pm 400
BlaP₁₉₇₋₁₉₈(Gln)₂₃	120 \pm 10	32 \pm 4	3700 \pm 300
BlaP₁₉₇₋₁₉₈(Gln)₃₀	87 \pm 7	27 \pm 3	3100 \pm 200
BlaP₁₉₇₋₁₉₈(Gln)₅₅	107 \pm 10	31 \pm 3	3500 \pm 500
BlaP₁₉₇₋₁₉₈(Gln)₇₉	77 \pm 2	27 \pm 2	2900 \pm 100

Cephalothin was used as the substrate. Errors are given as standard deviations.

4. The polyQ tract adopts a disordered structure and does not perturb the overall structure of BlaP₁₉₇₋₁₉₈

The effects of the insertion in position 197-198 of 23, 30, 55 and 79 glutamines on the tertiary structure of BlaP₁₉₇₋₁₉₈ were investigated by both intrinsic fluorescence and near-UV circular dichroism (CD) measurements. Fluorescence and near-UV CD signals of proteins mainly originate from aromatic residues, principally tryptophan (Trp) residues and are influenced by the environment of these residues. In this work, fluorescence emitted comes entirely from the Trp residues since the excitation wavelength used was 295 nm.

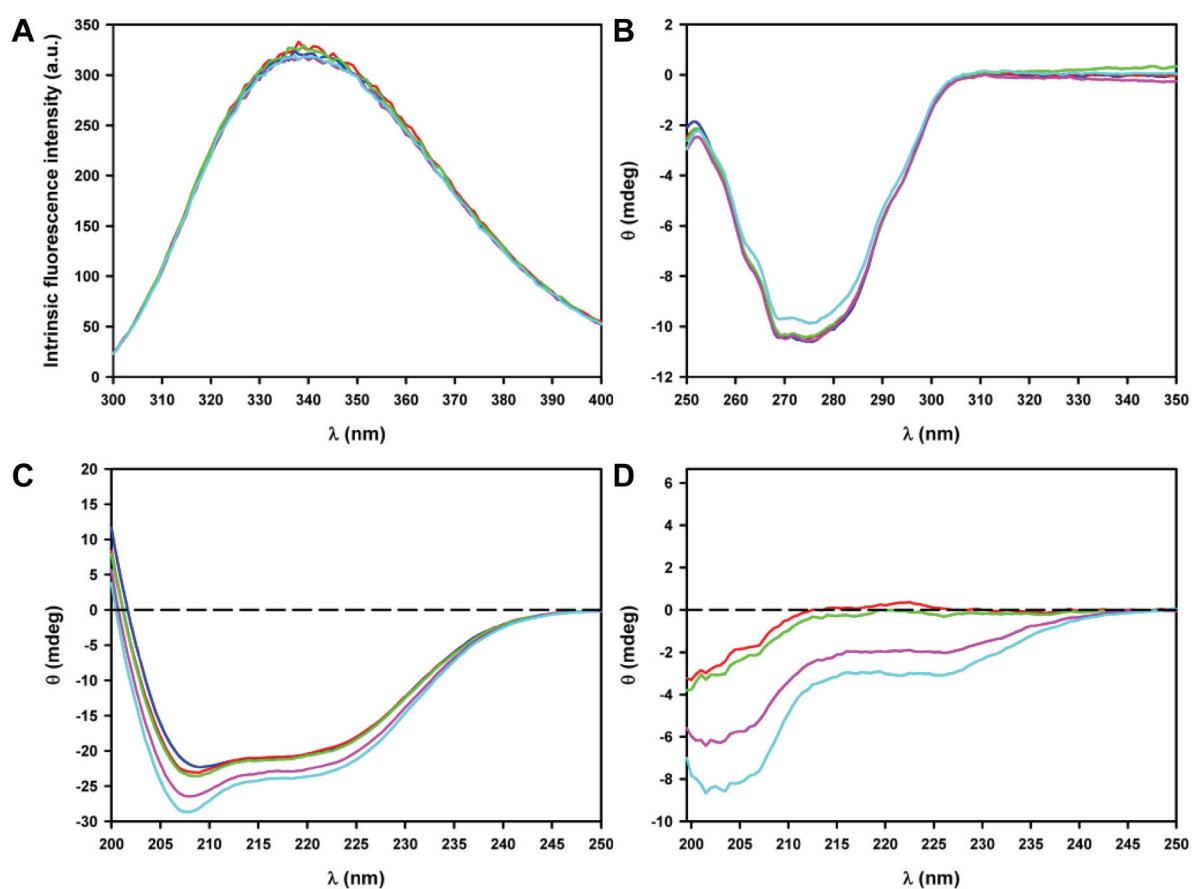


Figure 21: PolyQ insertions in position 197-198 have no effect on the overall structure of BlaP. (A) Intrinsic fluorescence, (B) near-UV CD and (C) far-UV CD spectra of BlaP₁₉₇₋₁₉₈ (blue) and the chimeras with 23 (red), 30 (green), 55 (pink) and 79 (cyan) glutamines inserted in position 197-198. a.u., arbitrary unit. (D) Difference spectra obtained by subtraction of the far-UV CD spectrum of BlaP₁₉₇₋₁₉₈ from that of each chimeric protein. Spectra were recorded at 25°C in PBS, pH 7.5, using protein concentrations of 4.6 μ M (fluorescence and far-UV CD) and 20 μ M (near-UV CD).

Intrinsic fluorescence and near-UV CD spectra of the chimeras are well superimposed to those of BlaP₁₉₇₋₁₉₈ (Figure 21A and B). These results suggest that the environment of the 3 Trp present in BlaP (Figure 22) is not modified by the insertion of glutamines in position 197-198.

This result is in good agreement with the observation that variable-sized polyQ insertions do not significantly modify the catalytic properties of the β -lactamase (Table 9).

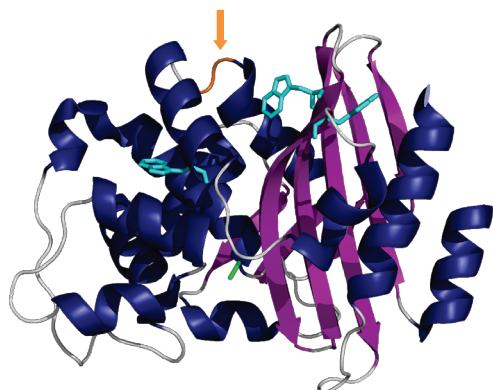


Figure 22: BlaP structure with Trp residues shown in cyan. The insertion site 197-198 is shown in orange

The effects of polyQ insertions on the secondary structure of BlaP₁₉₇₋₁₉₈ were examined by far-UV CD measurements. The CD signal of proteins in the far-UV region is dominated by contributions of the peptide bonds and gives information on the backbone structure of protein (i.e. their secondary structure). The far-UV CD spectra of BlaP₁₉₇₋₁₉₈ and its chimeras have a similar global shape but those of the chimeras are however characterized by a negative signal at ca. 207-208 nm, whose amplitude increases with the polyQ length (Figure 21C). The subtraction of the BlaP₁₉₇₋₁₉₈ spectrum from that of each chimera gives an indication of the structure adopted by the polyQ insert. These difference spectra display a negative peak with a minimum at ca. 202-203 nm, the amplitude of which increases with the number of glutamines (Figure 21D). This result suggests that the polyQ tract in position 197-198 does not significantly modify the secondary structure of BlaP₁₉₇₋₁₉₈ and that the polyQ tract, independently of the number of glutamines, adopts a disordered structure at the surface of the protein.

A similar absence of structural modification of the non-polyQ regions was observed when a polyQ tract is fused to the B domain of the protein A from *Staphylococcus aureus* (SpA) or apomyoglobin (Robertson *et al.*, 2008, Tobelmann and Murphy, 2011); however, radically different results were reported for inserted polyQ tracts. The most characterized chimeric proteins with inserted polyQ sequences were created from three all- α proteins: myoglobin (Tanaka *et al.*, 2001), SpA (Robertson *et al.*, 2008) and apomyoglobin (Tobelmann and Murphy, 2011). In these model proteins, the polyQ tract was inserted into a loop between α -helices (Table 5, page 27). The insertion of a polyQ tract into either myoglobin (Tanaka *et al.*, 2001) or SpA (Robertson *et al.*, 2008) induces changes in the tertiary structure of the host proteins, with no significant effect on their secondary structure. The insertion of at least 16Q

in apomyoglobin leads, however, to significant changes in secondary structure, besides a polyQ length-independent loss of tertiary structure (Tobelman and Murphy, 2011). This partial loss of structure was also observed when serine-glycine repeats were inserted at the same location, suggesting that it was not specific to polyQ insertion. As a likely consequence of these structural changes, the production yield for at least two of these model proteins (derived from SpA and apomyoglobin) was significantly lower than that of their respective wild-type counterparts, and this was shown to be directly related to the length of the polyQ tract inserted (Robertson *et al.*, 2008, Tobelman and Murphy, 2011). To date, the model protein with the longest inserted polyQ sequence reported in the literature consists of a chimeric myoglobin containing 50 glutamines only (Tanaka *et al.*, 2001). In contrast, the production yield of our BlaP chimeras remains unchanged (i.e. 10-20 mg per liter of culture), even when a 79-glutamine sequence is inserted at position 197-198. This makes BlaP an ideal scaffold to investigate the effects of long inserted polyQ sequences.

Our CD data suggest that the polyQ tracts of 23-79Q, inserted into the solvent-exposed loop of BlaP, adopt a disordered structure. This observation is in very good agreement with the results obtained with other model proteins containing variable-length polyQ tracts, e.g. (i) polyQ sequences fused either to SpA or GST (Klein *et al.*, 2007, Masino *et al.*, 2002, Robertson *et al.*, 2008) and (ii) the exon 1 of huntingtin fused to thioredoxin (Bennett *et al.*, 2002). In contrast, data obtained with myoglobin and apomyoglobin indicated that inserted polyQ sequences of sufficient length (*ca.* 16-28Q) could form β -structures (Tanaka *et al.*, 2001, Tobelman and Murphy, 2011). These findings suggest that the conformation adopted by a polyQ stretch is strongly context-dependent. The observation that a 79-residue long polyQ sequence is disordered when inserted into BlaP, suggests that it does not adopt any stable, aggregation-prone, β -sheet structure in solution prior to aggregation into amyloid-like fibrils. Because of the low sensitivity of the structural techniques used in this work, we cannot exclude, however, the possibility that the polyQ tract might rarely and/or transiently access more organized structures.

5. The polyQ tract destabilizes BlaP₁₉₇₋₁₉₈ and the extent of destabilization is largely independent of the polyQ length

To investigate the effects of polyQ inserts on the stability of BlaP₁₉₇₋₁₉₈, we carried out urea-induced equilibrium unfolding experiments. The loss of tertiary and secondary structures was followed by intrinsic fluorescence and far-UV CD measurements, respectively.

To ensure that the thermodynamic parameters could be determined, it is first necessary to demonstrate that: (i) the equilibrium is reached in all the samples before the fluorescence and CD measurements are carried out (i.e. after one night of incubation at 25°C) and (ii) the unfolding reaction is reversible. For these purposes, two sets of samples were prepared in urea concentrations ranging from 0.55 to 5.5 M as shown in the example presented in Figure 23.

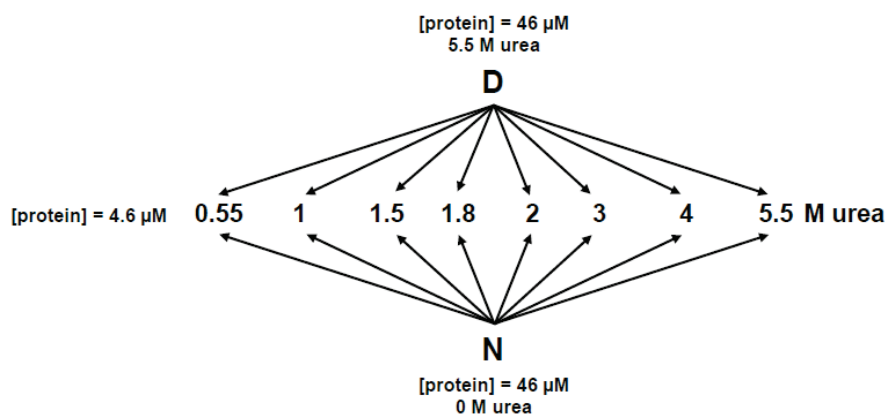


Figure 23: Scheme of samples preparation to demonstrate the reversibility and equilibrium conditions of the urea-unfolding experiment.

One series was prepared from the native protein and allows monitoring the unfolding process (lower part of the Figure 23). A protein sample unfolded in 5.5 M urea for 3 hours (under these conditions, BlaP₁₉₇₋₁₉₈ and all the chimeras were shown to be completely unfolded, data not shown) was used in the other series allowing to monitor the refolding process (upper part of the Figure 23). For each protein, the spectra obtained in the presence of 0.55 M urea for both the unfolding and refolding samples are superimposed (Figure 24, right column), indicating that the refolding was fully reversible. Moreover, the fluorescence data at 323 nm recorded with both sets of samples for each protein coincide well in the full range of urea concentrations tested (Figure 24, left column). This therefore indicates that urea-induced unfolding of BlaP₁₉₇₋₁₉₈ and its polyQ chimeras is fully thermodynamically reversible and that equilibrium is reached under the experimental conditions used. Far-UV CD signals at 222 nm of the two sets of samples were also measured and the data obtained also demonstrate that the equilibrium was reached and that the unfolding was fully reversible (Annexe 1).

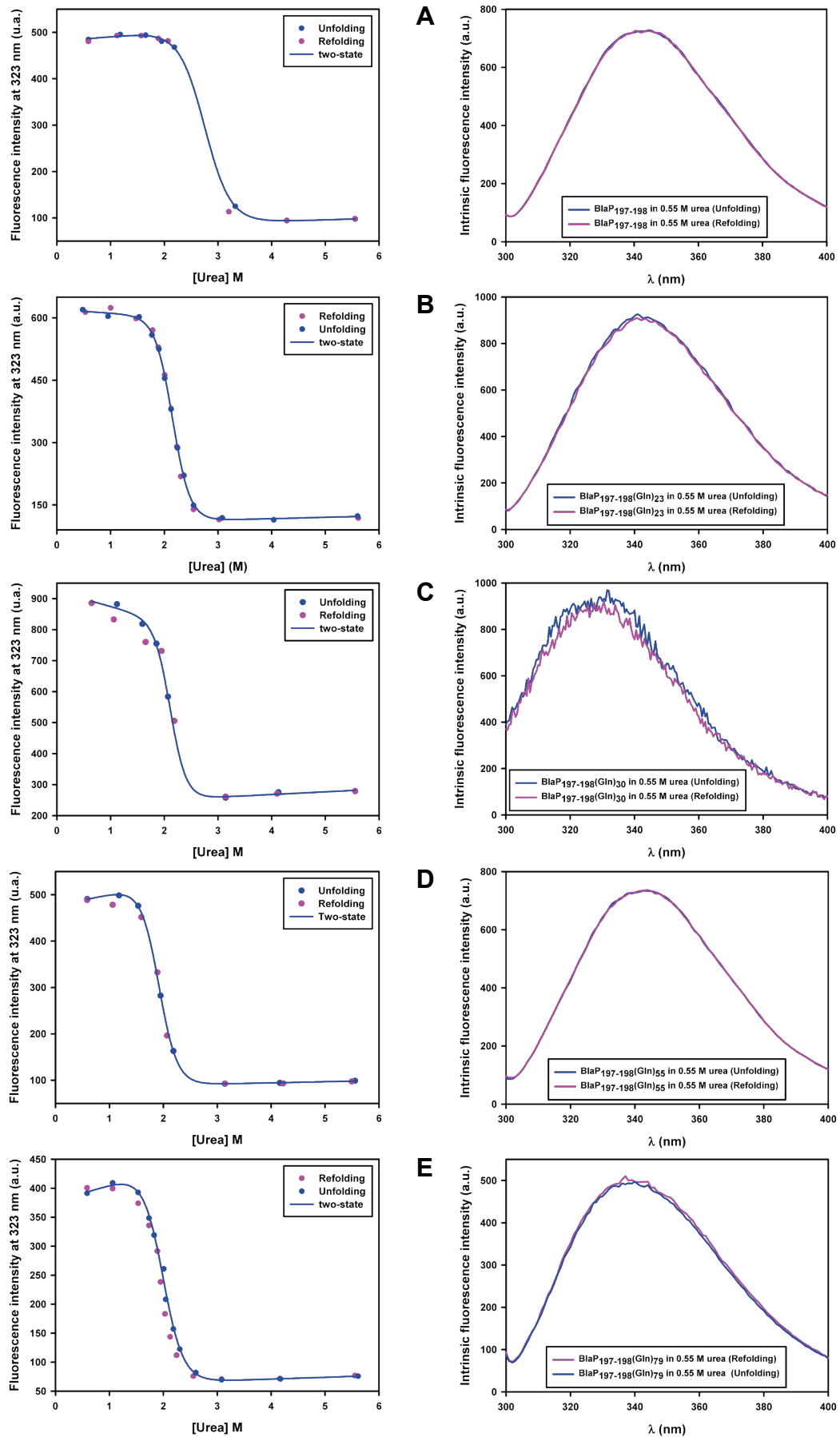


Figure 24: Unfolding (blue)/refolding (pink) transitions monitored by the change in fluorescence intensity at 323 nm (left column) and intrinsic fluorescence spectra (right column) of samples in 0.55 M urea from unfolding (blue) and refolding (pink) transitions. BlaP₁₉₇₋₁₉₈ (A) and chimeras with 23 (B), 30 (C), 55 (D) and 79Q (E). a.u., arbitrary unit.

Since both the reversibility and the equilibrium conditions were fulfilled, unfolding curves carried out with points at many more urea concentration (i.e. from 0 to 5.5 M by 0.1 M increments) were therefore performed in order to determine the values of the thermodynamic parameters. For each protein, single transitions were observed by both intrinsic fluorescence and far-UV CD measurements (as an example, transitions obtained for BlaP₁₉₇₋₁₉₈ are shown in Figure 25); the data were therefore analysed according to a two-state model using the equation 2 (Materials and Methods, section 9).

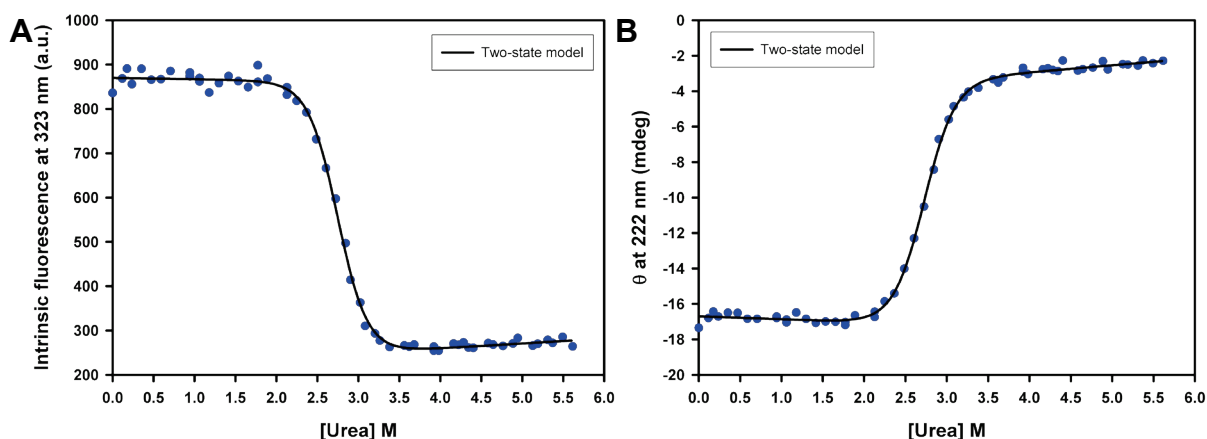


Figure 25: Urea-induced unfolding transitions obtained for BlaP₁₉₇₋₁₉₈ (4.6 μ M) in PBS, pH 7.5 and at 25°C and monitored by the change (A) in fluorescence intensity at 323 nm and (B) in ellipticity at 222 nm. a.u., arbitrary unit. The continuous line represents the fit of the equation of a two-state model to the experimental data.

Normalization of the data shows that the single transitions obtained by intrinsic fluorescence and far-UV CD coincide (Figure 26), indicating that the five proteins unfold according to a simple two-state mechanism, with no intermediate species significantly populated between the native and unfolded states.

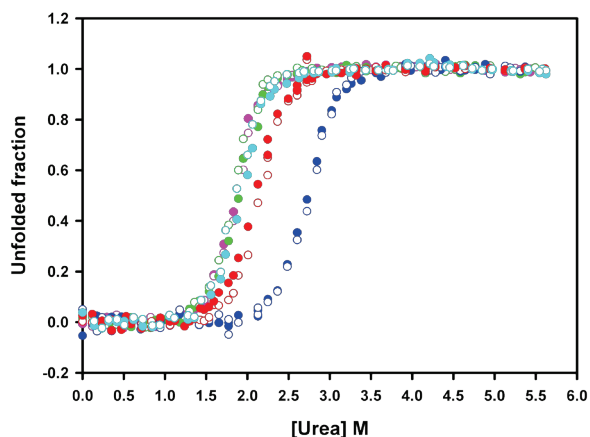


Figure 26: Normalized urea-induced unfolding transitions obtained for BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₂₃ (red), BlaP₁₉₇₋₁₉₈(Gln)₃₀ (green), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) at pH 7.5 and 25°C and monitored by the change in fluorescence intensity at 323 nm (filled circles) and in ellipticity at 222 nm (open circles). The protein concentration is 4.6 μ M.

The values of the thermodynamic parameters obtained by the two different spectroscopic techniques can be therefore averaged (Table 10). These results show that all four chimeras are destabilized with respect to the wild-type protein. Remarkably, the extent of destabilization is similar, i.e. 7.6-8.8 kJ·mol⁻¹, for all four chimeric proteins and thus is largely independent of the length of the polyQ tract.

Table 10: Thermodynamic parameters of unfolding of BlaP₁₉₇₋₁₉₈ and polyQ chimeras 197-198

	$\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$ (kJ·mol ⁻¹)	m_{NU} (kJ·mol ⁻¹ ·M ⁻¹)	C_m (M)	T_m^{app} (°C)
BlaP₁₉₇₋₁₉₈	37.0 ± 1.8	13.5 ± 0.6	2.74 ± 0.23	58.5 ± 0.35
BlaP₁₉₇₋₁₉₈(Gln)₂₃	29.3 ± 2.4	13.7 ± 0.9	2.13 ± 0.19	53.9 ± 0.7
BlaP₁₉₇₋₁₉₈(Gln)₃₀	29.4 ± 1.2	15.9 ± 0.7	1.85 ± 0.15	53.7 ± 0.3
BlaP₁₉₇₋₁₉₈(Gln)₅₅	29.0 ± 0.6	15.6 ± 0.6	1.86 ± 0.15	53.5 ± 0.6
BlaP₁₉₇₋₁₉₈(Gln)₇₉	28.2 ± 1.3	14.9 ± 0.6	1.89 ± 0.17	53.2 ± 0.3

The thermodynamic parameters were deduced from the analysis of unfolding transitions monitored by intrinsic fluorescence and far-UV CD measurements. The urea unfolding experiments were carried out at 25°C. $\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$, m_{NU} , C_m and T_m^{app} values result from the mean of the values obtained by far-UV CD and intrinsic fluorescence. Thermal unfolding was not fully reversible and only the apparent mid-unfolding temperature (T_m^{app}) was determined.

Thermal unfolding of BlaP₁₉₇₋₁₉₈ and its chimeras was also monitored by both intrinsic fluorescence and far-UV CD measurements and the curves obtained for each protein by both techniques are simple two-state transitions; the data obtained for BlaP₁₉₇₋₁₉₈ are shown in Figure 27 as a representative example.

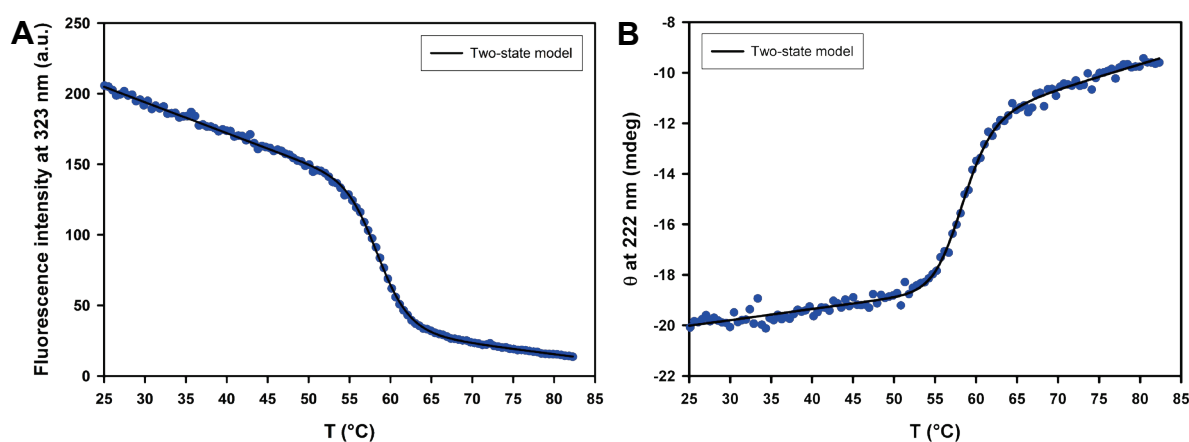


Figure 27: Heat-induced unfolding transitions obtained for BlaP₁₉₇₋₁₉₈ (4.6 μM) in PBS, pH 7.5 and monitored by the change (A) in fluorescence intensity at 323 nm and (B) in ellipticity at 222 nm. a.u., arbitrary unit. The continuous line represents the fit of the equation of a two-state model to the experimental data.

Although this process was not reversible (data not shown) for any of the proteins, data were analyzed based on a two-state model to determine the apparent temperatures of mid-transition (T_m^{app}) (Figure 27).

The normalized data obtained by both fluorescence and far-UV CD coincide (Figure 28) and T_m^{app} values were therefore averaged (Table 10).

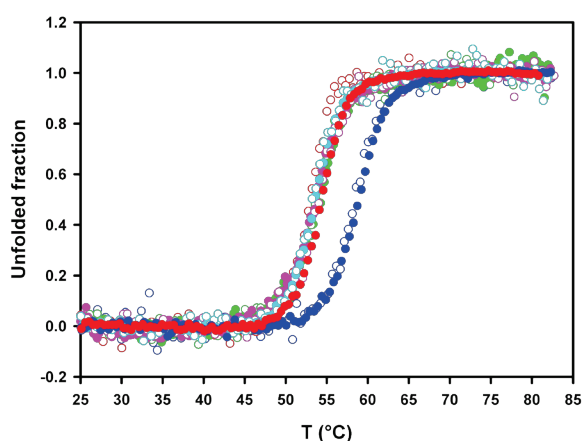


Figure 28: Normalized heat-induced unfolding transitions obtained for BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₂₃ (red), BlaP₁₉₇₋₁₉₈(Gln)₃₀ (green), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) at pH 7.5 and monitored by the change in fluorescence intensity at 323 nm (filled circles) and in ellipticity at 222 nm (open circles). The protein concentration is 4.6 μ M.

In good agreement with urea-unfolding data, all chimeras are less thermo-stable than the wild-type protein ($\Delta T_m^{app} = 4.5 - 5.3^\circ\text{C}$) and the extent of destabilization is independent of the number of glutamines.

The fact that the extent of BlaP destabilization is independent of the number of glutamines contrasts with observations on other model systems reporting: (i) no destabilization at all, e.g. the protein moiety SpA when fused to a polyQ tract up to 52 residues in length (Robertson *et al.*, 2008), (ii) selective destabilization, e.g. the cellular retinoic acid binding protein I (CRABPI) moiety when fused to Htt exon 1 containing polyQ stretches longer than a threshold size (Ignatova and Gierasch, 2006), (iii) cumulative destabilization, e.g. chymotrypsin inhibitor 2 (CI2) (Ladurner and Fersht, 1997), myoglobin (Tanaka *et al.*, 2001) and SpA (Robertson *et al.*, 2008) containing inserted polyQ tracts, where the extent of destabilization increases with increasing polyQ length, independently of a threshold. Finally, a polyQ-length independent destabilization was also described for apomyoglobin containing an inserted polyQ tract (Tobelman and Murphy, 2011) but this effect does not seem specific to the glutamines, since a similar destabilization was observed when serine-glycine repeats were inserted at the same location. In the case of BlaP, insertion of the 73-residue long chitin-binding domain of human macrophage chitotriosidase (Vandevenne *et al.*, 2007) destabilizes the host protein significantly less than the polyQ tracts (i.e. $3.2 \text{ kJ}\cdot\text{mol}^{-1}$ vs $7.6\text{-}8.8 \text{ kJ}\cdot\text{mol}^{-1}$).

This suggests that the extent of destabilization of BlaP depends more on the nature of the polypeptide inserted than its length, and that the insertion of the disordered polyQ stretch is more destabilizing than that of the folded chitin-binding domain of human macrophage chitotriosidase. Taken together, these data show that the effects of polyQ insertion on the stability of the host protein vary greatly depending on the properties of the latter.

The observation that the chimeric β -lactamases with 30, 55 and 79 glutamines are destabilized to the same extent relative to wild-type BlaP₁₉₇₋₁₉₈ (Table 10) gave us the unique opportunity to investigate independently the influence of i) the length of the polyQ sequence and ii) the conformational properties of the β -lactamase moiety, on the aggregating properties of the chimeras. Accordingly, the aggregation of BlaP₁₉₇₋₁₉₈ and the different chimeras was investigated under both native and denaturing conditions.

6. The threshold length of the polyQ tract above which chimeras aggregate into amyloid-like fibrils depends on the structural integrity of BlaP₁₉₇₋₁₉₈

6.1. Aggregation under denaturing conditions

The aggregating properties of the proteins were first investigated in the presence of 1.85 M urea (in PBS, pH 7.5) and at 25°C. Data in Figure 26 and Table 10 indicate that under these conditions, all BlaP₁₉₇₋₁₉₈ molecules are native, whereas ca. 18% and ca. 50% of, respectively, BlaP₁₉₇₋₁₉₈(Gln)₂₃ and the three other chimeric enzymes are unfolded. The kinetics of aggregation were monitored by measuring the amount of protein that remained soluble after different incubation times. BlaP₁₉₇₋₁₉₈, BlaP₁₉₇₋₁₉₈(Gln)₂₃ and BlaP₁₉₇₋₁₉₈(Gln)₃₀ display little, if any, tendency to aggregate, even after ca. 720 hours of incubation, whereas both BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ readily aggregate (Figure 29A). Aggregation of BlaP₁₉₇₋₁₉₈(Gln)₅₅ is characterized by a lag phase (ca. 50 hours) while very fast aggregation with no discernable lag phase is observed with BlaP₁₉₇₋₁₉₈(Gln)₇₉. The lag phase is consistent with the nucleation-polymerization mechanism that has been proposed for amyloid fibril formation (Harper and Lansbury, 1997) and corresponds to the nucleation phase. After ca. 720 and ca. 330 hours of incubation, fibrillar aggregates (Figure 29C) are clearly visible by TEM for BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉, respectively. In contrast, even after 720 hours of incubation, only minute amounts of amorphous aggregates, but no fibrils, were observed with BlaP₁₉₇₋₁₉₈, BlaP₁₉₇₋₁₉₈(Gln)₂₃ and BlaP₁₉₇₋₁₉₈(Gln)₃₀ (Figure 29C).

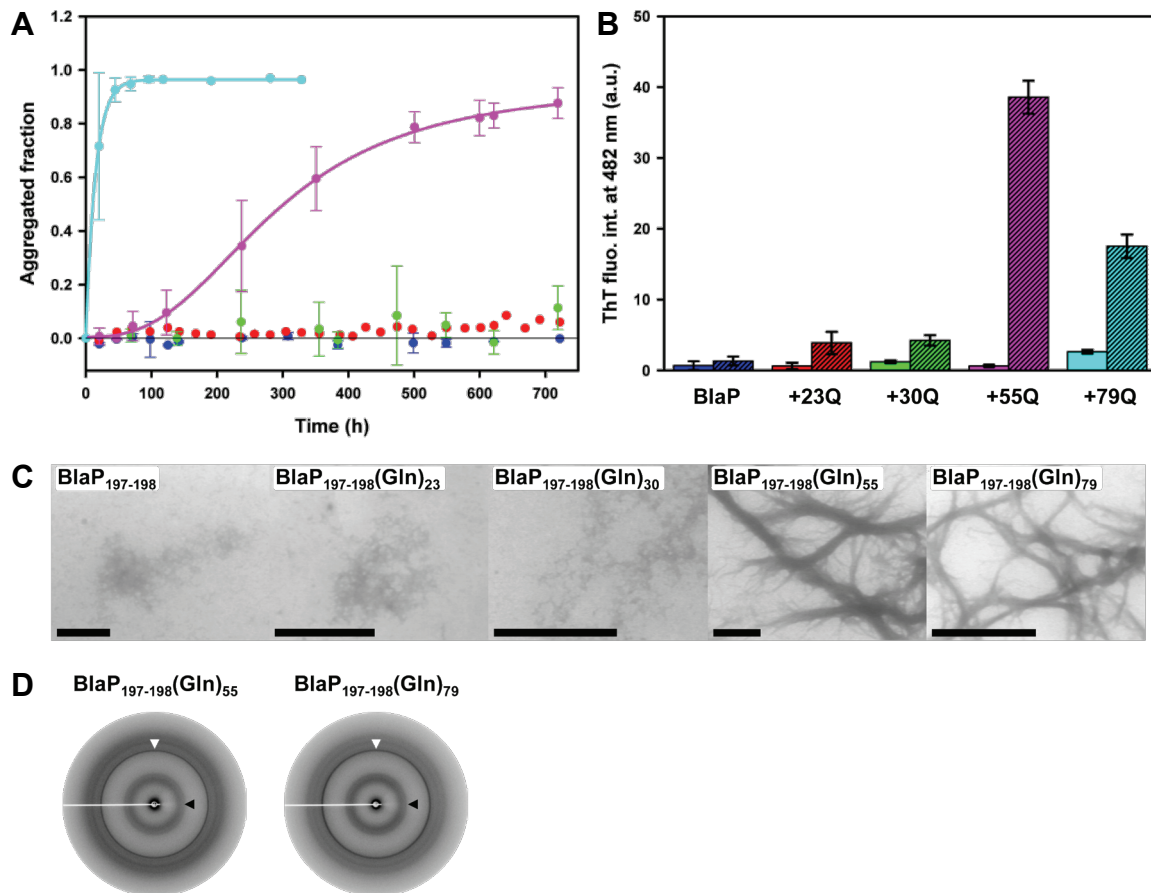


Figure 29: Only BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ form amyloid-like fibrils when incubated in the presence of 1.85 M urea. (A) Aggregation kinetics of 110 μ M BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₂₃ (red), BlaP₁₉₇₋₁₉₈(Gln)₃₀ (green), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) at 25°C in the presence of 1.85 M urea in PBS, pH 7.5, followed by measuring the concentration of protein remaining soluble. Time-points shown with an error bar are the average of three independent time-courses for BlaP₁₉₇₋₁₉₈(Gln)₃₀, BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ and two independent time-courses for BlaP₁₉₇₋₁₉₈. Error bars show the standard deviations. Only one time-course was carried out with BlaP₁₉₇₋₁₉₈(Gln)₂₃. (B) ThT fluorescence intensities at 482 nm in the presence of BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₂₃ (red), BlaP₁₉₇₋₁₉₈(Gln)₃₀ (green), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) samples at T₀ (solid bars) and T_f (Dashed bars). T₀ and T_f correspond to the initial and final points of one time-course for each protein. Data are the average of three measurements and error bars represent the standard deviations. a.u., arbitrary unit. (C) TEM images of the protein samples at T_f. The scale bar is 1 μ m. (D) X-ray fibre diffraction patterns from BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ fibrils. White and black arrows indicate meridional and equatorial reflections at 4.7 Å and ca. 9.5 Å, respectively.

The aggregates formed by BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ after respectively, 720 and 330 hours of incubation significantly bind ThT (Figure 29B), suggesting that the fibrils observed are amyloid-like. The degree of ThT binding observed for BlaP₁₉₇₋₁₉₈(Gln)₇₉ is, however, significantly lower than that observed for BlaP₁₉₇₋₁₉₈(Gln)₅₅, a surprising observation considering that BlaP₁₉₇₋₁₉₈(Gln)₇₉ aggregates to a higher extent than its 55-glutamine homolog (Figure 29A). The significantly lower ThT binding to BlaP₁₉₇₋₁₉₈(Gln)₇₉ fibrils may be due to an alternative overall packing of the aggregates. This is consistent with the observation that aggregates formed with BlaP₁₉₇₋₁₉₈(Gln)₇₉ were clearly larger and more difficult to resuspend in the ThT solution. For BlaP₁₉₇₋₁₉₈(Gln)₇₉, the ThT fluorescence measurements were

performed with samples incubated for *ca.* 330 hours and since the protein is fully aggregated after 100 hours, the aggregates could mature (e.g. laterally associate) during the last 200 hours of incubation. In the case of BlaP₁₉₇₋₁₉₈(Gln)₅₅, the aggregation is much slower and there is therefore less time for fibril maturation until the final point of the time-course (i.e. when samples were taken for ThT fluorescence measurements at *ca.* 720 hours) (Figure 29A). Aggregates formed by BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ were analyzed by X-ray fibre diffraction. Both the diffraction patterns obtained show two reflections, a sharp one at 4.7 Å and a broad and more diffuse one at ~9.5 Å (Figure 29D). Despite difficulties encountered in aligning the fibrils, these reflections are consistent with a cross-β structure characteristic of amyloid fibrils and reflect the distance between the β-strands in each of the sheets of the amyloid protofilament (4.7 Å) and the spacing between the β-sheets (~9.5 Å) (Serpell *et al.*, 1999).

These results show that, although the three chimeras BlaP₁₉₇₋₁₉₈(Gln)₃₀, BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ are equally destabilized under the conditions used to monitor aggregation (i.e. 50% of the molecules of each protein are unfolded at 25°C in the presence of 1.85 M of urea), only those with 55 and 79 glutamines significantly aggregate into amyloid-like fibrils within the time scale of the experiment. Hence, independently of the extent of destabilization, there seems to be a minimal length (i.e. a threshold) of the polyQ tract beyond which chimeric β-lactamases readily aggregate into amyloid-like fibrils. This value is comprised between 30 and 55 residues, which is reminiscent of the threshold number of glutamines above which other model proteins fused to a polyQ tract aggregate both *in vivo* and *in vitro* (Ignatova and Gierasch, 2006, Scherzinger *et al.*, 1997) and, more importantly, of the pathological threshold observed in polyQ diseases (Robertson and Bottomley, 2010).

BlaP₁₉₇₋₁₉₈ and all its chimeras (with 23 to 79 Q) were also incubated at 25°C in the presence of 3.5 M urea, a concentration at which all proteins are completely unfolded (Figure 26). Under these conditions, again, only BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ noticeably aggregate into fibrils that possess characteristics of amyloid such as ThT binding and cross-β structure (Figure 30A-D). Although BlaP₁₉₇₋₁₉₈, BlaP₁₉₇₋₁₉₈(Gln)₂₃ and BlaP₁₉₇₋₁₉₈(Gln)₃₀ remain soluble, small amounts of amorphous aggregates are observed by TEM; these aggregates do not bind ThT (Figure 30A-C).

Taken together, the results of these two experiments give a clear indication that the unfolding of the β-lactamase moiety is not the driving force underlying the fibrillar aggregation.

Rather, it is purely the expansion of the polyQ tract above a threshold length that promotes the formation of amyloid-like fibrils. Moreover, in the presence of 1.85 M or 3.5 M urea, the chimera with 79 glutamines aggregates faster than that containing 55 glutamines. Such a polyQ length-dependent rate of aggregation has been observed in all *in vitro* studies of polyQ peptides (Bhattacharyya *et al.*, 2006, Chen *et al.*, 2001) and different protein systems (Ignatova and Gierasch, 2006, Robertson *et al.*, 2008), and can be related to the so called “anticipation phenomenon” characteristic of polyQ diseases (Zoghbi and Orr, 2000).

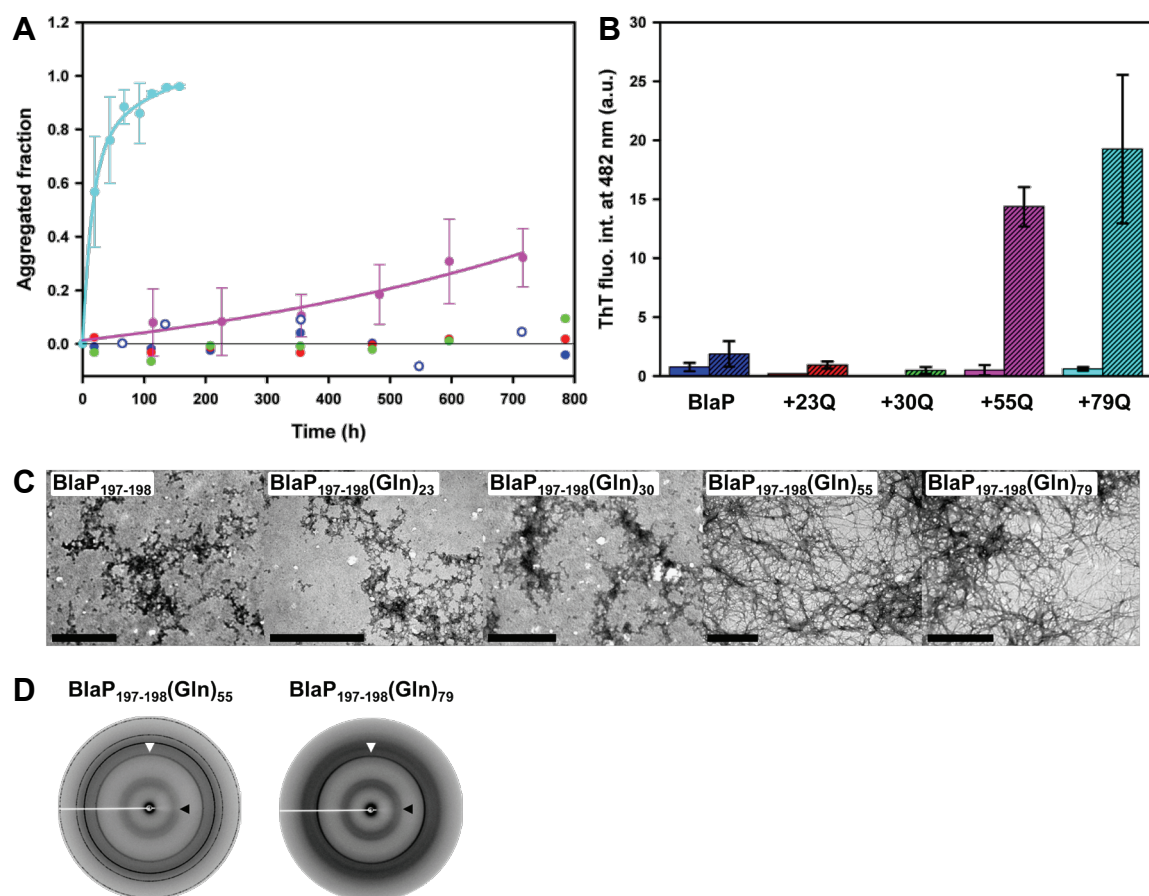


Figure 30: The unfolding of the β -lactamase moiety is not the critical factor that triggers the aggregation process.

(A) Aggregation kinetics of 110 μ M BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₂₃ (red), BlaP₁₉₇₋₁₉₈(Gln)₃₀ (green), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) at 25°C in the presence of 3.5 M urea in PBS pH 7.5, followed by measuring the concentration of protein remaining soluble. Time-points shown with an error bar are the average of three independent time-courses for BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉. The error bars show the standard deviations. For BlaP₁₉₇₋₁₉₈, two independent experiments were carried out (filled and open blue circles); however, since the times at which samples were taken differ from one time-course to the other, the data could not be averaged. For BlaP₁₉₇₋₁₉₈(Gln)₂₃ and BlaP₁₉₇₋₁₉₈(Gln)₃₀, only one time-course was carried out. (B) ThT fluorescence intensities at 482 nm in the presence of BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₂₃ (red), BlaP₁₉₇₋₁₉₈(Gln)₃₀ (green), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) samples at T_0 (solid bars) and at T_f (dashed bars). T_0 and T_f correspond to the initial and final points of one time-course for each protein. Data are the average of three measurements and error bars represent the standard deviations. a.u., arbitrary unit. (C) TEM images of the protein samples at T_f . The scale bar is 1 μ m. (D) X-ray fibre diffraction patterns from BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ fibrils. White and black arrows indicate meridional and equatorial reflections at 4.7 \AA and ca. 9.5 \AA , respectively.

6.2. Aggregation under native conditions

Finally, the kinetics of aggregation of BlaP₁₉₇₋₁₉₈, BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ were monitored under conditions favouring the native state (PBS, pH 7.5, 37°C) (Figure 28).

As observed under destabilizing conditions (Figures 29A and 30A), only chimeras with 55 and 79 glutamines aggregate, whereas BlaP₁₉₇₋₁₉₈ remains soluble throughout the duration of the experiment (Figure 31A). The kinetics of aggregation of BlaP₁₉₇₋₁₉₈(Gln)₇₉ was similar to that obtained in the presence of urea (Figures 29A and 30A) and the aggregates formed significantly bind ThT (Figure 31B), display a fibrillar morphology (Figure 31C), and exhibit a X-ray fibre diffraction pattern consistent with a cross- β structure (Figure 31D). These observations are all indicative of amyloid-like fibril formation by BlaP₁₉₇₋₁₉₈(Gln)₇₉ under native conditions. In contrast, BlaP₁₉₇₋₁₉₈(Gln)₅₅ aggregates to a lesser extent than in the presence of 1.85 M urea, and the resulting species bind ThT only weakly (Figure 31B), and appear amorphous when viewed by TEM (Figure 31C).

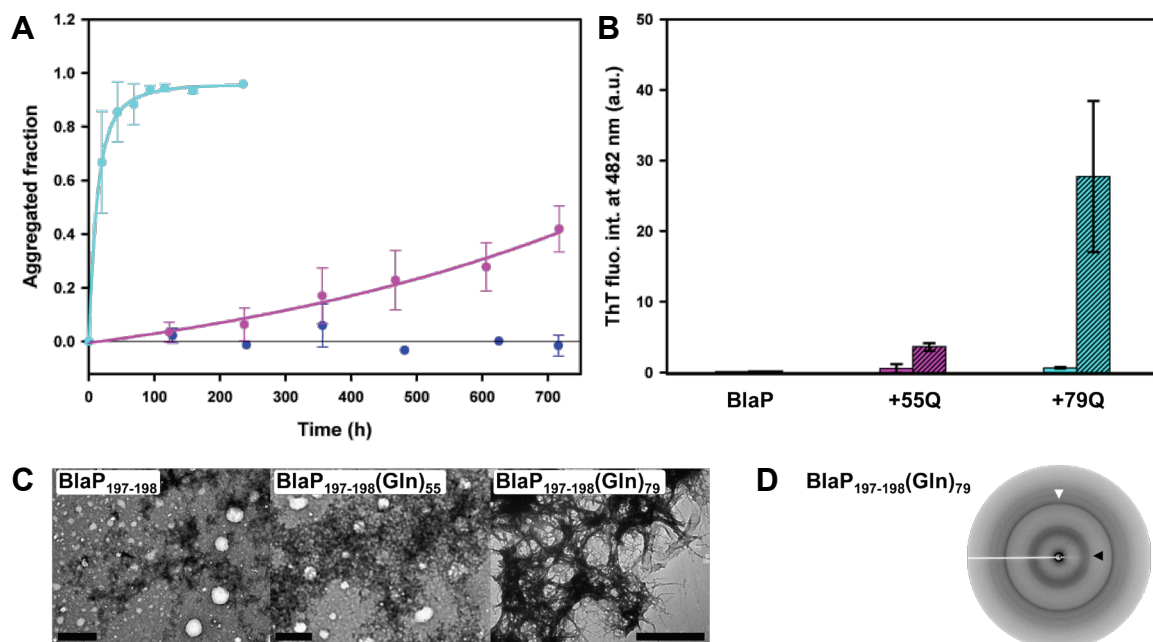


Figure 31: Only BlaP₁₉₇₋₁₉₈(Gln)₇₉ form amyloid-like fibrils when incubated in native conditions. (A) Aggregation kinetics of 110 μ M BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) at 37°C in PBS, pH 7.5 followed by measuring the concentration of protein remaining soluble. Time-points shown with an error bar are the average of three independent time-courses for BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ and two independent time-courses for BlaP₁₉₇₋₁₉₈. Error bars show the standard deviations. (B) ThT fluorescence intensities at 482 nm in the presence of BlaP₁₉₇₋₁₉₈ (blue), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (cyan) samples at T₀ (solid bars) and at T_f (Dashed bars). T₀ and T_f correspond to the initial and final points of one time-course for each protein. Data are the average of three measurements and error bars represent the standard deviations. ThT fluorescence intensities in the presence of BlaP₁₉₇₋₁₉₈ samples are weak and for more visibility, error bars (which are equal or less than 0.3) are not shown. a.u., arbitrary unit. (C) TEM images of the protein samples at T_f. The scale bar is 1 μ m. (D) X-ray fibre diffraction patterns from BlaP₁₉₇₋₁₉₈(Gln)₇₉ fibrils at T_f. White and black arrows indicate meridional and equatorial reflections at 4.7 Å and ca. 9.5 Å, respectively.

These aggregates did not mature into amyloid-like fibrils upon further incubation of up to 1500 hours (data not shown). Similarly, only small, disperse and amorphous aggregates that do not bind ThT are visible by TEM in BlaP₁₉₇₋₁₉₈ samples at the end of the experiment (Figure 31B and C).

These results show that under native conditions, only the chimeric β -lactamase with 79 glutamines forms amyloid-like fibrils. Our data also show that under native conditions, the polyQ threshold size for aggregation into fibrillar aggregates is higher than that observed in denaturing conditions, suggesting that its value depends on the structural integrity of the β -lactamase moiety. Indeed, BlaP₁₉₇₋₁₉₈(Gln)₅₅ is only able to form amyloid-like aggregates when BlaP₁₉₇₋₁₉₈ is unfolded (Figure 32A) whereas, remarkably, BlaP₁₉₇₋₁₉₈(Gln)₇₉ aggregates into amyloid-like fibrils at the same rate irrespective of the incubation conditions and, therefore, the conformation of BlaP₁₉₇₋₁₉₈ (Figure 32B).

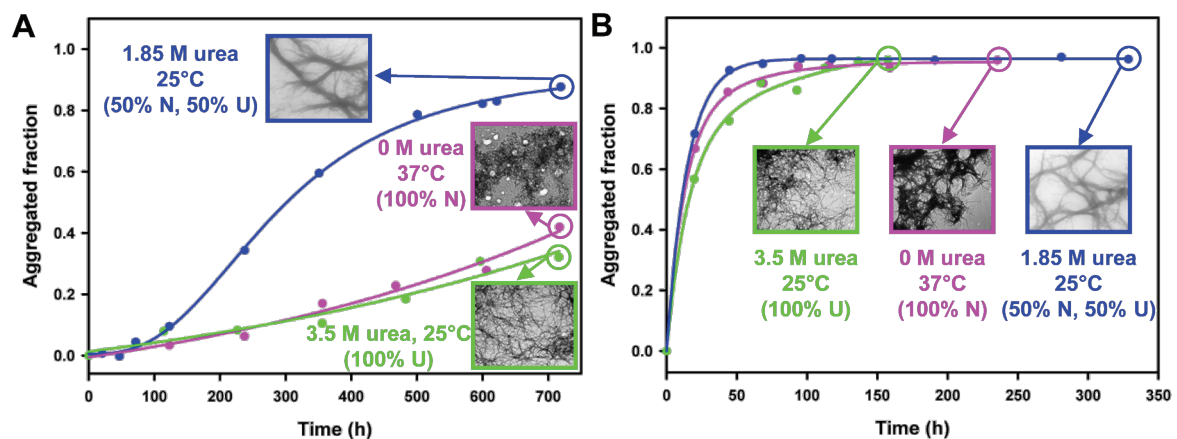


Figure 32: The ability of BlaP₁₉₇₋₁₉₈(Gln)₇₉ to form amyloid-like fibrils does not depend on the conformation of BlaP. Comparison between the aggregation kinetics and the morphology of the aggregates obtained with BlaP₁₉₇₋₁₉₈(Gln)₅₅ (A) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (B) in the following conditions of incubation; (i) PBS, pH 7.5 and 0 M urea at 37°C (pink), (ii) PBS, pH 7.5 and 1.85 M urea at 25°C (blue) and (iii) PBS, pH 7.5 and 3.5 M urea at 25°C (green). N is the native state and U is the unfolded state. The data correspond to those shown in Figures 29A, 30A and 31A without error bars.

This observation shows that the presence of folded BlaP₁₉₇₋₁₉₈ suppresses the intrinsic propensity of the 55-glutamine tract to trigger the self-association of the chimera into amyloid-like fibrils. In other words, in the presence of 55Q, the protective role of BlaP₁₉₇₋₁₉₈ in polyQ-triggered self-association into amyloid-like aggregates depends on the integrity of BlaP₁₉₇₋₁₉₈ structure. In contrast, the conformational state of BlaP₁₉₇₋₁₉₈ has no discernable effect on the ability of the 79-glutamine tract to promote the formation of amyloid-like fibrils.

The dependence of BlaP₁₉₇₋₁₉₈(Gln)₅₅ aggregation on the structural integrity of BlaP can result from both steric and conformational constraints. The folded β -lactamase moiety may limit the accessibility of the 55-glutamine tract via steric hindrances and thus abrogate its propensity to form the highly ordered intermolecular β -sheets characteristic of amyloid fibrils. The 55-glutamine tract in the presence of folded BlaP leads, nevertheless, to less organized intermolecular interactions and thus to the formation of amorphous aggregates. The unfolding of BlaP can however render the 55Q tract more accessible and thus permits the formation of amyloid-like fibrils. In contrast, when expanded to 79 glutamines, the polyQ tract is similarly accessible to interact with other monomers to form amyloid-like fibrils whether BlaP is denatured or not. The need to unfold BlaP in order to generate amyloid-like fibrils from the 55Q chimera is reminiscent of that of proteolysis observed with some proteins associated with polyQ diseases. For example, nuclear inclusions from HD patients have been shown to contain Htt fragments including the expanded polyQ tract rather than the full-length protein (DiFiglia *et al.*, 1997). It has been proposed that these fragments have a higher tendency to self-associate and thus may be crucial to initiate the aggregation phenomenon (Cooper *et al.*, 1998, Martindale *et al.*, 1998). Their greater propensity to aggregate is probably due to the removal by proteolysis of some of the steric hindrances otherwise exerted by the non-polyQ domains.

The protective effects mediated by the folded state of BlaP could also be due to the imposition of some conformational constraints on the polyQ tract. In a similar vein of what has been suggested by a number of experimental and computational studies (Bhattacharyya *et al.*, 2006, Crick *et al.*, 2006, Lakhani *et al.*, 2010, Vitalis *et al.*, 2007), we propose that the polyQ tract which extends from the solvent-exposed loop in BlaP exists as an ensemble of heterogeneous disordered conformations, in dynamic equilibrium with rare more structured conformational species, some of which are competent for amyloid fibril formation. The longer the polyQ repeat, the more dynamic it is and the wider the conformational space it can sample. As a consequence, more aggregation-prone conformations can be transiently sampled by longer polyQ tracts, thus explaining their higher propensity to aggregate. The length-dependent frequency at which the amyloid-competent conformations are visited, and thus the propensity of the polyQ tract to aggregate into amyloid-like fibrils, is further modulated by the conformational state of the β -lactamase moiety. Short polyQ stretches (≤ 30 residues) rarely or never access amyloid-competent conformations, even if the β -lactamase is in its unfolded state. In contrast, such conformations are equally sampled by the long polyQ tracts

(≥ 79 residues) regardless of whether BlaP is native or unfolded. Finally, for polyQ stretches of intermediate lengths (i.e. 30-79 residues), the frequency at which amyloid-competent conformations are adopted critically depends on the structural integrity of BlaP. For instance, the presence of folded BlaP prevents the 55-glutamine tract from accessing conformational precursors to amyloid-like fibril formation. Conversely, the unfolding of the β -lactamase moiety would allow the 55-glutamine tract to interconvert more freely between various conformations and hence access, more frequently, conformations that are competent for amyloid-like fibril formation. The probability of accessing these conformations is however lower for the 55- than the 79-glutamine tract, and thus a lag phase is observed for BlaP₁₉₇₋₁₉₈(Gln)₅₅ fibril formation. These results are consistent with previous reports showing the importance of the constraints imposed by adjacent sequences on the polyQ tract (Bhattacharyya *et al.*, 2006, Darnell *et al.*, 2007). For example, it was reported that the C-terminal addition of 10 or 11 prolines to polyQ peptides tends to decrease both aggregation rate and aggregate stability, and increases the threshold for fibril formation by disfavoring aggregation-competent conformations (Bhattacharyya *et al.*, 2006, Darnell *et al.*, 2007). The study of Darnell *et al.* strongly suggests that it is the formation of constraining polyproline type II helices by the C-terminal prolines that tips the balance in favor of aggregation-incompetent conformations (Darnell *et al.*, 2007). Moreover, a recent study has shown that when a polyQ tract of pathological length is positioned between two SpA domains, it triggers slower aggregation than when it is located at the N-terminus of one SpA domain (Robertson *et al.*, 2011b). The authors proposed that the reduction of the aggregation rate is due to lower conformational freedom and higher steric constraints.

Results obtained for the other model proteins with an inserted polyQ tract can tentatively be analyzed in terms of the balance between the intrinsic propensity of polyQ tracts to trigger aggregation into amyloid-like fibrils and the constraints imposed by the host protein on the polyQ tract. It is interesting to note that the chimera of myoglobin containing 50Q aggregates into amyloid-like fibrils when incubated under conditions similar to those where BlaP(Gln)₅₅ does not (Tanaka *et al.*, 2001). This observation suggests that steric/conformational constraints imposed on the polyQ tract in myoglobin are lower than those in BlaP. These lower constraints could, at least in part, originate from the fact that (i) the structure of myoglobin is perturbed by the insertion of the polyQ tract while the structure of BlaP is not and (ii) the loop in which the polyQ tract is inserted is significantly longer in myoglobin than in BlaP, essentially due to the addition of several amino acids from ataxin-3 at both sides of the polyQ tract (Tanaka *et al.*, 2001). In the case of apomyoglobin, the insertion of 38Q, the

longest tract used in the aggregation studies, does not induce amyloid-like fibril formation. This observation suggests that a 38Q tract inserted into apomyoglobin is not long enough to access amyloid-competent conformations and/or to overcome the steric hindrances exerted by apomyoglobin (Tobelman and Murphy, 2011).

Under native conditions, traces of dimeric species are observed for both BlaP and all chimeras with the highest amount being observed for BlaP₁₉₇₋₁₉₈ (Table 8) which does not form amyloid-like fibrils. It is therefore very unlikely that the observed dimers act as seeds to facilitate fibril formation by BlaP₁₉₇₋₁₉₈(Gln)₇₉ but not, for example, BlaP₁₉₇₋₁₉₈(Gln)₅₅. The high molecular weight oligomeric species are however observed only for BlaP₁₉₇₋₁₉₈(Gln)₇₉; since this protein is the only one to aggregate into amyloid-like fibrils under native conditions, the observed oligomers could indeed act as seeds of fibril formation and thus accelerate the process. A deeper characterization of the structural properties of the oligomeric species and of their role in the process of aggregation is under investigation. The potential of these species to act as seeds does not, however, invalidate the conclusions of the work. Indeed, this would purely imply that the constraints applied by the BlaP moiety act to prevent the formation of amyloid-like fibrils, at least in part, by preventing the formation of oligomeric species that are formed early on the pathway of aggregation.

7. Amyloid-like fibrils formed by BlaP₁₉₇₋₁₉₈(Gln)₇₉ are enzymatically inactive

BlaP₁₉₇₋₁₉₈(Gln)₇₉ samples were incubated 8 days under the three conditions described above. Then, the amyloid-like fibrils formed were separated from the remaining soluble protein by centrifugation. Under each condition, more than 95% of BlaP₁₉₇₋₁₉₈(Gln)₇₉ molecules are aggregated into fibrils and the rest of protein is still soluble as measured by absorbance measurement at 280 nm. Fibrils were resuspended in 50 mM sodium phosphate buffer, pH 7 (in a volume equivalent to that before centrifugation).

For each condition of incubation, initial rates of cephalothin (100 μM) hydrolysis were measured with:

- an aliquot of sample taken before centrifugation (Table 11: v₀ Mix) and its specific activity was calculated (Table 11: Specific activity Mix).
- the suspension of purified fibrils (Table 11: Fibrils) and
- the solution of remaining soluble proteins (Table 11: Supernatant).

Table 11: Initial rates of cephalothin hydrolysis by aggregated BlaP₁₉₇₋₁₉₈(Gln)₇₉ samples, their supernatant and their isolated fibrils

Incubation conditions	Specific activity Mix ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)	v_0 Mix ($\Delta\text{A}\cdot\text{min}^{-1}$)	v_0 Supernatant ($\Delta\text{A}\cdot\text{min}^{-1}$)	v_0 Fibrils ($\Delta\text{A}\cdot\text{min}^{-1}$)
Native 0 M urea, 37°C	0.038	- 0.0129	- 0.0132	- 0.0005
Denaturing (50% U) 1.85 M urea, 25°C	0.0035	- 0.0011	- 0.0008	- 0.0001
Denaturing (100% U) 3.5 M urea, 25°C	0.0044	- 0.0014	- 0.0008	- 0.00004

Mix corresponds to protein samples (110 μM) incubated 8 days under one of the three conditions and containing both aggregated and soluble proteins. U, unfolded state. 5 μl of samples, supernatant or resuspended isolated fibrils (not diluted) were added to 490 μl of a 100 μM cephalothin solution and absorbance at 260 nm was followed. Two initial rates were measured with the Mix and averaged and only one initial rate was measured with the isolated aggregates and the solution of remaining soluble protein.

First, the specific activity of BlaP₁₉₇₋₁₉₈(Gln)₇₉ in the sample incubated 8 days under native conditions ($0.038 \mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$) is very inferior to the expected specific activity of BlaP₁₉₇₋₁₉₈(Gln)₇₉ for a 100 μM cephalothin solution ($108 \mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$). This is a good indication that the fibrils formed by BlaP₁₉₇₋₁₉₈(Gln)₇₉ under native conditions are not active. As expected, the specific activities measured with samples incubated under denaturing conditions are significantly lower (> 8-fold) than that obtained with aggregated samples under native conditions. Moreover, the low residual activity detected in the aggregated sample seems to mainly come from the solution of remaining soluble protein (Table 11, v_0 Mix vs v_0 supernatant) and not from the aggregated form of BlaP₁₉₇₋₁₉₈(Gln)₇₉ (Table 11, v_0 Mix vs v_0 fibrils).

Taken together, these results show that fibrils formed by BlaP₁₉₇₋₁₉₈(Gln)₇₉, even under native conditions, are not active. These observations indicate that although the unfolding of BlaP is not the trigger of the aggregation into amyloid-like fibrils (see above), some structural reorganizations occur upon fibril formation and that within the fibrils, BlaP does not retain its native structure.

8. Conclusions

We have created and characterized a series of chimeras with 23, 30, 55 and 79 glutamines inserted into a solvent-exposed loop of a globular protein, the β -lactamase BlaP. The threshold number of glutamines above which the BlaP chimeras aggregate into amyloid-like fibrils critically depends on the structural integrity of the β -lactamase moiety. This result suggests that this threshold value results, at least in part, from a delicate balance between the intrinsic propensity of polyQ tracts to aggregate and the extrinsic protective conformational/steric constraints originating from BlaP. While it has been suggested that, in proteins associated with diseases, polyQ tracts are located in regions that are essentially unstructured and generally N- or C-terminal to a structured domain (Robertson and Bottomley, 2010), several studies have shown that sequences flanking the polyQ tract could, however, adopt some elements of secondary structure (Darnell *et al.*, 2007, Kim *et al.*, 2009). The latter could therefore exert constraints on the polyQ tract similar to those exerted by the BlaP moiety. The threshold for amyloid-like fibril formation by the BlaP chimeras under native conditions ($> 55Q$) is higher than the highest threshold length observed for proteins associated with diseases [49 residues in atrophin-1, (Robertson and Bottomley, 2010)]. This observation suggests that the constraints applied to the polyQ tract by BlaP in its folded state are higher than those imposed by the proteins associated with polyQ diseases, probably because BlaP is a more structured and rigid scaffold. The use of this globular protein has however allowed us to produce and characterize, for the first time, a model protein with an inserted 79Q tract. Interestingly, our results clearly show that the structural integrity of BlaP, and thus the constraints imposed on such a long tract, has negligible impact on the propensity of the latter to mediate amyloid-like fibril formation. Based on these observations, we speculate that the modulating effects of the protein context on the aggregating properties of proteins associated with polyQ diseases could also be negligible when a particularly long polyQ tract is present. PolyQ tracts of 88-306 residues were reported for some proteins associated with polyQ diseases (Orr and Zoghbi, 2007); the aggregating properties of proteins with such long tracts could be, therefore, dictated essentially by the intrinsic propensity of the polyQ expansion to form intermolecular β -sheets.

Finally, this study demonstrates that BlaP is an appropriate scaffold to further investigate the delicate balance between the propensity of polyQ tracts to trigger aggregation and the modulating effects of the host proteins.

Chapter 5. Creation and characterization of BlaP chimeras with polyQ tracts in position 216-217

1. Creation of chimeras with polyQ inserts in position 216-217

In order to compare the effects of polyQ insertions in position 216-217 on the properties of BlaP with those observed when the polyQ tracts are inserted in position 197-198, it is necessary to insert the same polyQ lengths in position 216-217 than those inserted in position 197-198 (i.e. 23, 30, 55 and 79Q). To achieve this with the strategy used to create the chimeras 197-198, was technically challenging for several reasons. Chimeras with polyQ tracts in position 197-198 have been constructed by cloning a library of polyCAG double-strand DNA fragments (obtained by an overlapping PCR strategy) into a SmaI restriction site previously introduced into the gene of BlaP (Chapter 4, section 1). In order to clone this library between the codons for residues 216 and 217, one should have first introduced a SmaI restriction site in this position. More problematic is the fact that the creation of the polyCAG double-strand DNA library and its successive cloning into the gene of BlaP does not allow controlling the number of CAG repeats that are inserted. Thus, in order to select clones with exactly 23, 30, 55 and 79 glutamines, we could have to screen a lot of clones without any guaranty of success. We therefore decided to order synthetic genes coding for:

(i) BlaP, with a SmaI restriction site between codons for residues 216 and 217. At the protein level, the introduction of this restriction site leads to the addition of a proline-glycine (PG) dipeptide between the amino acids 216 and 217 of BlaP. Although this restriction site was not anymore necessary at the gene level, we chose to keep it in order to reduce the number of factors varying from chimeras 197-198 to chimeras 216-217. The properties of this protein, referred to as BlaP₂₁₆₋₂₁₇, will be compared to those obtained with BlaP₁₉₇₋₁₉₈ (i.e. BlaP containing the insertion site, or in other words, the dipeptide PG in position 197-198). This comparison is important in order to dissociate the effects specific to the insertion of glutamines to those mainly due to the location of the insertion site (i.e. the dipeptide PG).

(ii) chimeric β -lactamases with 30, 55 and 79Q inserted in position 216-217, referred to as BlaP₂₁₆₋₂₁₇(Gln)₃₀, BlaP₂₁₆₋₂₁₇(Gln)₅₅ and BlaP₂₁₆₋₂₁₇(Gln)₇₉. These genes contain the codons for a proline and a glycine, upstream and downstream to the sequence coding for the polyQ tract, respectively. Since both BlaP₁₉₇₋₁₉₈(Gln)₂₃ and BlaP₁₉₇₋₁₉₈(Gln)₃₀ never aggregate into fibrils whatever the conditions of aggregation used, we decided to insert only one of these tracts (i.e.

30Q) in position 216-217. The nucleotidic sequence coding for a C-terminal glycine-proline followed by a (His)₅ tag was also introduced into the synthetic genes since these amino acids were present in BlaP₁₉₇₋₁₉₈ and its polyQ chimeras.

Finally, the genes of BlaP₂₁₆₋₂₁₇ and its polyQ chimeras were cloned into the constitutive expression vector pNY (Materials and Methods, section 1.2).

2. Production and purification of BlaP₂₁₆₋₂₁₇ and its polyQ chimeras

BlaP₂₁₆₋₂₁₇ and its corresponding polyQ chimeras with 30, 55 and 79Q were produced in the periplasm of *E. coli* cells using the same protocol than that described for the chimeras with inserts at position 197-198. Briefly, a 2 L culture was incubated one night at 37°C with agitation and then the periplasmic proteins were extracted by an osmotic shock (Materials and Methods, section 2). BlaP₂₁₆₋₂₁₇ and its chimeras were first purified by affinity chromatography on a Ni-PDC column via their C-terminal (His)₅ tag. Following elution with imidazole, the enzymatically active fractions containing more than 95% of the protein of interest as assessed by SDS-PAGE analysis were pooled, dialyzed twice against PBS pH 7.5 and stored at -20°C. However, for some production batches, contaminant proteins were still present in significant amount; in such cases, a second step of purification was performed using an anion exchange chromatography Q-HP column. After this second step of purification, fractions containing more than 95% of the protein of interest, as judged by SDS-PAGE analysis, were dialysed 2 times against PBS, pH 7.5, and stored at -20°C. Typical examples of production/purification yields obtained for the four proteins are given in the Table 12.

In general, the production/purification yields of BlaP₂₁₆₋₂₁₇ and its polyQ chimeras after the first step of purification are similar to those obtained for BlaP₁₉₇₋₁₉₈ and its polyQ chimeras (i.e. ~10-20 mg per liter of culture). However, the addition of the second step of purification reduces this yield by a factor up to 2.

The analysis by SDS-PAGE of purified BlaP₂₁₆₋₂₁₇ and its polyQ-containing chimeras is shown in the Figure 33. The migration of BlaP₂₁₆₋₂₁₇(Gln)₅₅ and BlaP₂₁₆₋₂₁₇(Gln)₇₉ is slower than that expected given their predicted molecular mass; on the other hand, BlaP₂₁₆₋₂₁₇ and BlaP₂₁₆₋₂₁₇(Gln)₃₀ migrated at the positions expected. The anomalous behaviour of BlaP₂₁₆₋₂₁₇(Gln)₅₅ and BlaP₂₁₆₋₂₁₇(Gln)₇₉ was previously observed for BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉ and for chimeras composed of myoglobin and polyQ sequences (Chapter 4, section 2). We have

previously proposed that long polyQ sequences interfere with SDS-binding. However, BlaP₂₁₆₋₂₁₇(Gln)₇₉ (Figure 33, Lane 5) has lower mobility than BlaP₁₉₇₋₁₉₈(Gln)₇₉ (Figure 33, Lane 7) while both proteins possess the same polyQ length. The reasons of this phenomenon are unclear.

Table 12: Purification of BlaP₂₁₆₋₂₁₇ and its polyQ chimeras (extracted from a 2 L culture)

Protein	Step	Total protein (mg)	Specific activity ¹ ($\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mgE}^{-1}$)	Total activity ($\mu\text{mol}\cdot\text{min}^{-1}$)	Purification factor ²	Yield ³ (%)
BlaP ₂₁₆₋₂₁₇	Periplasmic extract	213	4.3	920	-	-
	After one step of purification (Ni-PDC)	22	29	641	6.7	70
BlaP ₂₁₆₋₂₁₇ (Gln) ₃₀	Periplasmic extract	147	2.3	286	-	-
	After one step of purification (Ni-PDC)	31	7.5	233	3.3	81
BlaP ₂₁₆₋₂₁₇ (Gln) ₅₅	Periplasmic extract	138	1.3	174	-	-
	After the 1 st step of purification (Ni-PDC)	23	4.9	113	3.9	65
	After the 2 nd step of purification (Q-HP)	16	4.5	72	0.9	41
BlaP ₂₁₆₋₂₁₇ (Gln) ₇₉	Periplasmic extract	141	0.7	98	-	-
	After the 1 st step of purification (Ni-PDC)	15.5	5.5	86	7.9	87
	After the 2 nd step of purification (Q-HP)	9	3.8	34	0.7	40

¹The specific β -lactamase activity was measured using a 100 μM or a 500 μM cephalothin solution for BlaP₂₁₆₋₂₁₇ its polyQ chimeras, respectively (Materials and Methods, section 6).

²The purification factor is equal to: (specific activity measured after purification)/(specific activity measured before purification).

³The yield of purification is equal to: (total activity measured before purification)/(total activity measured after purification).

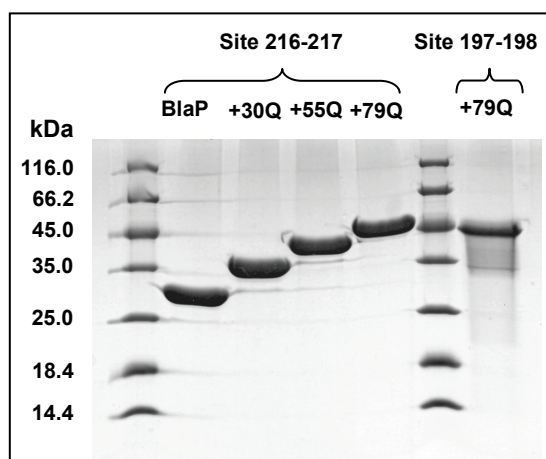


Figure 33: Purified BlaP₂₁₆₋₂₁₇ and its polyQ-chimeras separated on a 15% (w/v) acrylamide gel and stained with coomassie blue. The lanes 1 and 6 show protein markers with molecular masses as indicated. The lanes 2-5 show BlaP₂₁₆₋₂₁₇ and its polyQ chimeras, and the lane 7 gives a comparison with the chimera 197-198 with 79Q. Expected molecular masses are 30368, 34212, 37416 and 40491 Da for BlaP₂₁₆₋₂₁₇ and its chimeras with 30, 55 and 79 glutamines, respectively.

The integrity of the proteins was checked by mass spectrometry (ESI-QTOF-MS). For each protein, the majority of the protein population (> 65%) has the mass expected from the amino acid sequence. As for BlaP₁₉₇₋₁₉₈ and its chimeras, adducts of +182 Da and +200 Da are also observed in preparations of BlaP₂₁₆₋₂₁₇ and its chimeras. The resulting higher molecular mass species are present in similar amounts for all the proteins irrespective of the size or the location of the polyQ stretch (i.e. 10-20% for the +182 Da adduct and ≤ 15% for the +200 Da adduct). N-terminal sequencing was not carried out on sample preparations of BlaP₂₁₆₋₂₁₇ and its chimeras; the adduct of +200 Da is however likely to come from the same N-terminal misprocessing than that described for BlaP₁₉₇₋₁₉₈, its polyQ chimeras and the BlaP chimera with ChBD insert (Chapter 4, section 2) (Vandevenne *et al.*, 2007). Since the proportion of each molecular species is similar between various preparations of proteins related to the insertion sites 216-217 and 197-198, it is assumed that they do not interfere with the main conclusions of this work.

Preparations of BlaP₂₁₆₋₂₁₇ and its polyQ chimeras at different concentrations were analyzed by SEC. BlaP₂₁₆₋₂₁₇(Gln)₅₅ et BlaP₂₁₆₋₂₁₇(Gln)₇₉ were difficult to concentrate up to 120 μM due to a high tendency to aggregate. The percentages of the different species present in solution of BlaP₂₁₆₋₂₁₇ and its chimeras are compared to those observed in the preparations of BlaP₁₉₇₋₁₉₈ and its polyQ chimeras in Table 13. As a general observation, polyQ chimeras 216-217 have a higher tendency to oligomerize than polyQ chimeras 197-198; they form higher amount of dimers and a small proportion of trimers is also detected. Surprisingly, BlaP₂₁₆₋₂₁₇ does not form any oligomeric species while dimeric species were observed in BlaP₁₉₇₋₁₉₈ samples. Interestingly, no high molecular weight species were observed in BlaP₂₁₆₋₂₁₇(Gln)₇₉ samples while they were detected in samples of analyzed BlaP₁₉₇₋₁₉₈(Gln)₇₉ (Table 13). More analysis of samples at a protein concentration higher than 50 μM should however be performed in order to better characterize the propensity to oligomerize of BlaP₂₁₆₋₂₁₇ and its chimeras.

For BlaP₂₁₆₋₂₁₇ and its chimeras with 30 and 55Q, more than ca. 96% of the protein population is monomeric at concentrations as high as 50 μM, we assume that the presence of dimers, trimers did not significantly interfere with the determination of the catalytic and thermodynamic parameters, which were carried out with protein concentrations ≤ 4.6 μM. For BlaP₂₁₆₋₂₁₇(Gln)₇₉ which presents the higher proportion of dimers/trimers (7.7%) at 50 μM, decreasing the protein concentration from 50 to 10 μM increases the percentage of monomeric species from ca. 92 to ca. 94%. Thus, we can also assume that at protein concentration ≤ 4.6

μM , the percentages of dimers/trimers is small and that their presence have no or little influence on the determination of the catalytic and thermodynamic parameters for this protein

Table 13: Percentages of the different species observed by SEC

	[protein] (μM)	M (%)	D (%)	T (%)	O (%)
BlaP₂₁₆₋₂₁₇	50	100	0	0	0
(BlaP₁₉₇₋₁₉₈)	50	(95.1)	(4.9)	(0)	(0)
BlaP₂₁₆₋₂₁₇(Gln)₃₀	50	96.5	3	0.5	0
(BlaP₁₉₇₋₁₉₈(Gln)₃₀)	50	(99.6)	(0.4)	(0)	(0)
BlaP₂₁₆₋₂₁₇(Gln)₅₅	10	98.8	1.2	0	0
BlaP₂₁₆₋₂₁₇(Gln)₅₅	50	97.0	2.8	0.2	0
(BlaP₁₉₇₋₁₉₈(Gln)₅₅)	50	(98.9)	(1.1)	(0)	(0)
BlaP₂₁₆₋₂₁₇(Gln)₇₉	10	93.9	6.1	0	0
(BlaP₁₉₇₋₁₉₈(Gln)₇₉)	10	(97.3)	(0.9)	(0)	(1.8)
BlaP₂₁₆₋₂₁₇(Gln)₇₉	50	92.3	7.5	0.2	0
(BlaP₁₉₇₋₁₉₈(Gln)₇₉)	50	(96.4)	(1.1)	(0)	(2.5)

This analysis was carried out using a Superdex 200 GL 10/300 column using PBS buffer, pH 7.5. **M**, **D**, **T**, **O** stand for monomer, dimer, trimer and high molecular weight oligomeric species, respectively. The percentages given in the table correspond to: (peak area/total area in evaluated peaks)*100.

3. Effects of polyQ insertions in position 216-217 on the catalytic properties of BlaP

The effects of polyQ insertions in position 216-217 on the β -lactamase activity were investigated by monitoring the hydrolysis of cephalothin via the decrease in the absorbance at 260 nm. Complete time-courses of cephalothin hydrolysis ($[\text{cephalothin}] = 0.5 - 2 \text{ mM}$) were recorded for BlaP₂₁₆₋₂₁₇ and the values of K_m and k_{cat} were determined as previously described (De Meester *et al.*, 1987). For the polyQ chimeras, initial rates of hydrolysis were measured for cephalothin concentrations ranging from 0.1 to 2.5 mM. For all three chimeras, a linear dependence of the initial rates of hydrolysis on the cephalothin concentration was observed in the entire range of concentrations tested (Figure 34). This observation indicates that the K_m values for these proteins are higher than 2.5 mM ($[\text{S}] \ll K_m$) and that only the k_{cat}/K_m ratios can be determined ($v_0 = ([\text{S}] \cdot V_{max})/K_m$). The catalytic parameter values for cephalothin hydrolysis by BlaP₂₁₆₋₂₁₇ and its polyQ chimeras are presented in the Table 14 and compared to those of BlaP₁₉₇₋₁₉₈.

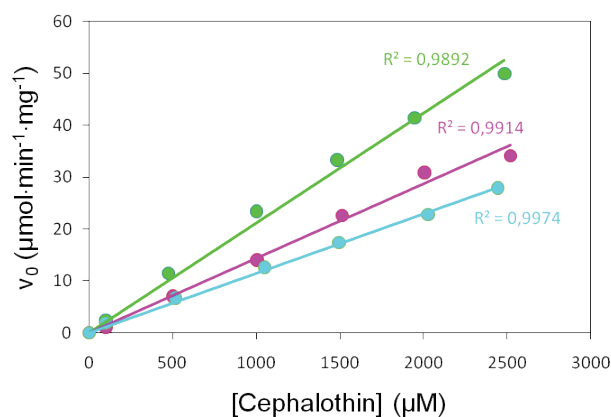


Figure 34: Initial rates (v_0) of cephalothin hydrolysis by BlaP₂₁₆₋₂₁₇(Gln)₃₀ (green), BlaP₂₁₆₋₂₁₇(Gln)₅₅ (pink) and BlaP₂₁₆₋₂₁₇(Gln)₇₉ (cyan) as a function of its concentration.

Table 14: Kinetic parameter values for BlaP₂₁₆₋₂₁₇ and its corresponding polyQ chimeras

	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($mM^{-1}\cdot s^{-1}$)
BlaP₂₁₆₋₂₁₇	76 ± 7	520 ± 50	150 ± 15
BlaP₂₁₆₋₂₁₇(Gln)₃₀	> 30	> 2500	12 ± 0.5
BlaP₂₁₆₋₂₁₇(Gln)₅₅	> 22	> 2500	9 ± 0.4
BlaP₂₁₆₋₂₁₇(Gln)₇₉	> 19	> 2500	7.7 ± 0.3
¹ BlaP₁₉₇₋₁₉₈	90 ± 10	25 ± 2.5	3600 ± 400

Cephalothin was used as the substrate. Errors are given as standard deviations. ¹Note that the catalytic parameters determined for polyQ chimeras 197-198 are similar to those of BlaP₁₉₇₋₁₉₈ (Chapter 4, section 3).

The k_{cat} value obtained for BlaP₂₁₆₋₂₁₇ is not significantly different from that of BlaP₁₉₇₋₁₉₈. The K_m value for the former protein is however more than twentyfold higher than that of the latter, considerably reducing the catalytic efficiency (k_{cat}/K_m). The simplest interpretation of these results is that the substrate accesses the catalytic site of BlaP₂₁₆₋₂₁₇ less easily than that of BlaP₁₉₇₋₁₉₈ but once bound, it is however cleaved with the same efficiency by both proteins. The difference between BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇ is the location of the PG dipeptide which is in position 197-198 in the former and in position 216-217 in the latter (Figure 35A). The loop containing the insertion site 197-198 (loop between H8 and H9) is on the opposite side of the enzyme molecule relatively to the active site. In contrast, the loop (between helices H9 and H10) containing the insertion site 216-217 of BlaP, although not directly involved in the catalytic site (Matagne *et al.*, 1998), is located in its vicinity (Figure 35A and B). The side chain of residues 216 and 217 (threonine 216 and glycine 217) are directed toward the entrance of the catalytic site (Figure 35B). It is therefore likely that the insertion of the PG dipeptide in position 216-217 renders the access to the active site more difficult.

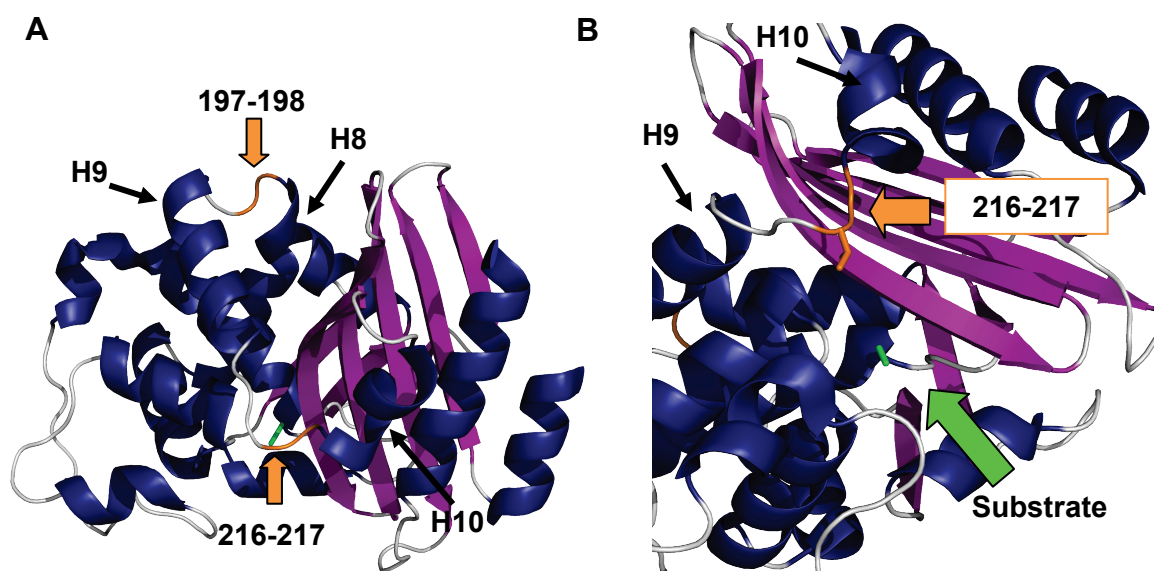


Figure 35: (A) Location of the insertion sites (orange arrows) 197-198 and 216-217 in BlaP. The active serine 70 is shown in green. H, helix. (B) Closer view of the catalytic site with the active serine and the entrance of catalytic site.

The insertion of glutamine repeats within the PG dipeptide in position 216-217 leads to a further increase in the K_m value (more than fivefold) (Table 14). Although the k_{cat} value cannot be determined, it is greater than $19\text{-}30\text{ s}^{-1}$ for the chimeras 216-217 and thus, not very different from the k_{cat} value of BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇. The higher K_m value for the chimeras is therefore probably responsible for the observed decrease of their k_{cat}/K_m ratio compared to that of BlaP₂₁₆₋₂₁₇. These results suggest that the insertion of glutamines renders the access to the catalytic site even more difficult. The k_{cat}/K_m ratio slightly decreases by a factor ~ 1.5 as the polyQ length increases from 30 to 79Q. This is however minor compared to the 25-fold and 12-fold decreases of the k_{cat}/K_m ratio due to the change in position of the dipeptide PG and the insertion of the shortest polyQ tract (30Q), respectively. Thus, the insertions of 55 and 79 glutamines do not appear to render the access to the active site significantly more difficult than the insertion of 30 glutamines.

4. Effects of polyQ insertions in position 216-217 on the overall structure of BlaP

First, intrinsic fluorescence spectrum of BlaP₂₁₆₋₂₁₇ was compared to that of BlaP₁₉₇₋₁₉₈ (Figure 36A). The spectrum of BlaP₂₁₆₋₂₁₇ (Figure 36A, blue) is red-shifted compared to that of BlaP₁₉₇₋₁₉₈ (Figure 36A, continuous grey line) by about 1.4 nm (Table 15). Although this value is small, it was observed with proteins originating from different batches of production. This result suggests that when located in position 216-217, the dipeptide PG perturbs the tertiary

structure of BlaP and modifies the environment of at least one tryptophan, exposing it to the solvent.

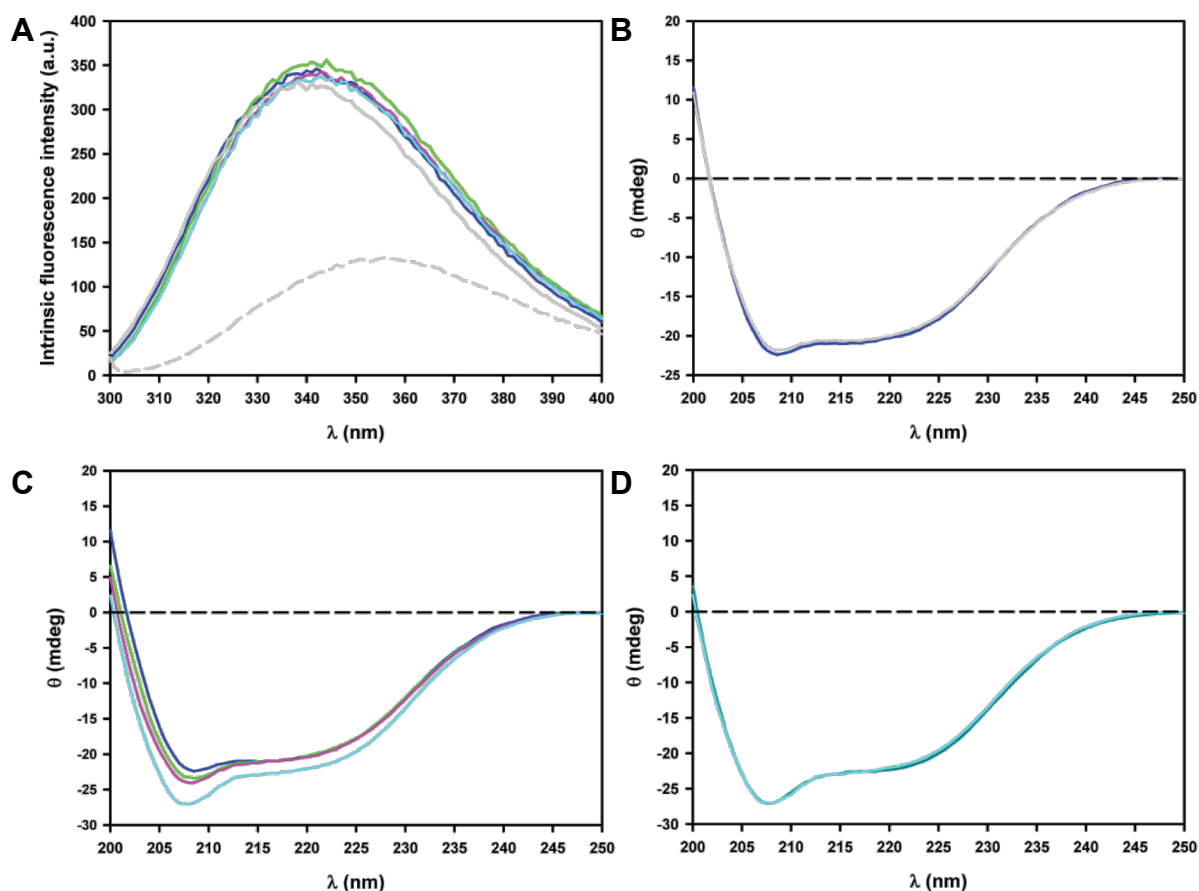


Figure 36: (A) Intrinsic fluorescence spectra of BlaP₂₁₆₋₂₁₇ (blue) and its chimeras with 30 (green), 55 (pink) and 79 (cyan) glutamines compared to that of BlaP₁₉₇₋₁₉₈ (continuous grey line). The intrinsic fluorescence spectrum of BlaP₁₉₇₋₁₉₈ unfolded 3 hours in 5.5 M urea is also shown (dashed grey line). (B) Superimposition of the far-UV CD spectra of BlaP₂₁₆₋₂₁₇ (blue) and BlaP₁₉₇₋₁₉₈ (grey). (C) Far-UV CD spectra of BlaP₂₁₆₋₂₁₇ (blue) and its chimeras with 30 (green), 55 (pink) and 79 (cyan) glutamines. (D) Superimposition of the far-UV CD spectra of BlaP₂₁₆₋₂₁₇(Gln)₇₉ (cyan) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (dark cyan). Spectra were recorded at 25°C in PBS, pH 7.5, using protein concentrations of 4.6 μM.

Table 15: Wavelengths of maximum fluorescence emission (λ_{\max})

	λ_{\max} (nm)
BlaP₁₉₇₋₁₉₈⁽¹⁾ (Unfolded)	340.0 355.4
BlaP₂₁₆₋₂₁₇	341.4
BlaP₂₁₆₋₂₁₇(Gln)₃₀	342.6
BlaP₂₁₆₋₂₁₇(Gln)₅₅	342.7
BlaP₂₁₆₋₂₁₇(Gln)₇₉	342.6

⁽¹⁾ Note that the wavelengths of maximum fluorescence emission of polyQ chimeras 197-198 are the same than that of BlaP₁₉₇₋₁₉₈ (i.e. 340.0 nm). A five-parameter weibull function (provided with the software SigmaPlot 9.0) was fitted to the fluorescence intensity spectra to obtain the associated λ_{\max} values.

The insertion of glutamines between the residues of the PG dipeptide further shifts the spectrum of the chimeras 216-217 to higher wavelengths than that of BlaP₂₁₆₋₂₁₇ (Table 15). This small shift (1.2-1.3 nm), which is also observed with different batches of protein, is independent of the number of glutamines inserted. This result therefore suggests that the insertion of a polyQ tract in position 216-217 further perturbs the tertiary structure of BlaP and this, independently of its length.

Although the fluorescence emission maximum is shifted to higher wavelength, the intensity of the signal does not decrease as it is observed upon unfolding of the BlaP moiety (Figure 36A, dashed grey line). This result suggests that even if the environment of at least one tryptophan is modified in BlaP₂₁₆₋₂₁₇ and its chimeras, the perturbations of the overall tertiary structure are minor. Such a perturbation of the tertiary structure was not observed when the glutamines were inserted in position 197-198.

On the other hand, the contents of secondary structure of BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇ are very similar as shown by the superimposition of their far-UV CD spectra (Figure 36B). Moreover, the far-UV CD spectra of the polyQ chimeras 216-217 have a similar shape than those of the polyQ chimeras 197-198 [Figure 36C and Figure 21C (page 58)]. These spectra are characterized by a negative peak at around 207-208 nm increasing with polyQ length. Figure 36D shows the superimposition of the far-UV CD spectra of BlaP₁₉₇₋₁₉₈(Gln)₇₉ and BlaP₂₁₆₋₂₁₇(Gln)₇₉. The good coincidence of these spectra indicates that these proteins have very similar secondary structure organisations and suggests that the polyQ tracts inserted in positions 197-198 and 216-217 are both equally disordered.

Taken together, these results indicate that the minor perturbations of the tertiary structure associated with the change in the location of the insertion site are however not associated with significant change in the secondary structure content of BlaP.

BlaP contains three tryptophans W210, W229 and W251 (Figure 37); W210 is located on the helix H9 and points to the interior of the structure, W229 is located on the β -strands B3 and is covered by the flexible loop connecting the helix H10 to the β -strand B3 and finally, W251 is located on the β -strands B4 and is covered by the helix H9. Thus, the introduction of the dipeptide PG in position 216-217 (loop between H9 and H10) and even more so, the insertion of glutamines between the residues of this dipeptide could induce a slight movement of the helices H9 and H10 (or only one of them) which would result in an increase of solvent exposition of at least one of the three tryptophans.

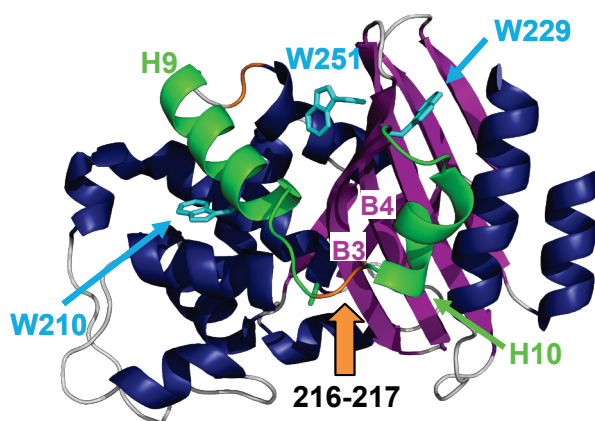


Figure 37: Three-dimensional structure of BlaP with the positions of the tryptophan residues (W) indicated. The active serine 70 is also shown (green). H, helix. B, β -strand.

Despite these structural perturbations, the results of the enzymatic activity measurements suggest that the loop connecting the helices H9 and H10 should remain close to the active site and still directed toward this latter.

5. Effects of polyQ insertions in position 216-217 on the stability of BlaP

The effects of polyQ insertions in position 216-217 on the stability of BlaP were investigated by urea-induced unfolding experiments. The loss of tertiary and secondary structures was followed by intrinsic fluorescence and far-UV CD measurements, respectively, as described previously (Chapter 4, section 5). First, we checked that the unfolding reaction was reversible and had reached the equilibrium before the measurements were made.

The signal of unfolded protein samples was therefore compared to that of unfolded/refolded samples, i.e. concentrated protein samples fully denatured for 3 hours in 5.5 M urea [4.5 M for BlaP₂₁₆₋₂₁₇(Gln)₇₉] and diluted in different final concentrations in urea (Figure 23, page 61). For all proteins and within experimental errors (essentially on the protein concentration), fluorescence (Figure 38) and far-UV CD (Annexe 2) signals of unfolded samples coincide well with those from unfolded/refolded samples, indicating that the urea unfolding reaction is completely reversible and that equilibrium is reached in all unfolded samples after one night.

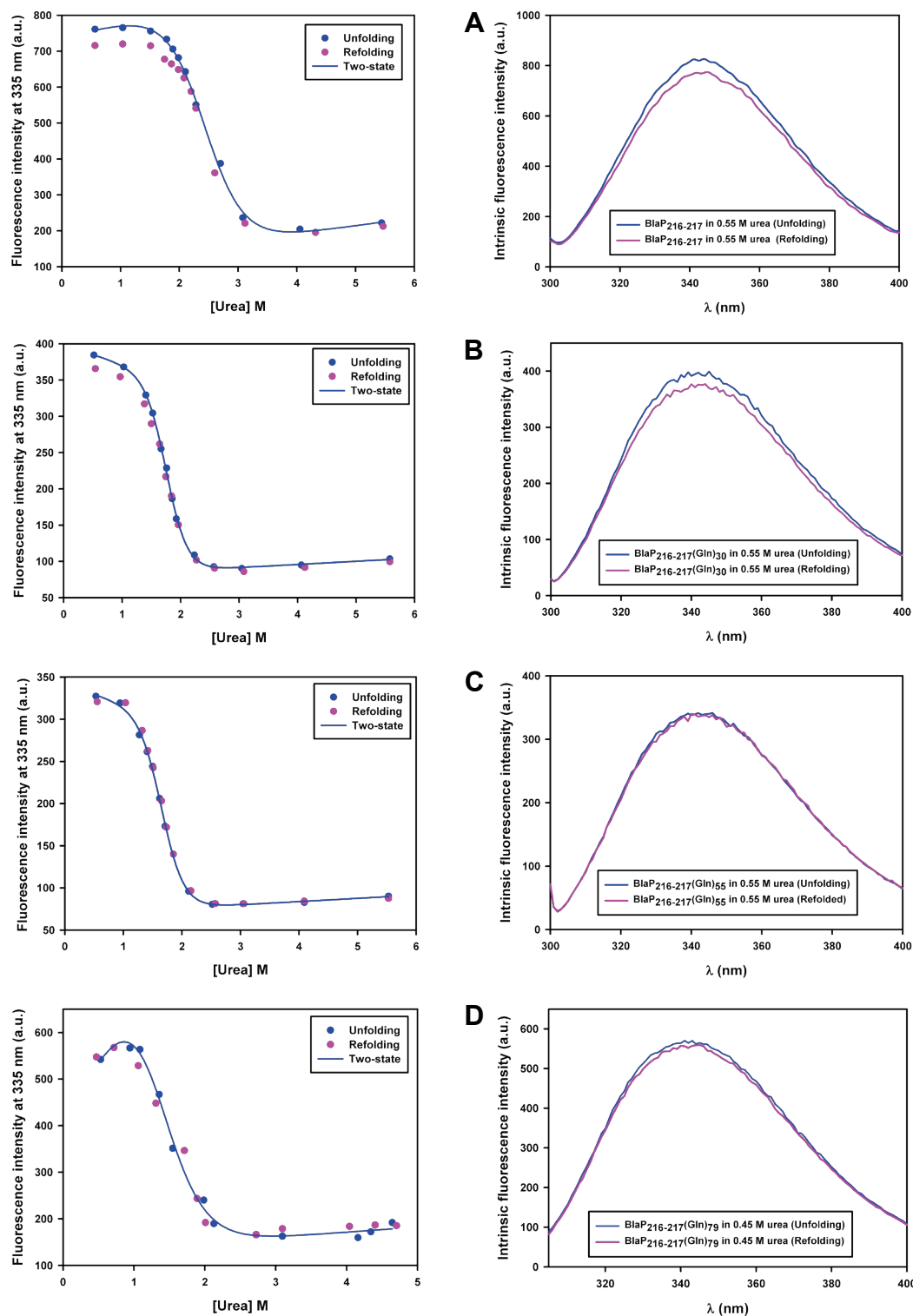


Figure 38: Unfolding (blue)/refolding (pink) transitions monitored by the change in fluorescence intensity at 335 nm (**left column**) and intrinsic fluorescence spectra (**right column**) of native (blue) and unfolded/refolded (pink) samples in 0.55 M urea (or in 0.45 M for BlaP₂₁₆₋₂₁₇(Gln)₇₉). BlaP₂₁₆₋₂₁₇ (**A**) and its chimeras with 30 (**B**), 55 (**C**) and 79Q (**D**). a.u., arbitrary unit. For BlaP₂₁₆₋₂₁₇(Gln)₅₅ and BlaP₂₁₆₋₂₁₇(Gln)₇₉, data from the unfolding experiments coincide well with those from unfolding/refolding experiments, indicating that the urea unfolding reaction is completely reversible and that equilibrium is reached in all unfolded samples after one night. In the cases of BlaP₂₁₆₋₂₁₇ and BlaP₂₁₆₋₂₁₇(Gln)₃₀, only 94% of the fluorescence and far-UV CD (Annexe 2) signals is however recovered upon refolding in 0.55 M urea. Since the unfolding reactions for BlaP₂₁₆₋₂₁₇(Gln)₅₅ and BlaP₂₁₆₋₂₁₇(Gln)₇₉ which have more inserted glutamines are reversible, we can assume that the ~6% of unrecovered signal after refolding of BlaP₂₁₆₋₂₁₇ and BlaP₂₁₆₋₂₁₇(Gln)₃₀ probably result from a difference in protein concentration between unfolded and refolded samples.

Then, unfolding curves with points at many more urea concentrations were performed to determine the values of the thermodynamic parameters. BlaP₂₁₆₋₂₁₇ and all the polyQ chimeras 216-217 unfold according to a simple two-state mechanism ($N \rightleftharpoons U$), with no intermediate species significantly populated between the native and unfolded states. Indeed, the curves obtained by both intrinsic fluorescence and far-UV CD are single transitions (the data obtained for BlaP₂₁₆₋₂₁₇ are shown in Figures 39A and B as a representative example) and the normalized data obtained by both spectroscopic techniques coincide well (Figure 40A).

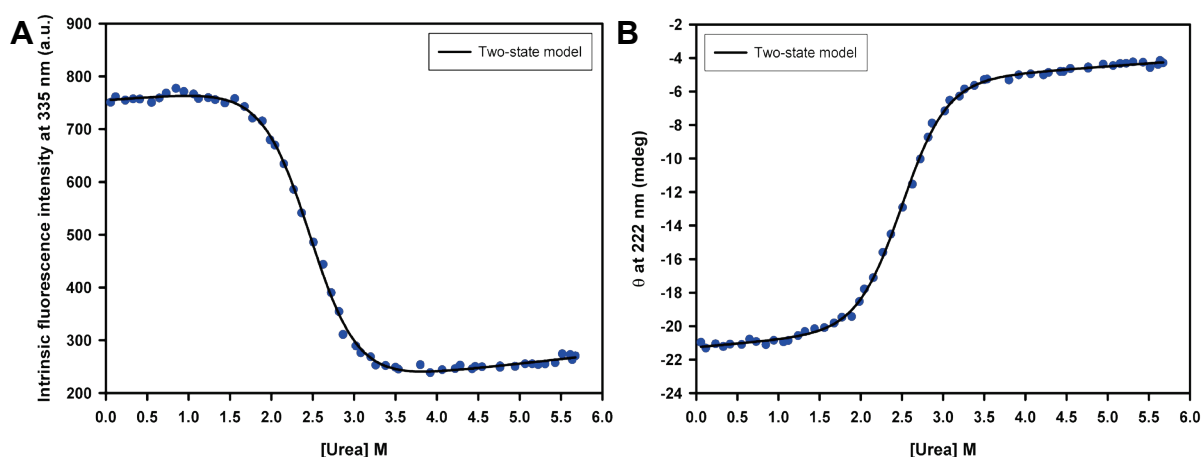


Figure 39: Urea-induced unfolding transitions obtained for BlaP₂₁₆₋₂₁₇ (4.6 μ M) in PBS, pH 7.5 and at 25°C and monitored by the change (A) in fluorescence intensity at 335 nm and (B) in ellipticity at 222 nm. The data were analyzed on the basis of a two-state model and the lines represent the best fit to the equation 2 (Materials and Methods, section 9)

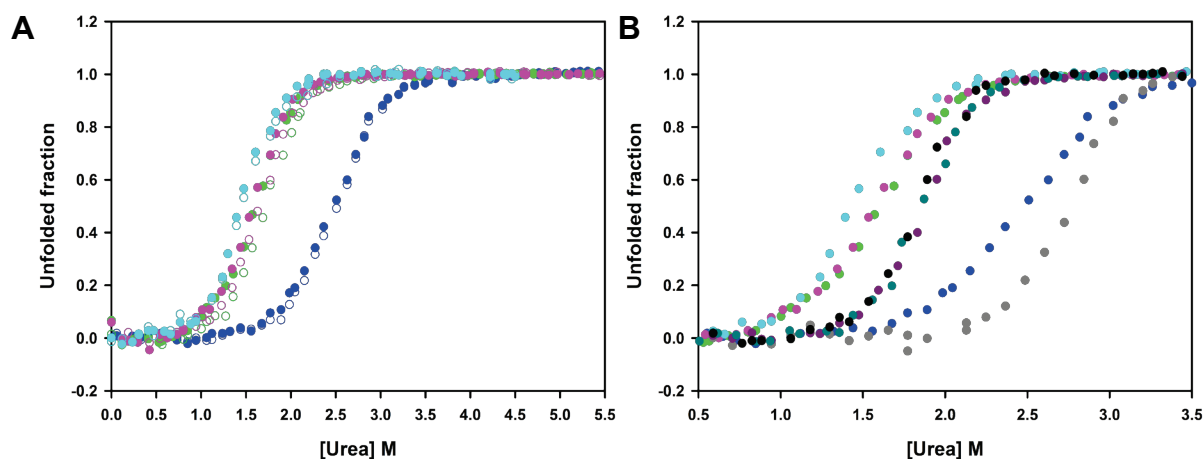


Figure 40: (A) Normalized urea-induced unfolding transitions obtained for BlaP₂₁₆₋₂₁₇ (blue), BlaP₂₁₆₋₂₁₇(Gln)₃₀ (green), BlaP₂₁₆₋₂₁₇(Gln)₅₅ (pink) and BlaP₂₁₆₋₂₁₇(Gln)₇₉ (cyan) in PBS, pH 7.5 and at 25°C and monitored by the change in fluorescence intensity at 335 nm (filled circles) and in ellipticity at 222 nm (open circles). The protein concentration is 4.6 μ M. (B) Comparison of the normalized urea unfolding transitions obtained for BlaP₂₁₆₋₂₁₇ and its polyQ chimeras [see legend in (A)] with those obtained for BlaP₁₉₇₋₁₉₈ (grey), BlaP₁₉₇₋₁₉₈(Gln)₃₀ (black), BlaP₁₉₇₋₁₉₈(Gln)₅₅ (violet) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (dark cyan). Only data derived from fluorescence measurements are shown.

The averaged values of the thermodynamic parameters resulting from fitting the equation of a two-state model to the experimental data are presented in Table 16.

Analysis of these parameters shows that BlaP₂₁₆₋₂₁₇ is significantly destabilized compared to BlaP₁₉₇₋₁₉₈ [Table 16, the $\Delta\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$ between BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇ is 15 kJ·mol⁻¹]. Analysis of the X-ray structure of the wild-type BlaP (i.e. the protein as produced by *Bacillus licheniformis*) suggests that the loop containing the position 216-217 (residues 213-217) is less flexible than the loop containing the position 197-198 (residues 197-200), while having similar lengths; the temperature factor of the main chain is indeed 5.9 Å² for the former and 9.6 Å² for the latter. Thus, it is not surprising that the PG dipeptide when located in position 216-217 destabilizes significantly more the enzyme than when it is located in position 197-198.

Table 16: Comparison between the thermodynamic parameters of unfolding of BlaP₂₁₆₋₂₁₇/chimeras 216-217 and BlaP₁₉₇₋₁₉₈/chimeras 197-198

	$\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$ (kJ·mol ⁻¹)	$\Delta\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$ 197-216 (kJ·mol ⁻¹)	m_{NU} (kJ·mol ⁻¹ ·M ⁻¹)	C_m (M)
BlaP₂₁₆₋₂₁₇	22.0 ± 1.1	15.0 ± 2.9	8.8 ± 0.4	2.48 ± 0.15
BlaP₁₉₇₋₁₉₈	37.0 ± 1.8		13.5 ± 0.6	2.74 ± 0.23
BlaP₂₁₆₋₂₁₇(Gln)₃₀	18.1 ± 0.9	11.3 ± 3.3	10.9 ± 0.4	1.65 ± 0.12
BlaP₁₉₇₋₁₉₈(Gln)₃₀	29.4 ± 2.4		15.9 ± 0.7	1.85 ± 0.15
BlaP₂₁₆₋₂₁₇(Gln)₅₅	17.9 ± 0.7	11.1 ± 1.3	11.1 ± 0.4	1.61 ± 0.12
BlaP₂₁₆₋₂₁₇(Gln)₅₅	29.0 ± 0.6		15.6 ± 0.6	1.86 ± 0.15
BlaP₂₁₆₋₂₁₇(Gln)₇₉	16.7 ± 1.2	11.5 ± 2.5	11.5 ± 0.8	1.46 ± 0.15
BlaP₂₁₆₋₂₁₇(Gln)₇₉	28.2 ± 1.3		14.9 ± 0.6	1.89 ± 0.17

The thermodynamic parameters were deduced from the analysis of urea unfolding transitions monitored by intrinsic fluorescence and far-UV CD measurements using equation 2. The urea unfolding experiments were carried out at 25°C. $\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$, m_{NU} and C_m values result from the mean of the values obtained by far-UV CD and intrinsic fluorescence. $\Delta\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$ 197-216 is the differences between, respectively, $\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$ obtained with proteins related to site 197-198 and those obtained for proteins related to site 216-217.

The addition of glutamines at this latter position further increases the extent of destabilization; considering the errors on the $\Delta G^{\circ}_{\text{NU}}(\text{H}_2\text{O})$, this latter is however largely independent of the length of the polyQ (i.e. 3.9-5.3 kJ·mol⁻¹). The destabilizing effect of polyQ insertions in position 216-217 is therefore less pronounced than that of polyQ insertions in position 197-198 (7.6-8.8 kJ·mol⁻¹).

The m is the measure dependence of the free energy on the denaturant concentration. Its value has been correlated, at least for globular proteins, with the difference between solvent accessible surface areas (ASA) in the unfolded and folded states ($\Delta\text{ASA} = \text{ASA}_U - \text{ASA}_N$) (Myers *et al.*, 1995), and it is expected to be directly proportional to the change in solvent

accessible surface area upon unfolding. Thus, a change in ΔASA due to a mutation could originate from a change in the accessible surface area of the native state (A_N) and/or of the denatured state (A_U) of the protein. The ΔASA has been correlated with the number of residues in proteins and can be estimated to calculate the theoretical m value. BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇ have the same theoretical m value since they have the same number of amino acids (Table 17). However, the experimental m value obtained for BlaP₁₉₇₋₁₉₈ is in relatively good agreement with the theoretical value while that obtained for BlaP₂₁₆₋₂₁₇ is significantly lower (Table 17). The difference between the experimental m values of these proteins ($-4.7 \text{ kJ}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$) corresponds to a ca. 40% decrease in the surface exposed upon unfolding of BlaP₂₁₆₋₂₁₇ compared to that exposed upon unfolding of BlaP₁₉₇₋₁₉₈. It is rather unlikely that the minor perturbations in the tertiary structure observed for BlaP₂₁₆₋₂₁₇ can totally account for such a large change in ΔASA . These results therefore suggest that the denatured state of the two proteins could significantly differ; that of BlaP₂₁₆₋₂₁₇ should be less expanded. On the other hand, a lower m value could also be due to a deviation from a two-state mechanism and thus, the presence of intermediates. The unfolding curves obtained by both fluorescence and CD measurements are simple transitions (Figure 39 and data not shown) whose normalizations coincide together (Figure 40A). Moreover, ANS fluorescence measurements (data not shown) have revealed no significant binding of this probe which specifically binds to molten globule intermediate. Thus, the lower m value could be attributed to the formation of a series of discrete intermediates between the native and unfolded state. A difference in the experimental m values obtained for polyQ chimeras 216-217 and 197-198 is also observed and is similar to that observed between BlaP₂₁₆₋₂₁₇ and BlaP₁₉₇₋₁₉₈. Thus, the same hypothesis can be made to explain it. Note that if discrete intermediate species are indeed populated, the thermodynamic parameter values derived from equation 2 are erroneous and should be considered as apparent parameters.

Table 17: Theoretical and experimental m values and theoretical ΔASA for BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇.

	BlaP ₁₉₇₋₁₉₈	BlaP ₂₁₆₋₂₁₇
m theoretical ($\text{kJ}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$)⁽¹⁾	12.8	12.8
m experimental ($\text{kJ}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$)	13.5 ± 0.6	8.8 ± 0.4
ΔASA theoretical (\AA^2) determined from the m experimental⁽²⁾	25918	15711

⁽¹⁾ m theoretical values are determined using the empirical equation of Myers *et al.* (Myers *et al.*, 1995): $m = 373 + 0.11(\Delta\text{ASA})$ with $\Delta\text{ASA} = -907 + 93(\#\text{res})$, $\#\text{res}$ is the number of residues in the protein.

⁽²⁾ The theoretical values of ΔASA presented in the table are calculated from the experimental m value to estimate the difference in the surface areas exposed upon unfolding for BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇.

Thermal-induced unfolding of BlaP₂₁₆₋₂₁₇ and its corresponding polyQ chimeras was also investigated. For all proteins, single thermal transitions were obtained by both fluorescence and far-UV CD measurements (as an example, the data obtained for BlaP₂₁₆₋₂₁₇ are shown in Figure 41A and B) and the normalized data obtained by both techniques are superimposable (Figure 42), indicating that the proteins unfold according to a simple two-state model (N=U).

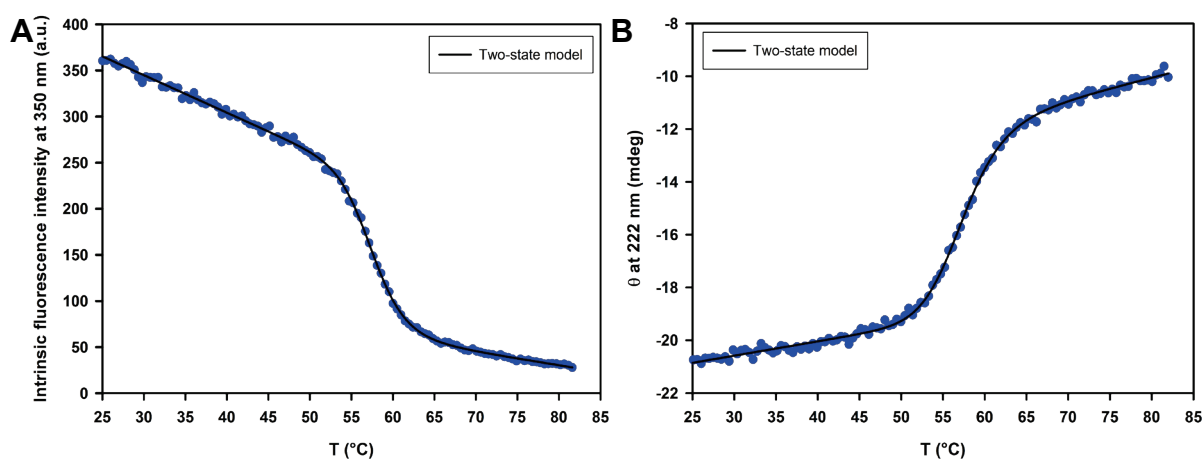


Figure 41: Heat-induced unfolding transitions obtained for BlaP₂₁₆₋₂₁₇ (4.6 μ M) in PBS, pH 7.5 and monitored by the change (A) in fluorescence intensity at 350 nm and (B) in ellipticity at 222 nm. The data were analyzed on the basis of a two-state model and the lines represent the best fit to the equation 3 (Materials and Methods, section 10)

Since thermal unfolding is not reversible for BlaP₂₁₆₋₂₁₇ and its chimeras (data not shown), only apparent T_m values were deduced from fitting the equation of a two-state model to the experimental data (Table 18); the values obtained by fluorescence and far-UV measurements were averaged.

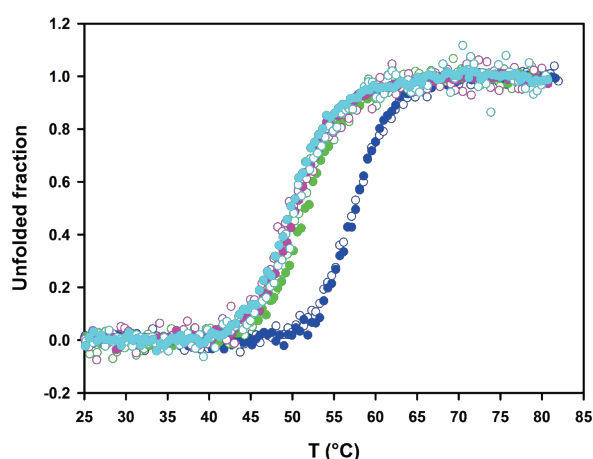


Figure 42: Normalized heat-induced unfolding transitions obtained for BlaP₂₁₆₋₂₁₇ (blue), BlaP₂₁₆₋₂₁₇(Gln)₃₀ (green), BlaP₂₁₆₋₂₁₇(Gln)₅₅ (pink) and BlaP₂₁₆₋₂₁₇(Gln)₇₉ (cyan) at pH 7.5 and monitored by the change in fluorescence intensity at 350 nm (filled circles) and in ellipticity at 222 nm (open circles). The protein concentration is 4.6 μ M.

Table 18: Comparison between mid-unfolding temperature of BlaP₂₁₆₋₂₁₇/chimeras 216-217 and BlaP₁₉₇₋₁₉₈/chimeras 216-217

	T_m^{app} (°C)
BlaP₂₁₆₋₂₁₇	57.2 ± 0.7
BlaP₁₉₇₋₁₉₈	58.5 ± 0.4
BlaP₂₁₆₋₂₁₇(Gln)₃₀	51.5 ± 0.6
BlaP₁₉₇₋₁₉₈(Gln)₃₀	53.7 ± 0.3
BlaP₂₁₆₋₂₁₇(Gln)₅₅	50.5 ± 0.8
BlaP₂₁₆₋₂₁₇(Gln)₅₅	53.5 ± 0.6
BlaP₂₁₆₋₂₁₇(Gln)₇₉	50.4 ± 0.9
BlaP₂₁₆₋₂₁₇(Gln)₇₉	53.2 ± 0.3

Thermal unfolding was not fully reversible and only the apparent mid-unfolding temperature (T_m^{app}) was determined. T_m^{app} values result from the mean of the values obtained by both fluorescence and far-UV CD measurements.

Two information can be extracted from these results:

- (i) the stability against temperature of polyQ chimeras 216-217 is lower than that of BlaP₂₁₆₋₂₁₇ and the extent of destabilization is largely independent of the length of the insert ($\Delta Tm^{app} = 5.7 - 6.8^\circ C$). This is in good agreement with the results obtained for the urea-induced unfolding.
- (ii) more importantly, all proteins related to site 216-217 are still native at 37°C. This implies that the aggregation of the chimeras under native conditions could be carried out under the same conditions as those used for BlaP₁₉₇₋₁₉₈ and its chimera (i.e. at 37°C).

6. Effects of polyQ insertions in position 216-217 on the aggregating properties of BlaP

Due to a lack of time to carry out the aggregation experiments with all the chimeras, we decided to focus on the following major question: would the different effects on the structure and stability of BlaP depending on the location of the insertion site correlate with a change in the polyQ threshold length for amyloid-like fibril formation? For polyQ chimeras 197-198, this threshold is comprised between:

- 55 and 79Q under native conditions (the chimera with 79Q forms amyloid-like fibrils whereas the one with 55Q only forms amorphous aggregates).
- 30 and 55Q under denaturing conditions (the chimera with 55Q forms amyloid-like fibrils whereas the one with 30Q does not significantly aggregate).

To answer this question, we should at least characterize the aggregation properties of BlaP₂₁₆₋₂₁₇(Gln)₃₀, BlaP₂₁₆₋₂₁₇(Gln)₅₅ and BlaP₂₁₆₋₂₁₇(Gln)₇₉ and to compare them with those of their homolog from site 197-198 [BlaP₁₉₇₋₁₉₈(Gln)₃₀, BlaP₁₉₇₋₁₉₈(Gln)₅₅ and BlaP₁₉₇₋₁₉₈(Gln)₇₉]. However, problems encountered during the concentration of the chimera BlaP₂₁₆₋₂₁₇(Gln)₇₉ and the lack of time to produce BlaP₁₉₇₋₁₉₈(Gln)₃₀ have deprived us of these proteins for the experience.

For each protein, one time course of aggregation was monitored in two different sets of incubation conditions:

- (i) under native conditions: PBS buffer, pH 7.5 and at 37°C
- (ii) under denaturing conditions: PBS buffer, pH 7.5, at 25°C and in the presence of a concentration in urea for which 50% of the molecules of a given chimera are unfolded (C_m M urea, Table 19).

Table 19: Urea concentration used for the aggregation under denaturing conditions

	[Urea] (M)
BlaP₂₁₆₋₂₁₇(Gln)₃₀	1.65
BlaP₂₁₆₋₂₁₇(Gln)₅₅	1.61
BlaP₁₉₇₋₁₉₈(Gln)₃₀	1.85
BlaP₁₉₇₋₁₉₈(Gln)₅₅	1.85

6.1. Aggregation under native conditions

The aggregating properties of BlaP₂₁₆₋₂₁₇(Gln)₃₀ and BlaP₂₁₆₋₂₁₇(Gln)₅₅ were investigated in PBS, pH 7.5, at 37°C and compared in the same experiment with those of BlaP₁₉₇₋₁₉₈(Gln)₅₅. Under these conditions, very fast aggregation with no lag phase is observed for BlaP₂₁₆₋₂₁₇(Gln)₅₅ while BlaP₁₉₇₋₁₉₈(Gln)₅₅ aggregates significantly more slowly (Figure 43A). Analysis by TEM of the aggregates morphology at the end-point of the time-course (Figure 43B) indicates that BlaP₂₁₆₋₂₁₇(Gln)₅₅ forms fibrils while BlaP₁₉₇₋₁₉₈(Gln)₅₅ only forms amorphous aggregates as previously observed. Under the same conditions, BlaP₂₁₆₋₂₁₇(Gln)₃₀ slowly forms amorphous aggregates (Figure 43A and B).

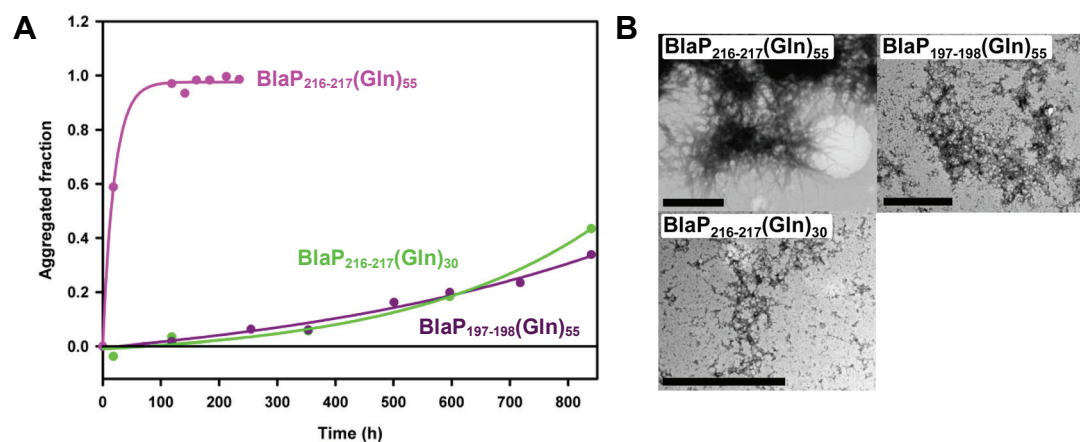


Figure 43: Aggregation properties of polyQ chimeras 216-217 under native conditions. (A) Aggregation kinetics of 110 μM BlaP₂₁₆₋₂₁₇(Gln)₃₀ (green), BlaP₂₁₆₋₂₁₇(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₅₅ (violet) at 37°C, in PBS, pH 7.5, monitored by measuring the concentration of protein remaining soluble. (B) TEM images of the aggregates formed by the proteins at the end of the incubation in PBS, pH 7.5 and at 37°C. The scale bar is 1 μm.

Taken together, these results suggest that the polyQ threshold length above which the chimeras 216-217 aggregate into fibrils under native conditions is comprised between 30 and 55Q. Very interestingly, this threshold value is therefore lower than that observed for the polyQ chimeras 197-198, which is between 55 and 79Q (Chapter 4, section 6.2).

As discussed in section 4, the native structure of the chimeras 216-217 is perturbed in the region of the insertion site. Consequently, we propose that the steric and/or conformational constraints applied to the polyQ tract are lower when the tract is in position 216-217, allowing the latter to be more flexible and accessible compared to a polyQ tract of the same length inserted in position 197-198. Thus, a 55Q tract inserted in 216-217 could fluctuate more freely and adopt much more frequently amyloid-competent conformations and/or be more accessible for the formation of the highly ordered intermolecular β -sheets characteristic of amyloid fibrils than when it is inserted in position 197-198.

Another interesting result is that BlaP₂₁₆₋₂₁₇(Gln)₅₅ forms fibrils at the same rate than BlaP₁₉₇₋₁₉₈(Gln)₇₉ (Figure 44). In the previous chapter, we showed that the structural integrity of BlaP₁₉₇₋₁₉₈ and thus the constraints imposed on the polyQ tract has negligible effect on the rate at which BlaP₁₉₇₋₁₉₈(Gln)₇₉ form fibrils. If fibril formation by both proteins under native conditions was only mediated by the polyQ interactions, BlaP₂₁₆₋₂₁₇(Gln)₅₅ should therefore aggregate more slowly than BlaP₁₉₇₋₁₉₈(Gln)₇₉. A possible explanation for our unexpected observation is that interactions between some regions of the BlaP moiety somehow increase the rate of fibril formation of chimeras 216-217 compared to that of chimeras 197-198.

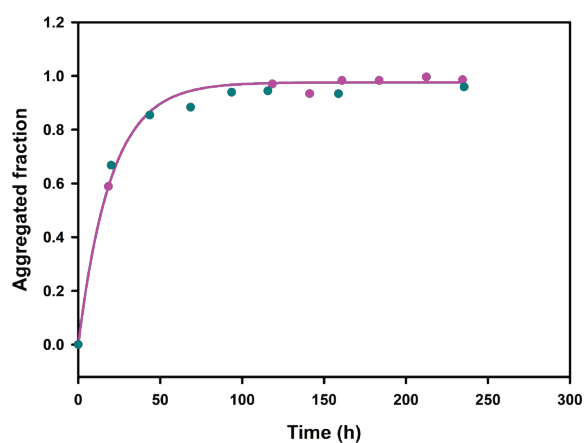


Figure 44: Comparison between the aggregation time-courses of BlaP₂₁₆₋₂₁₇(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₇₉ (dark cyan) under native conditions: PBS, pH 7.5 and 0 M urea at 37°C. The data correspond to those shown in Figure 43 for BlaP₂₁₆₋₂₁₇(Gln)₅₅ and to those shown in Figure 32 (page 71) for BlaP₁₉₇₋₁₉₈(Gln)₇₉.

6.2. Aggregation under denaturing conditions

Aggregation time-courses of BlaP₂₁₆₋₂₁₇(Gln)₃₀ and BlaP₂₁₆₋₂₁₇(Gln)₅₅ were also monitored in PBS, pH 7.5, 25°C, at a concentration in urea corresponding to their respective C_m (Table 19) and compared in the same experiment with that of BlaP₁₉₇₋₁₉₈(Gln)₅₅.

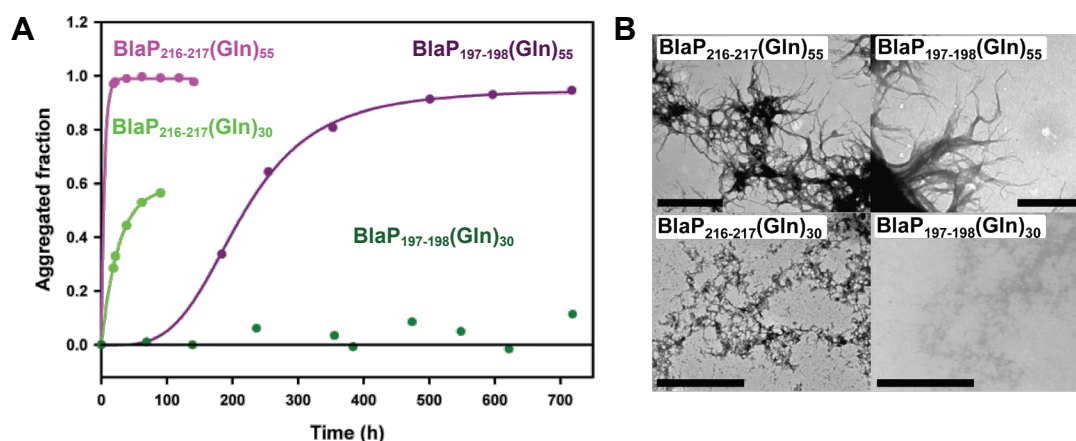


Figure 45: Aggregation properties of chimeras 216-217 under denaturing conditions. (A) Aggregation time-courses of 110 μM BlaP₂₁₆₋₂₁₇(Gln)₃₀ (green), BlaP₂₁₆₋₂₁₇(Gln)₅₅ (pink) and BlaP₁₉₇₋₁₉₈(Gln)₅₅ (violet) in PBS, pH 7.5, 25°C and in the presence of a urea concentration corresponding to their respective C_m (Table 19), monitored by measuring the concentration of protein remaining soluble. The aggregation time-course of BlaP₁₉₇₋₁₉₈(Gln)₃₀ (dark green) in PBS, pH 7.5, 25°C and C_m M urea coming from previous experiments (Figure 29, page 67) is shown here without error bar. (B) TEM images of the aggregates formed by the proteins at the end of the incubation at 25°C, in PBS, pH 7.5 and in the presence of C_m M urea. The scale bar is 1 μm.

Under these denaturing conditions, while the aggregation time-course of BlaP₁₉₇₋₁₉₈(Gln)₅₅ has the sigmoidal appearance characteristic of a nucleation-polymerization process (as described in Chapter 4, section 6.1), no lag phase is observed for BlaP₂₁₆₋₂₁₇(Gln)₅₅ (Figure 45A). The rate of aggregation of BlaP₂₁₆₋₂₁₇(Gln)₅₅ is ~73-fold faster than that of BlaP₁₉₇₋₁₉₈(Gln)₅₅ (i.e. the time for mid-aggregation is *ca.* 3 hours and *ca.* 220 hours, respectively). Aggregates formed by both chimeras with 55Q at the end-point of the time-course are fibrillar (Figure 45B).

BlaP₂₁₆₋₂₁₇(Gln)₃₀ also aggregates with no lag phase but in a lower extent than BlaP₂₁₆₋₂₁₇(Gln)₅₅ (Figure 45A). This protein is more aggregation-prone than its homolog from site 197-198 [BlaP₁₉₇₋₁₉₈(Gln)₃₀] since this latter does not significantly aggregates (Figure 45A). BlaP₂₁₆₋₂₁₇(Gln)₃₀ does not however form fibrils and only amorphous aggregates are visualized by TEM, as for BlaP₁₉₇₋₁₉₈(Gln)₃₀ (Figure 45B).

These results therefore indicate that under denaturing conditions (C_m M urea), the polyQ threshold length for fibril formation does not depend on the location of the insertion site. This threshold length is indeed comprised between 30 and 55Q for both chimeras 197-198 and

chimeras 216-217. As proposed in the section 6 of the previous chapter, these results suggest that short polyQ stretches (≤ 30 residues) rarely or never access amyloid-competent conformations, even if BlaP is in its unfolded state. Moreover, as observed for chimeras containing polyQ tracts in position 197-198, these results indicate that the unfolding of the β -lactamase moiety is not the driving force for underlying the fibrillar aggregation and that the polyQ length is a critical determinant for fibril formation.

Interestingly, the rate at which BlaP₂₁₆₋₂₁₇(Gln)₃₀ forms amorphous aggregates is much faster when the protein is incubated in the presence of a urea concentration corresponding to its C_m than when it is incubated under native conditions (Figure 46).

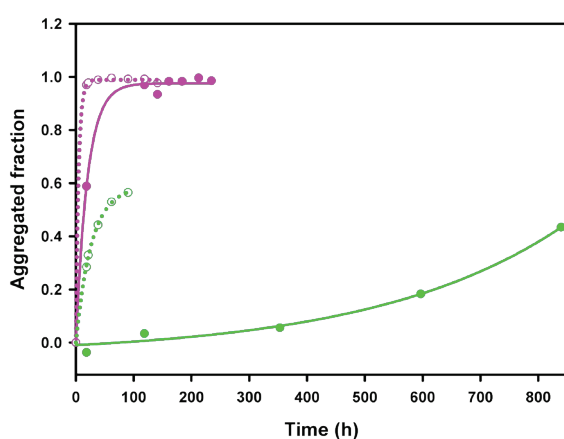


Figure 46: Comparison between the aggregation time-courses obtained with BlaP₂₁₆₋₂₁₇(Gln)₃₀ (green) and BlaP₂₁₆₋₂₁₇(Gln)₅₅ (pink) under native conditions (continuous line): PBS, pH 7.5 and 0 M urea at 37°C and under denaturing conditions (dotted line): PBS, pH 7.5, 25°C and at a concentration in urea corresponding to their respective C_m (Table 19).

This observation suggests that although the unfolding of BlaP is not the driving factor for fibril formation, it significantly enhances the formation of amorphous aggregates. The urea-unfolding transitions of BlaP₂₁₆₋₂₁₇ and its chimeras are characterized by lower m values than those of BlaP₁₉₇₋₁₉₈ and its chimeras. As already mentioned, this could arise from (i) the formation of a series of discrete partially unfolded species as the urea concentration increases or (ii) a more compact unfolded state. Thus, it is tempting to speculate that the interactions between BlaP regions exposed to the solvent in these species increase the rate of amorphous aggregation. Such interactions between BlaP regions could also rationalize, at least in part, the fact that:

- (i) BlaP₂₁₆₋₂₁₇(Gln)₅₅ aggregates faster into fibrils when the protein is incubated in 1.61 M urea than when it is incubated under native conditions, respectively (Figure 46),
- (ii) BlaP₂₁₆₋₂₁₇(Gln)₅₅ aggregates faster than BlaP₁₉₇₋₁₉₈(Gln)₅₅ when they are incubated in the presence of a urea concentration corresponding to their respective C_m (Figure 45A).

The interactions between regions of the BlaP moiety could enhance the rate at which BlaP₂₁₆₋₂₁₇(Gln)₅₅ aggregate into fibrils by a mechanism resembling a multi-domain aggregation process such as those described for Htt exon1 and ataxin 3.

7. Conclusions

In this chapter, we have characterized the catalytic, structural, thermodynamic and aggregating properties of chimeras with 30, 55 and 79 glutamines inserted in position 216-217 of BlaP (polyQ chimeras 216-217). The results obtained indicate that the insertion of the PG dipeptide destabilizes BlaP and slightly modifies its tertiary structure. These effects are more pronounced when a 30Q tract is inserted in this position; extension of the polyQ tract up to 79Q has however no further effect on the structure and stability of BlaP. Under all the conditions of aggregation investigated (i.e. in the absence or presence of urea), only the chimeras with 55 and 79Q are able to form fibrillar aggregates, clearly indicating that the polyQ length is the critical determinant for fibril formation.

Very interestingly, the threshold number of glutamines above which the chimeras 216-217 aggregate into fibrils does not apparently depend on the unfolding of the BlaP moiety; it is between 30 and 55 under both native and denaturing conditions. The threshold length for fibril formation under native conditions of chimeras 216-217 is therefore more similar to the pathological polyQ threshold length observed in polyQ diseases [around 35-45Q, Table 2 (page 7)] than that of chimeras 197-198 (between 55 and 79Q). The polyQ tract is probably more flexible and/or accessible in chimeras 216-217 than in chimeras 197-198 due to the structural changes induced in the BlaP moiety. Thus, the steric and/or conformational constraints imposed by BlaP in polyQ chimeras 216-217 may resemble more to those imposed on the polyQ tract in proteins associated with polyQ diseases.

On the other hand, the results we obtained suggest that interactions between regions of the BlaP moiety could assist the aggregation of the polyQ chimeras 216-217. Although this phenomenon should be investigated in more details, the mechanism of aggregation of the chimeras 216-217 could resemble to a multi-domain aggregation mechanism such as those described for Htt exon1 and ataxin 3.

Chapter 6. Conclusions and perspectives

1. Conclusions

We have created and characterized two sets of chimeras made of the β -lactamase BlaP and polyglutamine (polyQ) inserts: the polyQ chimeras 197-198 with 23, 30, 55 and 79Q inserted in position 197-198 and the polyQ chimeras 216-217 with 30, 55 and 79Q inserted in position 216-217. To our knowledge, it is the first time that:

- polyQ chimeras with inserted polyQ lengths higher than 50 residues were successfully produced, purified and characterized,
- the aggregation time courses of a polyQ-containing protein have been systematically monitored under different conditions of incubation (i.e. under conditions favoring different conformations of the host protein),
- two insertion sites were used in the same host protein to create polyQ chimeras. As summarized below and in Table 20, the properties of these two series of chimeras significantly differ. In both cases, the polyQ tracts however adopt a random conformation independently of their length and the length of the polyQ tract above a threshold is the determinant factor for the aggregation into fibrils.

Table 20: Summary of the main results obtained in this work

	Chimeras 197-198	Chimeras 216-217
EFFECTS OF POLYQ INSERTIONS ON THE:		
Catalytic properties	-	+
Tertiary structure	-	+
Secondary structure	-	(in the region of the insertion site)
STRUCTURE OF THE POLYQ TRACT	disordered	disordered
AGGREGATING PROPERTIES		
PolyQ threshold length for fibril formation under: native conditions	55-79	30-55
denaturing conditions	30-55	30-55
Modulating role of the BlaP moiety	+(inhibiting)	+(assisting)

+, effect. -, no effect.

The insertion of polyQ tracts in position 197-198 does not perturb the overall structure of BlaP but destabilizes the protein in an extent largely independent of the polyQ length. The investigation of the aggregating properties of these polyQ chimeras under both native and denaturing conditions indicates that there is a threshold number of glutamines above which the polyQ chimeras 197-198 aggregate into amyloid-like fibrils and that it depends on the

structural integrity of BlaP. The value of this threshold is indeed comprised between 55 and 79 glutamines and between 30 and 55 glutamines under native and denaturing conditions, respectively (Table 20). These results indicate that the native conformation of BlaP moiety reduces the ability of the 55Q tract to mediate aggregation into amyloid-like fibrils. We proposed that the modulating effects of BlaP on the propensity of the polyQ tract to trigger aggregation result from the imposition of steric and/or conformational constraints on the polyQ tract by the BlaP moiety. Interestingly, our results clearly show that the structural integrity of BlaP, and thus the constraints imposed on a 79Q tract, has negligible impact on the propensity of the latter to mediate amyloid-like fibril formation.

In contrast, the results obtained with polyQ chimeras 216-217 indicate that the tertiary structure of the BlaP moiety is slightly perturbed in these chimeras in a polyQ length-independent manner. Based on the X-ray structure of BlaP, we propose that these minor structural perturbations are essentially localized to the polypeptide region at both sides of the insertion site 216-217. The polyQ chimeras 216-217 are more destabilized than the polyQ chimeras 197-198. The destabilization of the BlaP moiety induces by polyQ insertions in position 216-217 is largely independent of the polyQ length as observed for that induced by polyQ insertion in position 197-198. However, the urea-induced equilibrium unfolding of BlaP₂₁₆₋₂₁₇ and its polyQ chimeras may diverge from a two-state process or the unfolded state of these proteins may significantly differ from that of chimeras 197-198. Finally, our preliminary results indicate that the aggregating properties of the chimeras 216-217 significantly differ from those of chimeras 197-198:

- under native conditions, the threshold number of glutamines above which chimeras 216-217 form fibrils is comprised between 30 and 55Q and is therefore lower than that observed for chimeras 197-198 under the same conditions of incubation (55 and 79Q). Thus the 55Q tract is able to trigger fibril formation when it is inserted in position 216-217 but not when inserted in position 197-198. We propose that the structural perturbations observed in chimeras 216-217 decrease the steric and/or conformational constraints imposed on the polyQ tract and this allows the 55Q tract to promote fibril formation. Moreover, the results we obtained so far also suggest that interactions between regions of the BlaP moiety could enhance the rate of fibril formation of chimeras 216-217 compared to chimeras 197-198.

- when chimeras are incubated at a urea concentration corresponding to their respective C_m , the polyQ threshold length for fibril formation of polyQ chimeras 216-217 is similar to that of chimeras 197-198 (between 30 and 55Q). All together, the aggregating properties of BlaP₂₁₆₋₂₁₇(Gln)₃₀ and BlaP₂₁₆₋₂₁₇(Gln)₃₀ under these conditions suggest that interactions involving the β -lactamase moiety may assist their aggregation. When incubated at a urea concentration corresponding to their respective C_m , chimeras 216-217 could aggregate into fibrils by a mechanism resembling a multi-domain aggregation process such as that described for the polyQ disease-associated protein, ataxin 3 and the Htt exon 1.

Taken together, the results obtained for both sets of chimeras highlight a very complex interplay between the propensity of the polyQ tract to mediate aggregation and the modulating effect of BlaP. The two sets of chimeras with polyQ tracts inserted either in position 197-198 or in position 216-217 present therefore valuable models to investigate in details which and how the properties of the host protein influence the aggregation behavior of polyQ-containing proteins. However, the data we accumulated so far have only allowed us to propose hypotheses to explain the influence of host protein factors in the aggregation of the chimeras. Thus, a range of experiments have to be carried out to further understand how the protein context influences the aggregating behavior of polyQ-containing protein.

2. Perspectives

This work resulted in a number of very interesting findings, including the different effects (polyQ) insertions in position 197-198 and 216-217 on the structure and the urea unfolding pathway of BlaP, and the different aggregating behaviors of chimeras 197-198 and 216-217. A number of experiments will be however required in order to get a comprehensive understanding of the delicate balance between the propensity of the polyQ to trigger aggregation and the modulating (inhibiting or assisting) effects of the BlaP moiety.

(a) Can we characterize and compare at the molecular level the effects of the insertion of the PG dipeptide and the glutamines in position 197-198 and 216-217 on the structure and dynamic of BlaP?

It is of great importance to get structural information at the molecular level for BlaP₁₉₇₋₁₉₈, BlaP₂₁₆₋₂₁₇ and their chimeras in order to assess the effects of the insertion on the structure and flexibility of the BlaP moiety, especially in the region of the insertion site. Those information should allow to rationalize at least in part the difference in the aggregation behavior of the

chimeras 197-198 and 216-217. The structure of BlaP₁₉₇₋₁₉₈ and BlaP₂₁₆₋₂₁₇ and their chimeras with 23Q could be determined by X-ray crystallography while NMR is probably more suitable for chimeras with longer polyQ tracts. Preliminary data obtained in the laboratory indicate that at least the chimeras 197-198 with up to 79Q are enough soluble and stable to carry out a range of NMR measurements.

In addition, limited proteolysis experiments will also be performed to assess the effect of insertions on the flexibility of the regions flanking the insertion site.

Note that the wild-type protein (i.e. without the PG dipeptide at any location) should also be characterized as a reference protein.

(b) Urea-induced equilibrium unfolding: Can we clarify the cause of the lower m values obtained with BlaP₂₁₆₋₂₁₇ and its chimeras? Is it due to the presence of partially unfolded intermediate species or to a more compact unfolded state?

Dynamic light scattering (DLS) measurements on protein samples completely unfolded should provide information on the compactness of the unfolded state. Moreover, monitoring the urea-unfolding process by NMR can provide evidence for the formation of partially unfolded species.

(c) Can we investigate in more details the aggregation process of BlaP₂₁₆₋₂₁₇ and its chimeras that resemble a multi-domain process?

The aggregation time-courses of BlaP₂₁₆₋₂₁₇ and its chimeras will be monitored under native and various denaturing conditions. The early step of aggregation will be particularly investigated using size-exclusion chromatography (SEC), dynamic light scattering (DLS), TEM and dot blot using monoclonal antibodies specific to polyQ sequences and BlaP. The results of these studies should not only clarify the assisting role of the BlaP moiety in the aggregation of the chimeras 216-217 but should also indicate which species (i.e. the partially unfolded intermediate species or the unfolded state) is the main species involved.

(d) Do the high molecular weight oligomers observed by SEC in BlaP₁₉₇₋₁₉₈(Gln)₇₉ samples play a role in the aggregation of this latter?

Techniques such as SEC and DLS will be used to further investigate the formation of oligomeric species and their role in the aggregation process. Moreover, we recently found that the high molecular weight oligomers can be removed from the solution by filtration with

filters of 0.02 μm porosity. Thus, the aggregation time-courses of the chimeras can be monitored in the absence and in the presence of these oligomers in the initial sample.

References

- Alba, M.M. & Guigo, R. 2004. Comparative analysis of amino acid repeats in rodents and humans. *Genome Res*, 14, 549-554.
- Albrecht, M., Hoffmann, D., Evert, B.O., Schmitt, I., Wullner, U. & Lengauer, T. 2003. Structural modeling of ataxin-3 reveals distant homology to adaptins. *Proteins*, 50, 355-370.
- Ambler, R.P., Coulson, A.F., Frere, J.M., Ghuysen, J.M., Joris, B., Forsman, M., Levesque, R.C., Tiraby, G. & Waley, S.G. 1991. A standard numbering scheme for the class A beta-lactamases. *Biochem J*, 276 (Pt 1), 269-270.
- Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R. & Finkbeiner, S. 2004. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature*, 431, 805-810.
- Bauer, H.H., Aebi, U., Haner, M., Hermann, R., Muller, M. & Merkle, H.P. 1995. Architecture and polymorphism of fibrillar supramolecular assemblies produced by in vitro aggregation of human calcitonin. *J Struct Biol*, 115, 1-15.
- Bauer, P.O. & Nukina, N. 2009. The pathogenic mechanisms of polyglutamine diseases and current therapeutic strategies. *J Neurochem*, 110, 1737-1765.
- Bennett, M.J., Huey-Tubman, K.E., Herr, A.B., West, A.P., Jr., Ross, S.A. & Bjorkman, P.J. 2002. Inaugural Article: A linear lattice model for polyglutamine in CAG-expansion diseases. *Proc Natl Acad Sci U S A*, 99, 11634-11639.
- Bevivino, A.E. & Loll, P.J. 2001. An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel beta -fibrils. *Proc Natl Acad Sci U S A*, 98, 11955-11960.
- Bhattacharyya, A., Thakur, A.K., Chellgren, V.M., Thiagarajan, G., Williams, A.D., Chellgren, B.W., Creamer, T.P. & Wetzel, R. 2006. Oligoproline effects on polyglutamine conformation and aggregation. *J Mol Biol*, 355, 524-535.
- Bhattacharyya, A.M., Thakur, A.K. & Wetzel, R. 2005. polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction. *Proc Natl Acad Sci U S A*, 102, 15400-15405.
- Bulone, D., Masino, L., Thomas, D.J., San Biagio, P.L. & Pastore, A. 2006. The interplay between PolyQ and protein context delays aggregation by forming a reservoir of protofibrils. *PLoS One*, 1, e111.
- Campioni, S., Monsellier, E. & Chiti, F. 2010. Why proteins misfold. In: Ramirez-Alvarado, M., Kelly, J.W. & Dobson, C.M. (eds.) *Protein Misfolding Diseases: Current and Emerging principles and Therapies*. John Wiley and Sons, Inc.
- Chen, S., Berthelie, V., Yang, W. & Wetzel, R. 2001. Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. *J Mol Biol*, 311, 173-182.
- Chen, S., Ferrone, F.A. & Wetzel, R. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc Natl Acad Sci U S A*, 99, 11884-11889.
- Chen, Y.W., Allen, M.D., Veprintsev, D.B., Lowe, J. & Bycroft, M. 2004. The structure of the AXH domain of spinocerebellar ataxin-1. *J Biol Chem*, 279, 3758-3765.
- Chevigne, A., Yilmaz, N., Gaspard, G., Giannotta, F., Francois, J.M., Frere, J.M., Galleni, M. & Filee, P. 2007. Use of bifunctional hybrid beta-lactamases for epitope mapping and immunoassay development. *J Immunol Methods*, 320, 81-93.
- Chiti, F. & Dobson, C.M. 2006. Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem*, 75, 333-366.
- Chow, M.K., Ellisdon, A.M., Cabrita, L.D. & Bottomley, S.P. 2004. Polyglutamine expansion in ataxin-3 does not affect protein stability: implications for misfolding and disease. *J Biol Chem*, 279, 47643-47651.

- Cooper, J.K., Schilling, G., Peters, M.F., Herring, W.J., Sharp, A.H., Kaminsky, Z., Masone, J., Khan, F.A., Delanoy, M., Borchelt, D.R., Dawson, V.L., Dawson, T.M. & Ross, C.A. 1998. Truncated N-terminal fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. *Hum Mol Genet*, 7, 783-790.
- Crick, S.L., Jayaraman, M., Frieden, C., Wetzel, R. & Pappu, R.V. 2006. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc Natl Acad Sci U S A*, 103, 16764-16769.
- Cummings, C.J., Mancini, M.A., Antalffy, B., DeFranco, D.B., Orr, H.T. & Zoghbi, H.Y. 1998. Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nat Genet*, 19, 148-154.
- Darnell, G., Orgel, J.P., Pahl, R. & Meredith, S.C. 2007. Flanking polyproline sequences inhibit beta-sheet structure in polyglutamine segments by inducing PPII-like helix structure. *J Mol Biol*, 374, 688-704.
- De Chiara, C., Menon, R.P., Adinolfi, S., De Boer, J., Ktistaki, E., Kelly, G., Calder, L., Kioussis, D. & Pastore, A. 2005a. The AXH domain adopts alternative folds the solution structure of HBP1 AXH. *Structure*, 13, 743-753.
- De Chiara, C., Menon, R.P., Dal Piaz, F., Calder, L. & Pastore, A. 2005b. Polyglutamine is not all: the functional role of the AXH domain in the ataxin-1 protein. *J Mol Biol*, 354, 883-893.
- De Meester, F., Joris, B., Reckinger, G., Bellefroid-Bourguignon, C., Frere, J.M. & Waley, S.G. 1987. Automated analysis of enzyme inactivation phenomena. Application to beta-lactamases and DD-peptidases. *Biochem Pharmacol*, 36, 2393-2403.
- Dehay, B. & Bertolotti, A. 2006. Critical role of the proline-rich region in Huntingtin for aggregation and cytotoxicity in yeast. *J Biol Chem*, 281, 35608-35615.
- Diaz-Hernandez, M., Moreno-Herrero, F., Gomez-Ramos, P., Moran, M.A., Ferrer, I., Baro, A.M., Avila, J., Hernandez, F. & Lucas, J.J. 2004. Biochemical, ultrastructural, and reversibility studies on huntingtin filaments isolated from mouse and human brain. *J Neurosci*, 24, 9361-9371.
- Difiglia, M., Sapp, E., Chase, K., Schwarz, C., Meloni, A., Young, C., Martin, E., Vonsattel, J.P., Carraway, R., Reeves, S.A. & Et Al. 1995. Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron*, 14, 1075-1081.
- Difiglia, M., Sapp, E., Chase, K.O., Davies, S.W., Bates, G.P., Vonsattel, J.P. & Aronin, N. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science*, 277, 1990-1993.
- Dobson, C.M. 2003. Protein folding and misfolding. *Nature*, 426, 884-890.
- Dobson, C.M. 2004a. Experimental investigation of protein folding and misfolding. *Methods*, 34, 4-14.
- Dobson, C.M. 2004b. Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol*, 15, 3-16.
- Duennwald, M.L., Jagadish, S., Muchowski, P.J. & Lindquist, S. 2006. Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc Natl Acad Sci U S A*, 103, 11045-11050.
- Dumoulin, M. & Dobson, C.M. 2004. Probing the origins, diagnosis and treatment of amyloid diseases using antibodies. *Biochimie*, 86, 589-600.
- Dumoulin, M., Kumita, J.R. & Dobson, C.M. 2006. Normal and aberrant biological self-assembly: Insights from studies of human lysozyme and its amyloidogenic variants. *Acc Chem Res*, 39, 603-610.

- Dumoulin, M., Last, A.M., Desmyter, A., Decanniere, K., Canet, D., Larsson, G., Spencer, A., Archer, D.B., Sasse, J., Muyltermans, S., Wyns, L., Redfield, C., Matagne, A., Robinson, C.V. & Dobson, C.M. 2003. A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature*, 424, 783-788.
- Duyao, M.P., Auerbach, A.B., Ryan, A., Persichetti, F., Barnes, G.T., Mcneil, S.M., Ge, P., Vonsattel, J.P., Gusella, J.F., Joyner, A.L. & Et Al. 1995. Inactivation of the mouse Huntington's disease gene homolog Hdh. *Science*, 269, 407-410.
- El Hajjaji, H., Dumoulin, M., Matagne, A., Colau, D., Roos, G., Messens, J. & Collet, J.F. 2009. The zinc center influences the redox and thermodynamic properties of Escherichia coli thioredoxin 2. *J Mol Biol*, 386, 60-71.
- Ellgaard, L. & Helenius, A. 2001. ER quality control: towards an understanding at the molecular level. *Curr Opin Cell Biol*, 13, 431-437.
- Ellisdon, A.M., Pearce, M.C. & Bottomley, S.P. 2007. Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J Mol Biol*, 368, 595-605.
- Ellisdon, A.M., Thomas, B. & Bottomley, S.P. 2006. The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J Biol Chem*, 281, 16888-16896.
- Faux, N.G., Bottomley, S.P., Lesk, A.M., Irving, J.A., Morrison, J.R., De La Banda, M.G. & Whisstock, J.C. 2005. Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res*, 15, 537-551.
- Gauthier, L.R., Charrin, B.C., Borrell-Pages, M., Dompierre, J.P., Rangone, H., Cordelieres, F.P., De Mey, J., Macdonald, M.E., Lessmann, V., Humbert, S. & Saudou, F. 2004. Huntingtin controls neurotrophic support and survival of neurons by enhancing BDNF vesicular transport along microtubules. *Cell*, 118, 127-138.
- Gething, M.J. & Sambrook, J. 1992. Protein folding in the cell. *Nature*, 355, 33-45.
- Glabe, C.G. & Kaye, R. 2006. Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. *Neurology*, 66, S74-78.
- Gourfinkel-An, I., Cancel, G., Duyckaerts, C., Faucheux, B., Hauw, J.J., Trotter, Y., Brice, A., Agid, Y. & Hirsch, E.C. 1998. Neuronal distribution of intranuclear inclusions in Huntington's disease with adult onset. *Neuroreport*, 9, 1823-1826.
- Graham, R.K., Deng, Y., Slow, E.J., Haigh, B., Bissada, N., Lu, G., Pearson, J., Shehadeh, J., Bertram, L., Murphy, Z., Warby, S.C., Doty, C.N., Roy, S., Wellington, C.L., Leavitt, B.R., Raymond, L.A., Nicholson, D.W. & Hayden, M.R. 2006. Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. *Cell*, 125, 1179-1191.
- Gunawardena, S., Her, L.S., Bruschi, R.G., Laymon, R.A., Niesman, I.R., Gordesky-Gold, B., Sintasath, L., Bonini, N.M. & Goldstein, L.S. 2003. Disruption of axonal transport by loss of huntingtin or expression of pathogenic polyQ proteins in Drosophila. *Neuron*, 40, 25-40.
- Gutekunst, C.A., Li, S.H., Yi, H., Mulroy, J.S., Kuemmerle, S., Jones, R., Rye, D., Ferrante, R.J., Hersch, S.M. & Li, X.J. 1999. Nuclear and neuropil aggregates in Huntington's disease: relationship to neuropathology. *J Neurosci*, 19, 2522-2534.
- Hallet, B., Sherratt, D.J. & Hayes, F. 1997. Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein. *Nucleic Acids Res*, 25, 1866-1867.
- Han, K.K., Tetaert, D., Debuire, B., Dautrevaux, M. & Biserte, G. 1977. [Sequential Edman degradation]. *Biochimie*, 59, 557-576.
- Hands, S.L. & Wyttenbach, A. 2010. Neurotoxic protein oligomerisation associated with polyglutamine diseases. *Acta Neuropathol*, 120, 419-437.

- Harper, J.D. & Lansbury, P.T., Jr. 1997. Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu Rev Biochem*, 66, 385-407.
- Harrison, P.M. 2006. Exhaustive assignment of compositional bias reveals universally prevalent biased regions: analysis of functional associations in human and *Drosophila*. *BMC Bioinformatics*, 7, 441.
- Hodgson, J.G., Agopyan, N., Gutekunst, C.A., Leavitt, B.R., Lepiane, F., Singaraja, R., Smith, D.J., Bissada, N., Mccutcheon, K., Nasir, J., Jamot, L., Li, X.J., Stevens, M.E., Rosemond, E., Roder, J.C., Phillips, A.G., Rubin, E.M., Hersch, S.M. & Hayden, M.R. 1999. A YAC mouse model for Huntington's disease with full-length mutant huntingtin, cytoplasmic toxicity, and selective striatal neurodegeneration. *Neuron*, 23, 181-192.
- Hollenbach, B., Scherzinger, E., Schweiger, K., Lurz, R., Lehrach, H. & Wanker, E.E. 1999. Aggregation of truncated GST-HD exon 1 fusion proteins containing normal range and expanded glutamine repeats. *Philos Trans R Soc Lond B Biol Sci*, 354, 991-994.
- Holmberg, M., Duyckaerts, C., Durr, A., Cancel, G., Gourfinkel-An, I., Damier, P., Faucheux, B., Trottier, Y., Hirsch, E.C., Agid, Y. & Brice, A. 1998. Spinocerebellar ataxia type 7 (SCA7): a neurodegenerative disorder with neuronal intranuclear inclusions. *Hum Mol Genet*, 7, 913-918.
- Ignatova, Z. & Gierasch, L.M. 2006. Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein. *J Biol Chem*, 281, 12959-12967.
- Ignatova, Z., Thakur, A.K., Wetzel, R. & Gierasch, L.M. 2007. In-cell aggregation of a polyglutamine-containing chimera is a multistep process initiated by the flanking sequence. *J Biol Chem*, 282, 36736-36743.
- Irwin, S., Vandelft, M., Pinchev, D., Howell, J.L., Graczyk, J., Orr, H.T. & Truant, R. 2005. RNA association and nucleocytoplasmic shuttling by ataxin-1. *J Cell Sci*, 118, 233-242.
- Johnson, S.M., Wiseman, R.L., Sekijima, Y., Green, N.S., Adamski-Werner, S.L. & Kelly, J.W. 2005. Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidoses. *Acc Chem Res*, 38, 911-921.
- Kar, K., Jayaraman, M., Sahoo, B., Kodali, R. & Wetzel, R. 2011. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat Struct Mol Biol*, 18, 328-336.
- Katti, S.K., Lemaster, D.M. & Eklund, H. 1990. Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. *J Mol Biol*, 212, 167-184.
- Kim, M.W., Chelliah, Y., Kim, S.W., Otwinowski, Z. & Bezprozvanny, I. 2009. Secondary structure of Huntingtin amino-terminal region. *Structure*, 17, 1205-1212.
- Klein, F.A., Pastore, A., Masino, L., Zeder-Lutz, G., Nierengarten, H., Oulad-Abdelghani, M., Altschuh, D., Mandel, J.L. & Trottier, Y. 2007. Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *J Mol Biol*, 371, 235-244.
- Knox, J.R. & Moews, P.C. 1991. Beta-lactamase of *Bacillus licheniformis* 749/C. Refinement at 2 Å resolution and analysis of hydration. *J Mol Biol*, 220, 435-455.
- Ko, J., Ou, S. & Patterson, P.H. 2001. New anti-huntingtin monoclonal antibodies: implications for huntingtin conformation and its binding proteins. *Brain Res Bull*, 56, 319-329.

- Ladurner, A.G. & Fersht, A.R. 1997. Glutamine, alanine or glycine repeats inserted into the loop of a protein have minimal effects on stability and folding rates. *J Mol Biol*, 273, 330-337.
- Lakhani, V.V., Ding, F. & Dokholyan, N.V. 2010. Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences. *PLoS Comput Biol*, 6, e1000772.
- Lee, J.M., Ivanova, E.V., Seong, I.S., Cashorali, T., Kohane, I., Gusella, J.F. & Macdonald, M.E. 2007. Unbiased gene expression analysis implicates the huntingtin polyglutamine tract in extra-mitochondrial energy metabolism. *PLoS Genet*, 3, e135.
- Legleiter, J., Mitchell, E., Lotz, G.P., Sapp, E., Ng, C., Difiglia, M., Thompson, L.M. & Muchowski, P.J. 2010. Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner in vitro and in vivo. *J Biol Chem*, 285, 14777-14790.
- Li, M., Chevalier-Larsen, E.S., Merry, D.E. & Diamond, M.I. 2007. Soluble androgen receptor oligomers underlie pathology in a mouse model of spinobulbar muscular atrophy. *J Biol Chem*, 282, 3157-3164.
- Li, W., Serpell, L.C., Carter, W.J., Rubinsztein, D.C. & Huntington, J.A. 2006. Expression and characterization of full-length human huntingtin, an elongated HEAT repeat protein. *J Biol Chem*, 281, 15916-15922.
- Lunkes, A. & Mandel, J.L. 1998. A cellular model that recapitulates major pathogenic steps of Huntington's disease. *Hum Mol Genet*, 7, 1355-1361.
- Lunkes, A., Trottier, Y., Fagart, J., Schultz, P., Zeder-Lutz, G., Moras, D. & Mandel, J.L. 1999. Properties of polyglutamine expansion in vitro and in a cellular model for Huntington's disease. *Philos Trans R Soc Lond B Biol Sci*, 354, 1013-1019.
- Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., Lawton, M., Trottier, Y., Lehrach, H., Davies, S.W. & Bates, G.P. 1996. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*, 87, 493-506.
- Martindale, D., Hackam, A., Wieczorek, A., Ellerby, L., Wellington, C., Mccutcheon, K., Singaraja, R., Kazemi-Esfarjani, P., Devon, R., Kim, S.U., Bredesen, D.E., Tufaro, F. & Hayden, M.R. 1998. Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. *Nat Genet*, 18, 150-154.
- Masino, L., Kelly, G., Leonard, K., Trottier, Y. & Pastore, A. 2002. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett*, 513, 267-272.
- Masino, L., Musi, V., Menon, R.P., Fusi, P., Kelly, G., Frenkiel, T.A., Trottier, Y. & Pastore, A. 2003. Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Lett*, 549, 21-25.
- Masino, L., Nicastro, G., Menon, R.P., Dal Piaz, F., Calder, L. & Pastore, A. 2004. Characterization of the structure and the amyloidogenic properties of the Josephin domain of the polyglutamine-containing protein ataxin-3. *J Mol Biol*, 344, 1021-1035.
- Matagne, A., Lamotte-Brasseur, J. & Frere, J.M. 1998. Catalytic properties of class A beta-lactamases: efficiency and diversity. *Biochem J*, 330 (Pt 2), 581-598.
- Matagne, A., Misselyn-Bauduin, A.M., Joris, B., Erpicum, T., Granier, B. & Frere, J.M. 1990. The diversity of the catalytic properties of class A beta-lactamases. *Biochem J*, 265, 131-146.
- Matilla, A., Roberson, E.D., Banfi, S., Morales, J., Armstrong, D.L., Burrig, E.N., Orr, H.T., Sweatt, J.D., Zoghbi, H.Y. & Matzuk, M.M. 1998. Mice lacking ataxin-1 display learning deficits and decreased hippocampal paired-pulse facilitation. *J Neurosci*, 18, 5508-5516.

- Mccampbell, A., Taylor, J.P., Taye, A.A., Robitschek, J., Li, M., Walcott, J., Merry, D., Chai, Y., Paulson, H., Sobue, G. & Fischbeck, K.H. 2000. CREB-binding protein sequestration by expanded polyglutamine. *Hum Mol Genet*, 9, 2197-2202.
- Mossuto, M.F., Dhulesia, A., Devlin, G., Frare, E., Kumita, J.R., De Laureto, P.P., Dumoulin, M., Fontana, A., Dobson, C.M. & Salvatella, X. 2010. The non-core regions of human lysozyme amyloid fibrils influence cytotoxicity. *J Mol Biol*, 402, 783-796.
- Myers, J.K., Pace, C.N. & Scholtz, J.M. 1995. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci*, 4, 2138-2148.
- Nagai, Y., Inui, T., Popiel, H.A., Fujikake, N., Hasegawa, K., Urade, Y., Goto, Y., Naiki, H. & Toda, T. 2007. A toxic monomeric conformer of the polyglutamine protein. *Nat Struct Mol Biol*, 14, 332-340.
- Nagai, Y., Tucker, T., Ren, H., Kenan, D.J., Henderson, B.S., Keene, J.D., Strittmatter, W.J. & Burke, J.R. 2000. Inhibition of polyglutamine protein aggregation and cell death by novel peptides identified by phage display screening. *J Biol Chem*, 275, 10437-10442.
- Nasir, J., Floresco, S.B., O'kusky, J.R., Diewert, V.M., Richman, J.M., Zeisler, J., Borowski, A., Marth, J.D., Phillips, A.G. & Hayden, M.R. 1995. Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygotes. *Cell*, 81, 811-823.
- Natalello, A., Frana, A.M., Relini, A., Apicella, A., Invernizzi, G., Casari, C., Gliozzi, A., Doglia, S.M., Tortora, P. & Regonesi, M.E. 2011. A major role for side-chain polyglutamine hydrogen bonding in irreversible ataxin-3 aggregation. *PLoS One*, 6, e18789.
- Nicastro, G., Masino, L., Esposito, V., Menon, R.P., De Simone, A., Fraternali, F. & Pastore, A. 2009. Josephin domain of ataxin-3 contains two distinct ubiquitin-binding sites. *Biopolymers*, 91, 1203-1214.
- Nicastro, G., Menon, R.P., Masino, L., Knowles, P.P., Mcdonald, N.Q. & Pastore, A. 2005. The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. *Proc Natl Acad Sci U S A*, 102, 10493-10498.
- Nikolov, D.B., Chen, H., Halay, E.D., Hoffman, A., Roeder, R.G. & Burley, S.K. 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc Natl Acad Sci U S A*, 93, 4862-4867.
- Nilsson, M.R. 2004. Techniques to study amyloid fibril formation in vitro. *Methods*, 34, 151-160.
- Nozaki, K., Onodera, O., Takano, H. & Tsuji, S. 2001. Amino acid sequences flanking polyglutamine stretches influence their potential for aggregate formation. *Neuroreport*, 12, 3357-3364.
- Nucifora, F.C., Jr., Sasaki, M., Peters, M.F., Huang, H., Cooper, J.K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V.L., Dawson, T.M. & Ross, C.A. 2001. Interference by huntingtin and atrophin-1 with cbp-mediated transcription leading to cellular toxicity. *Science*, 291, 2423-2428.
- Olshina, M.A., Angley, L.M., Ramdzan, Y.M., Tang, J., Bailey, M.F., Hill, A.F. & Hatters, D.M. 2010. Tracking mutant huntingtin aggregation kinetics in cells reveals three major populations that include an invariant oligomer pool. *J Biol Chem*, 285, 21807-21816.
- Ordway, J.M. & Detloff, P.J. 1996. In vitro synthesis and cloning of long CAG repeats. *Biotechniques*, 21, 609-610, 612.
- Ordway, J.M., Tallaksen-Greene, S., Gutekunst, C.A., Bernstein, E.M., Cearley, J.A., Wiener, H.W., Dure, L.S.T., Lindsey, R., Hersch, S.M., Jope, R.S., Albin, R.L. & Detloff, P.J.

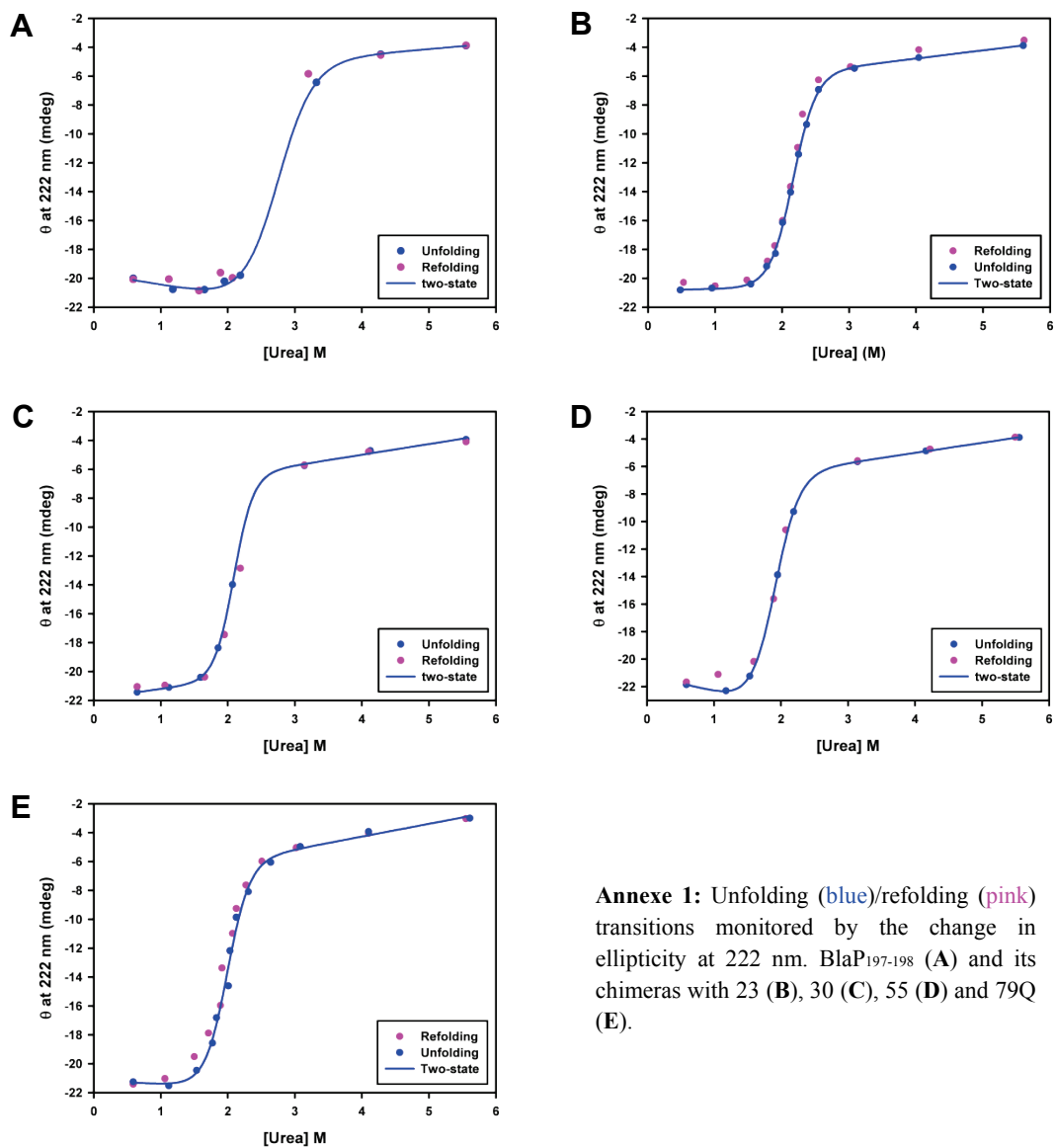
1997. Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell*, 91, 753-763.
- Orr, H.T. & Zoghbi, H.Y. 2007. Trinucleotide repeat disorders. *Annu Rev Neurosci*, 30, 575-621.
- Pace, C.N. & Scholtz, M. 1997. Measuring the conformational stability of a protein. In: Creighton, T.E. (ed.) *Protein structure: a practical approach*. Second Edition ed.: IRL press.
- Padiath, Q.S., Srivastava, A.K., Roy, S., Jain, S. & Brahmachari, S.K. 2005. Identification of a novel 45 repeat unstable allele associated with a disease phenotype at the MJD1/SCA3 locus. *Am J Med Genet B Neuropsychiatr Genet*, 133B, 124-126.
- Papaleo, E. & Invernizzi, G. 2011. Conformational diseases: structural studies of aggregation of polyglutamine proteins. *Curr Comput Aided Drug Des*, 7, 23-43.
- Paulson, H.L., Das, S.S., Crino, P.B., Perez, M.K., Patel, S.C., Gotsdiner, D., Fischbeck, K.H. & Pittman, R.N. 1997a. Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. *Ann Neurol*, 41, 453-462.
- Paulson, H.L., Perez, M.K., Trotter, Y., Trojanowski, J.Q., Subramony, S.H., Das, S.S., Vig, P., Mandel, J.L., Fischbeck, K.H. & Pittman, R.N. 1997b. Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron*, 19, 333-344.
- Pedersen, J.S., Dikov, D., Flink, J.L., Hjuler, H.A., Christiansen, G. & Otzen, D.E. 2006. The changing face of glucagon fibrillation: structural polymorphism and conformational imprinting. *J Mol Biol*, 355, 501-523.
- Pereira De Jesus-Tran, K., Cote, P.L., Cantin, L., Blanchet, J., Labrie, F. & Breton, R. 2006. Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. *Protein Sci*, 15, 987-999.
- Perutz, M.F., Johnson, T., Suzuki, M. & Finch, J.T. 1994. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A*, 91, 5355-5358.
- Peters-Libeu, C., Miller, J., Rutenber, E., Newhouse, Y., Krishnan, P., Cheung, K., Hatters, D., Brooks, E., Widjaja, K., Tran, T., Mitra, S., Arrasate, M., Mosquera, L.A., Taylor, D., Weisgraber, K.H. & Finkbeiner, S. 2012. Disease-Associated Polyglutamine Stretches in Monomeric Huntingtin Adopt a Compact Structure. *J Mol Biol*.
- Peters-Libeu, C., Newhouse, Y., Krishnan, P., Cheung, K., Brooks, E., Weisgraber, K. & Finkbeiner, S. 2005. Crystallization and diffraction properties of the Fab fragment of 3B5H10, an antibody specific for disease-causing polyglutamine stretches. *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 61, 1065-1068.
- Pinsky, L., Trifiro, M., Kaufman, M., Beitel, L.K., Mhatre, A., Kazemi-Esfarjani, P., Sabbaghian, N., Lumbroso, R., Alvarado, C., Vasiliou, M. & Et Al. 1992. Androgen resistance due to mutation of the androgen receptor. *Clin Invest Med*, 15, 456-472.
- Poirier, M.A., Li, H., Macosko, J., Cai, S., Amzel, M. & Ross, C.A. 2002. Huntingtin spheroids and protofibrils as precursors in polyglutamine fibrilization. *J Biol Chem*, 277, 41032-41037.
- Reddy, P.H., Williams, M., Charles, V., Garrett, L., Pike-Buchanan, L., Whetsell, W.O., Jr., Miller, G. & Tagle, D.A. 1998. Behavioural abnormalities and selective neuronal loss in HD transgenic mice expressing mutated full-length HD cDNA. *Nat Genet*, 20, 198-202.
- Robertson, A.L., Bate, M.A., Androulakis, S.G., Bottomley, S.P. & Buckle, A.M. 2011a. PolyQ: a database describing the sequence and domain context of polyglutamine repeats in proteins. *Nucleic Acids Res*, 39, D272-276.

- Robertson, A.L., Bate, M.A., Buckle, A.M. & Bottomley, S.P. 2011b. The rate of polyQ-mediated aggregation is dramatically affected by the number and location of surrounding domains. *J Mol Biol*, 413, 879-887.
- Robertson, A.L. & Bottomley, S.P. 2010. Towards the treatment of polyglutamine diseases: the modulatory role of protein context. *Curr Med Chem*, 17, 3058-3068.
- Robertson, A.L., Horne, J., Ellisdon, A.M., Thomas, B., Scanlon, M.J. & Bottomley, S.P. 2008. The structural impact of a polyglutamine tract is location-dependent. *Biophys J*, 95, 5922-5930.
- Ross, C.A. 1997. Intranuclear neuronal inclusions: a common pathogenic mechanism for glutamine-repeat neurodegenerative diseases? *Neuron*, 19, 1147-1150.
- Ruth, N., Quinting, B., Mainil, J., Hallet, B., Frere, J.M., Huygen, K. & Galleni, M. 2008. Creating hybrid proteins by insertion of exogenous peptides into permissive sites of a class A beta-lactamase. *FEBS J*, 275, 5150-5160.
- Saiki, M., Honda, S., Kawasaki, K., Zhou, D., Kaito, A., Konakahara, T. & Morii, H. 2005. Higher-order molecular packing in amyloid-like fibrils constructed with linear arrangements of hydrophobic and hydrogen-bonding side-chains. *J Mol Biol*, 348, 983-998.
- Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M.E. 1998. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell*, 95, 55-66.
- Saunders, H.M. & Bottomley, S.P. 2009. Multi-domain misfolding: understanding the aggregation pathway of polyglutamine proteins. *Protein Eng Des Sel*, 22, 447-451.
- Saunders, H.M., Gilis, D., Rومان, M., Dehouck, Y., Robertson, A.L. & Bottomley, S.P. 2011. Flanking domain stability modulates the aggregation kinetics of a polyglutamine disease protein. *Protein Sci*, 20, 1675-1681.
- Schaefer, M.H., Wanker, E.E. & Andrade-Navarro, M.A. 2012. Evolution and function of CAG/polyglutamine repeats in protein-protein interaction networks. *Nucleic Acids Res*.
- Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G.P., Davies, S.W., Lehrach, H. & Wanker, E.E. 1997. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. *Cell*, 90, 549-558.
- Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G.P., Lehrach, H. & Wanker, E.E. 1999. Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology. *Proc Natl Acad Sci U S A*, 96, 4604-4609.
- Schiene, C. & Fischer, G. 2000. Enzymes that catalyse the restructuring of proteins. *Curr Opin Struct Biol*, 10, 40-45.
- Schmidt, T., Lindenberg, K.S., Krebs, A., Schols, L., Laccone, F., Herms, J., Rechsteiner, M., Riess, O. & Landwehrmeyer, G.B. 2002. Protein surveillance machinery in brains with spinocerebellar ataxia type 3: redistribution and differential recruitment of 26S proteasome subunits and chaperones to neuronal intranuclear inclusions. *Ann Neurol*, 51, 302-310.
- Seidel, K., Den Dunnen, W.F., Schultz, C., Paulson, H., Frank, S., De Vos, R.A., Brunt, E.R., Deller, T., Kampinga, H.H. & Rub, U. 2010. Axonal inclusions in spinocerebellar ataxia type 3. *Acta Neuropathol*, 120, 449-460.
- Serpell, L.C., Fraser, P.E. & Sunde, M. 1999. X-ray fiber diffraction of amyloid fibrils. *Methods Enzymol*, 309, 526-536.
- Serpell, L.C., Sunde, M., Benson, M.D., Tennent, G.A., Pepys, M.B. & Fraser, P.E. 2000. The protofilament substructure of amyloid fibrils. *J Mol Biol*, 300, 1033-1039.

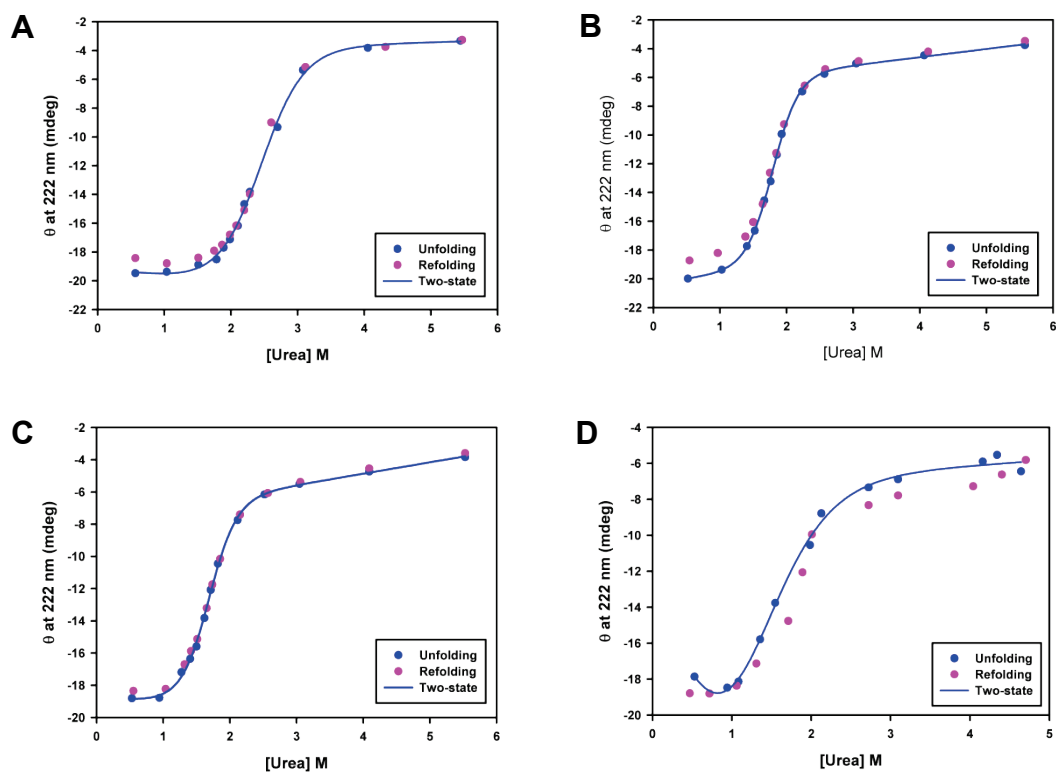
- Sharma, D., Sharma, S., Pasha, S. & Brahmachari, S.K. 1999. Peptide models for inherited neurodegenerative disorders: conformation and aggregation properties of long polyglutamine peptides with and without interruptions. *FEBS Lett*, 456, 181-185.
- Skinner, P.J., Koshy, B.T., Cummings, C.J., Klement, I.A., Helin, K., Servadio, A., Zoghbi, H.Y. & Orr, H.T. 1997. Ataxin-1 with an expanded glutamine tract alters nuclear matrix-associated structures. *Nature*, 389, 971-974.
- Song, A.X., Zhou, C.J., Peng, Y., Gao, X.C., Zhou, Z.R., Fu, Q.S., Hong, J., Lin, D.H. & Hu, H.Y. 2010. Structural transformation of the tandem ubiquitin-interacting motifs in ataxin-3 and their cooperative interactions with ubiquitin chains. *PLoS One*, 5, e13202.
- Stefani, M. & Dobson, C.M. 2003. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med*, 81, 678-699.
- Stenoien, D.L., Mielke, M. & Mancini, M.A. 2002. Intranuclear ataxin1 inclusions contain both fast- and slow-exchanging components. *Nat Cell Biol*, 4, 806-810.
- Sunde, M. & Blake, C. 1997. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv Protein Chem*, 50, 123-159.
- Takahashi, T., Kikuchi, S., Katada, S., Nagai, Y., Nishizawa, M. & Onodera, O. 2008. Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic. *Hum Mol Genet*, 17, 345-356.
- Tanaka, M., Machida, Y., Nishikawa, Y., Akagi, T., Hashikawa, T., Fujisawa, T. & Nukina, N. 2003. Expansion of polyglutamine induces the formation of quasi-aggregate in the early stage of protein fibrillization. *J Biol Chem*, 278, 34717-34724.
- Tanaka, M., Morishima, I., Akagi, T., Hashikawa, T. & Nukina, N. 2001. Intra- and intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases. *J Biol Chem*, 276, 45470-45475.
- Thakur, A.K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V.M., Byeon, I.J., Anjum, D.H., Kodali, R., Creamer, T.P., Conway, J.F., Gronenborn, A.M. & Wetzel, R. 2009. Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat Struct Mol Biol*, 16, 380-389.
- Thompson, J.R., Bratt, J.M. & Banaszak, L.J. 1995. Crystal structure of cellular retinoic acid binding protein I shows increased access to the binding cavity due to formation of an intermolecular beta-sheet. *J Mol Biol*, 252, 433-446.
- Tobelman, M.D., Kerby, R.L. & Murphy, R.M. 2008. A strategy for generating polyglutamine 'length libraries' in model host proteins. *Protein Eng Des Sel*, 21, 161-164.
- Tobelman, M.D. & Murphy, R.M. 2011. Location trumps length: polyglutamine-mediated changes in folding and aggregation of a host protein. *Biophys J*, 100, 2773-2782.
- Tsai, C.C., Kao, H.Y., Mitzutani, A., Banayo, E., Rajan, H., Mckeown, M. & Evans, R.M. 2004. Ataxin 1, a SCA1 neurodegenerative disorder protein, is functionally linked to the silencing mediator of retinoid and thyroid hormone receptors. *Proc Natl Acad Sci USA*, 101, 4047-4052.
- Uversky, V.N. & Fink, A.L. 2004. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta*, 1698, 131-153.
- Vandevenne, M., Filee, P., Scarafone, N., Cloes, B., Gaspard, G., Yilmaz, N., Dumoulin, M., Francois, J.M., Frere, J.M. & Galleni, M. 2007. The *Bacillus licheniformis* BlaP beta-lactamase as a model protein scaffold to study the insertion of protein fragments. *Protein Sci*, 16, 2260-2271.

- Vandevenne, M., Gaspard, G., Yilmaz, N., Giannotta, F., Frere, J.M., Galleni, M. & Filee, P. 2008. Rapid and easy development of versatile tools to study protein/ligand interactions. *Protein Eng Des Sel*, 21, 443-451.
- Vitalis, A., Wang, X. & Pappu, R.V. 2007. Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys J*, 93, 1923-1937.
- Walsh, D.M. & Selkoe, D.J. 2004. Oligomers on the brain: the emerging role of soluble protein aggregates in neurodegeneration. *Protein Pept Lett*, 11, 213-228.
- Wetzel, R. 2012. Physical Chemistry of Polyglutamine: Intriguing Tales of a Monotonous Sequence. *J Mol Biol*.
- Williams, A.J. & Paulson, H.L. 2008. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci*, 31, 521-528.
- Wu, X.C., Lee, W., Tran, L. & Wong, S.L. 1991. Engineering a *Bacillus subtilis* expression-secretion system with a strain deficient in six extracellular proteases. *J Bacteriol*, 173, 4952-4958.
- Yue, S., Serra, H.G., Zoghbi, H.Y. & Orr, H.T. 2001. The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding activity that is inversely affected by the length of its polyglutamine tract. *Hum Mol Genet*, 10, 25-30.
- Zander, C., Takahashi, J., El Hachimi, K.H., Fujigasaki, H., Albanese, V., Lebre, A.S., Stevanin, G., Duyckaerts, C. & Brice, A. 2001. Similarities between spinocerebellar ataxia type 7 (SCA7) cell models and human brain: proteins recruited in inclusions and activation of caspase-3. *Hum Mol Genet*, 10, 2569-2579.
- Zeitlin, S., Liu, J.P., Chapman, D.L., Papaioannou, V.E. & Efstratiadis, A. 1995. Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. *Nat Genet*, 11, 155-163.
- Zoghbi, H.Y. & Orr, H.T. 2000. Glutamine repeats and neurodegeneration. *Annu Rev Neurosci*, 23, 217-247.
- Zurdo, J. 2005. Polypeptide models to understand misfolding and amyloidogenesis and their relevance in protein design and therapeutics. *Protein Pept Lett*, 12, 171-187.

Annexes

Annexe 1

Annexe 1: Unfolding (blue)/refolding (pink) transitions monitored by the change in ellipticity at 222 nm. Blap₁₉₇₋₁₉₈ (A) and its chimeras with 23 (B), 30 (C), 55 (D) and 79Q (E).

Annexe 2

Annexe 2: Unfolding (blue)/refolding (pink) transitions monitored by the change in ellipticity at 222 nm. BlaP₂₁₆₋₂₁₇ (A) and its chimeras with 30 (B), 55 (C) and 79Q (D).

Annexe 3

Amyloid-like fibril formation by polyQ proteins: a critical balance between the polyQ length and the constraints imposed by the host protein

Natacha Scarafone¹, Coralie Pain¹, Anthony Fratamico^{1,3}, Gilles Gaspard^{2,4}, Nursel Yilmaz^{2,4}, Patrice Filée^{2,4}, Moreno Galleni², André Matagne¹ and Mireille Dumoulin^{1*}

¹Laboratory of Enzymology and Protein Folding, and ²Biological Macromolecules, Centre for Protein Engineering, Institute of Chemistry B6, University of Liège, 4000 Liège, Belgium - Current addresses: ³Laboratory of Plant Biochemistry and Photobiology, B22, University of Liège, 4000 Liège, Belgium; ⁴ProGenesis S.A., Boulevard du Rectorat, 27b, Sart-Tilman, 4000 Liège, Belgium.

Abstract

Nine neurodegenerative disorders, called polyglutamine (polyQ) diseases, are characterized by the formation of intranuclear amyloid-like aggregates by nine proteins containing a polyQ tract above a threshold length. These insoluble aggregates and/or some of their soluble precursors are thought to play a role in the pathogenesis. The mechanism by which polyQ expansions trigger the aggregation of the relevant proteins remains, however, unclear. In this work, polyQ tracts of different lengths were inserted into a solvent-exposed loop of the β -lactamase BlaP and the effects of these insertions on the properties of BlaP were investigated by a range of biophysical techniques. The insertion of up to 79 glutamines does not modify the structure of BlaP; it does, however, significantly destabilize the enzyme. The extent of destabilization is largely independent of the polyQ length, allowing us to study independently the effects intrinsic to the polyQ length and those related to the structural integrity of BlaP on the aggregating properties of the chimeras. Only chimeras with 55Q and 79Q readily form amyloid-like fibrils; therefore, similarly to the proteins associated with diseases, there is a threshold number of glutamines above which the chimeras aggregate into amyloid-like fibrils. Most importantly, the chimera containing 79Q forms amyloid-like fibrils at the same rate whether BlaP is folded or not, whereas the 55Q chimera aggregates into amyloid-like fibrils only if BlaP is unfolded. The threshold value for amyloid-like fibril formation depends, therefore, on the structural integrity of the β -lactamase moiety and thus on the steric and/or conformational constraints applied to the polyQ tract. These constraints have, however, no significant effect on the propensity of the 79Q tract to trigger fibril formation. These results suggest that the influence of the protein context on the aggregating properties of polyQ disease-associated proteins could be negligible when the latter contain particularly long polyQ tracts.

