Polyglutamine Monomer Structure and its Implications for Molecular Self-Assembly

Bryan VanSchouwen, MSc

Biotechnology (Chemical)

Submitted in partial fulfillment
of the requirements for the degree of

MSc in Biotechnology

Faculty of Graduate Studies, Brock University
St. Catharines, Ontario

Abstract:

Polyglutamine is a naturally occurring peptide found within several proteins in neuronal cells of the brain, and its aggregation has been implicated in several neurodegenerative diseases, including Huntington's disease. The resulting aggregates have been demonstrated to possess β-sheet structure, and aggregation has been shown to start with a single misfolded peptide. The current project sought to computationally examine the structural tendencies of three mutant polyglutamine peptides that were studied experimentally, and found to aggregate with varying efficiencies. Low-energy structures were generated for each peptide by simulated annealing, and were analyzed quantitatively by various geometry- and energy-based methods. According to the results, the experimentally-observed inhibition of aggregation appears to be due to localized conformational restraint placed on the peptide backbone by inserted prolines, which in turn confines the peptide to native coil structure, discouraging transition towards the β-sheet structure required for aggregation. Such knowledge could prove quite useful to the design of future treatments for Huntington's and other related diseases.

Acknowledgements:

A great deal of thanks goes out to my supervisors, Dr. Stuart M. Rothstein and Dr. Heather Gordon, for all of the helpful guidance and suggestions that they have provided over the course of my thesis research. Thanks also goes out to Dr. Adonis Skandalis, Daniel Oblinsky, and to our Japanese collaborators, Dr. Shigenori Tanaka and his students.

TABLE OF CONTENTS

List of Tables:

List of Figures:

1. INTRODUCTION

1.1. Reason for Studying Polyglutamine: Huntington's Disease and the Huntingtin Protein

Huntington's disease (HD) is a hereditary neurodegenerative disease of the brain. It is characterized by behaviour problems, personality changes and dementia that typically begin in midadulthood, and results in death within a few years of onset.[1,2,3] The disease is caused by an abnormal expansion of CAG trinucleotide repeats, which encode the amino acid L-glutamine, within exon 1 of the *IT-15* gene. This in turn results in production of huntingtin protein containing an abnormally long sequence of L-glutamine amino acid residues, known as an expanded poly-L-glutamine sequence, near its amino terminus.[1,2,4,5]

Huntingtin is a protein that is normally found in healthy neuronal cells, although its normal biological function remains unclear. Some studies have argued that huntingtin is important in the uptake and utilization of iron within neurons, assisting in the conversion of iron to a bioavailable form.[6,7] Others suggest that huntingtin is involved in intracellular transport, in conjunction with other known transport proteins.[8,9] More notably, a study of mice lacking the gene for huntingtin has demonstrated that huntingtin is important in proper embryonic development.[10] Nevertheless, abnormal expansion of the poly-L-glutamine region of huntingtin leads to a characteristic formation of insoluble proteinaceous plaques in neuronal cells, which is believed to play a key role in the neuronal cell death associated with HD. While shorter poly-L-glutamine sequences are not associated with occurrence of the disease, HD almost invariably occurs when the poly-L-glutamine sequence exceeds 37 residues in length, with progressively longer sequences resulting in onset of the disease at an exponentially earlier age.[1,2,4,5,11]

Poly-L-glutamine (polyQ) is a naturally occurring polypeptide found in a number of different proteins, and polyQ sequences exceeding a certain critical length have been implicated in several neurodegenerative diseases, including HD.[2,12,13] The presence of such extended polyQ sequences leads to conformational changes within the protein, and subsequent protein aggregation via polyQ sequence association (see Figure 1), ultimately resulting in formation of the disease-associated protein plaques.[1,3,12,13,14,15] However, it should be noted that some studies have found the soluble oligomers of abnormally-folded polyQ to cause cell death even in the absence of larger aggregates, suggesting the individual oligomers themselves to be toxic to neuronal cells.[1,16,17] In either case, toxicity may result from disruptive interactions with, or intra-aggregate incorporation of, other cellular proteins or disruption of the cell membrane.[18,19,20,21]

Recently, it has been suggested that the polyQ sequence of huntingtin may serve as an osmotically-stable glutamine reservoir, which can be cleaved from huntingtin and degraded to provide a source of glutamine, and its metabolite glutamate, when the levels of these vital amino acids within the host neuronal cells become too low.[22] Expansion of the polyQ sequence may gradually overwhelm the normal degradation mechanism with age, resulting in the release of intact polyQ chains which could then assemble into oligomers and larger aggregates. Indeed, others have suggested that the polyQ chain is cleaved from huntingtin prior to aggregation,[1,3,23] and several *in vitro* and *in vivo* aggregation studies of synthetic polyQ peptides have demonstrated a high capacity of isolated polyQ for aggregation, thus supporting the idea that polyQ forms aggregates when isolated from its host protein.[24,25,26] As a result of such abnormal polyQ processing, the supply of glutamine and glutamate provided by the polyQ chains would be

progressively cut off, thus potentially leading to a harmful deficiency of glutamine and glutamate contributing to neuronal cell death.[22]

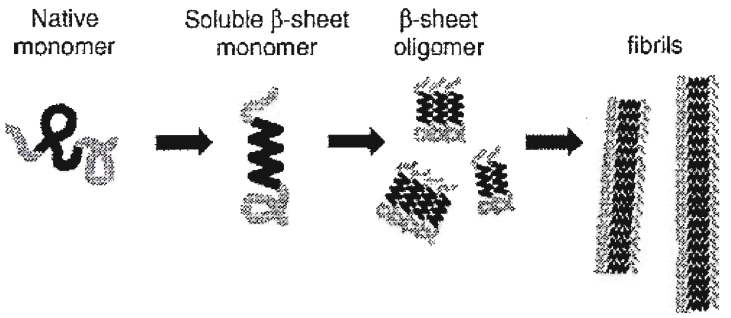


Figure 1: General schematic inferred for the aggregation of polyglutamine-containing proteins.[16] A conformational shift within the native structure converts the polyglutamine chain (highlighted in black) to an abnormal β-sheet structure, which then self-associates into β-sheet oligomers, fibrillar aggregates, and eventually, plaques.

Whatever the mechanism of toxicity, it is clear that aggregation of abnormally-folded polyQ is somehow associated with the neuronal cell death of HD. Indeed, Poirier *et al.* (2005)[27] found that huntingtin amino-terminal fragments capable of aggregation within mammalian neuronal cells (*via* polyQ association) induced visible cytotoxicity upon aggregation, while huntingtin amino-terminal fragments that failed to form visible aggregates induced no such cytotoxicity. In addition, *in vitro* and *in vivo* aggregation studies of polyQ performed by Scherzinger *et al.* (1997),[3] Ignatova and Gierasch (2006)[28] and others[5,15,29] have indicated that a minimum polyQ chain length of approximately 37-40 residues is required for abnormal polyQ folding and aggregation to occur spontaneously under normal physiological conditions, in agreement with clinical data indicating the occurrence of HD at such polyQ chain lengths. Finally, the rate of

aggregation increases as polyQ chain length increases, which may explain the decrease in the age of HD onset that occurs as the polyQ sequence increases in length.[25,28]

1.2. Polyglutamine Structure and Aggregation

Polyglutamine (polyQ) chains have been shown by a number of experimental studies to assemble into highly ordered and stable fibrillar aggregates, following transition from harmless native states to abnormal molecular structures (as outlined in Figure 1). The native state of polyQ is believed to be a coil structure, as suggested from circular dichroism and NMR experiments on polyQ-containing proteins, although the true native structure of polyQ in huntingtin is not known with certainty.[21,25,30] However, X-ray diffraction, electron microscopy, circular dichroism and solid-state NMR data have indicated that when polyQ aggregates, it forms linear, unbranched fibrils with a characteristic "cross-β" structure.[3,15,18,24,31,32,33,34,35,36] This structure consists of β-sheets oriented parallel to the fibril axis, with their β-strands nearly perpendicular to the axis (as outlined in Figure 2).

In particular, X-ray diffraction patterns for polyQ aggregates have invariably demonstrated a characteristic 4.8 Å meridional reflection, indicating the presence of hydrogen-bonded β-strands oriented perpendicular to the fibril axis. Circular dichroism data and internuclear couplings observed in NMR data have also confirmed the presence of β-strand-containing structures within the aggregates, and electron microscopy observations have highlighted the linear nature of the fibrils. Also, the binding of Congo red dye to polyQ aggregates produces an apple-green birefringence upon examination with polarized light microscopy, indicating an alignment of the dye molecules along

linear, regular fibril structures.[37,38,39,40] Illustrations of example cross-β structures for an antiparallel β-sheet and a parallel β-helix are provided in Figure 2.

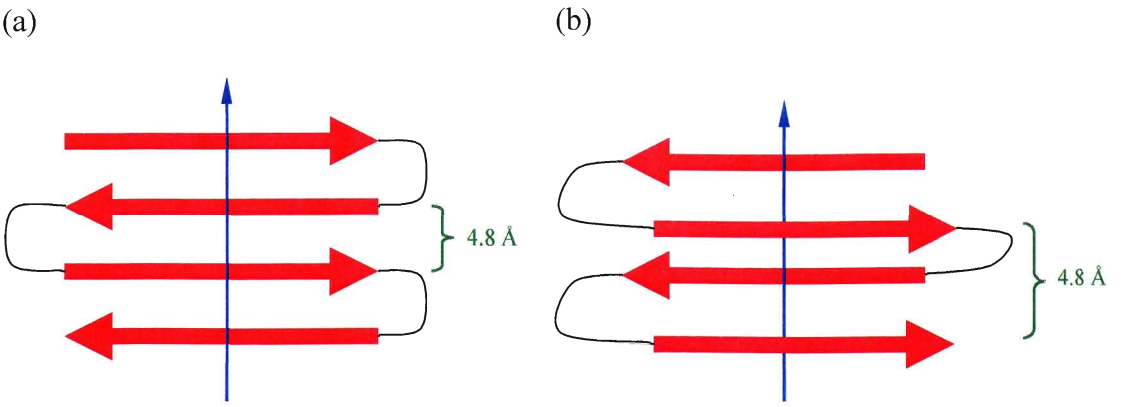


Figure 2: Schematic illustrations of cross-β structures for (a) an antiparallel β-sheet, and (b) a parallel β-helix. All β-strands are illustrated as red arrows, the direction of the fibril axis is indicated by a blue arrow, and the 4.8 Å spacing between hydrogen-bonded β-strands (corresponding to the meridional 4.8 Å X-ray diffraction feature) is labelled.

Within the inferred β-sheet structure, neighbouring β-strands are bound to each other by stacks of hydrogen bond interactions between both the peptide bond functional groups of the polyQ backbone, and the amide groups of the glutamine side chains (see Figure 3). These hydrogen-bond stacks span the entire length of the fibril, forming a so-called "polar zipper" which is crucial to fibril formation and stabilization.[24,34,41,42] Such participation of the side chains in hydrogen bonding was confirmed by NMR data, in which distinct ^{13}C and ^{15}N couplings indicated the presence of an ordered arrangement of the side chain amides.[43,44] In addition, mutation studies of fibril-forming proteins, including huntingtin, have demonstrated that amino acid substitutions within the fibril-forming segments of these proteins typically diminished aggregation, highlighting the importance of the side chains to aggregate assembly.[37,45] Once formed, the polyQ fibrils can then associate with each other via hydrophobic and/or electrostatic interactions to form larger fibres, and eventually, the proteinaceous plaques associated with HD.

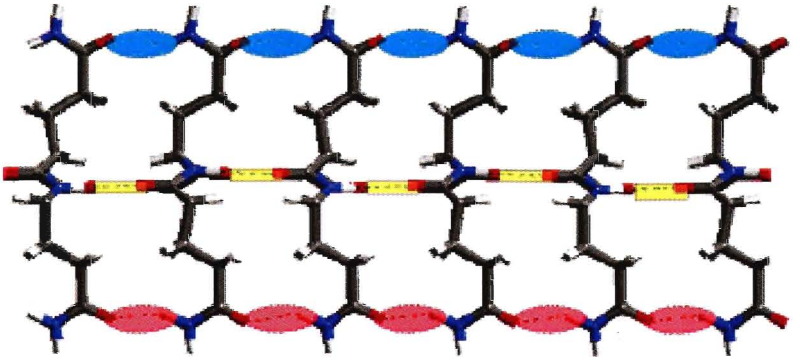


Figure 3: The polar zipper of polyglutamine β-sheet structures.[35] Backbone hydrogen bonds are highlighted as yellow and red bars, and side chain hydrogen bonds are highlighted in pink and cyan. The individual β-strands are shown in side view, and illustrated using the standard atomic colour scheme: gray = carbon; white = hydrogen; blue = nitrogen; red = oxygen.

It has also been demonstrated that polyQ assembles into fibrils by means of a cooperative nucleation-growth mechanism.[13,14,21,23,25,28,29,34, 46] In this mechanism, a critical nucleus for aggregation, consisting of a small section of the β-sheet fibril structure, is first formed from a small number of abnormally-folded polyQ chains (in a process called nucleation). Then, the aggregate grows quickly into a fibril as additional polyQ chains are added onto the initial nucleus, which serves as a template for misfolding and binding of the additional polyQ chains. This mechanism was supported by the observation of a lag period in aggregation kinetics studies.[14,21,25,29,33,46] During the lag period, the nucleus would assemble slowly due to an energetically unfavourable loss of entropy as the polyQ structure becomes more ordered, as well as kinetic barriers in the transition to the required β-sheet structure. After the lag phase, however, aggregation would be more rapid due to favourable enthalpy changes, which result from the formation of progressively more extensive polar-zipper hydrogen bonding as more polyQ chains are added to the developing fibril.

Furthermore, Chen *et al.* (2002)[25] and Thakur and Wetzel (2002)[26] both deduced from measured aggregation kinetic parameters that a single polyQ chain serves as the critical nucleus for fibril formation. Additionally, Bhattacharyya *et al.* (2005)[47] deduced a free energy of nucleation of approximately 12.2 kcal/mol from the measured equilibrium constant for nucleation. Thus, nucleation likely involves an unfavourable structure transition within the polyQ monomer, during which the monomer assembles into a β-sheet structure that acts as the required nucleus for further aggregation.

1.3. Proposed Structural Models for Polyglutamine Aggregates

Various models demonstrating cross-β structure (see Section 1.2) have been proposed for the β-sheet structure of polyQ and its aggregates, and a controversy still exists as to which model is correct.[11] A major reason for such controversy is the insolubility of polyQ in many experimental solvents, which makes structural examination by standard, high-resolution methods (such as X-ray crystallography and solution NMR) quite difficult.[13,45] Consequently, there has been a great deal of reliance on theoretical studies, and on lower-resolution experimental methods such as X-ray fibril diffraction and solid-state NMR, with rather sparse experimental evidence of higher-resolution structure. In addition, variations in fibril packing within polyQ aggregates could influence the observed X-ray and NMR results, leading to variations between the results – and subsequent interpretations – of different experiments. Therefore, verification of any one proposed structural model, and rejection of the others, with certainty has proven infeasible at this point in time.[33,48]

PolyQ peptides have been reported as being able to form a number of abnormal structures. For instance, a number of studies have reported the presence of various

intermediate species that arise during the aggregation process, ranging from small oligomers to spherical and amorphous aggregates, although some of these intermediates may be off-pathway polyQ assemblies that compete with the main fibril formation pathway.[4,20,23,49] In addition, some theoretical[11,50] and experimental[4,51,52,53] studies have suggested that polyQ peptides can assemble into annular structures. In this type of structure, shorter peptides form tube-like assemblies that resemble β-barrels, with their polyQ chains inclined at approximately 45° to the tube axis. Longer peptides, on the other hand, assemble into large ring-like structures that resemble disordered β-helices, with their polyQ chains nearly perpendicular to the ring axis. For instance, atomic force microscopy work by Wacker *et al.* (2004)[4] found that huntingtin containing an expanded polyQ sequence could form small annular structures, and X-ray diffraction work by Elam *et al.* (2003)[53] found that the polyQ segment of mutant SOD1 also formed annular structures. To date, however, two predominant classes of underlying structure have been proposed for polyQ aggregates: ribbon-like fibrils, and β-helix-based fibrils.[54]

1.3.1. Ribbon-like Polyglutamine Fibrils

Some researchers have argued that polyQ chains may form aggregates similar to those formed by the proteins associated with Alzheimer's, Parkinson's, and several other so-called "amyloid diseases".[15,31,35,36,37,42,55] In this model, the chains first assemble side-by-side into long β-sheets stabilized by polar-zipper hydrogen bonding.[23,31,32,37,42,49,56] Two or more β-sheets then stack together, with the meeting side chains of each pair of adjacent sheets interdigitating to form "steric zippers" between the sheets (see Figure 4). This gives rise to a ribbon-like fibril, which can then associate with others to form larger aggregates. Within such aggregates, the constituent polyQ chains could potentially form

one of a number of possible subunit structures, including extended β-strands, β-hairpins, or compact β-sheets (see Figure 5).[54] However, compact β-sheets would be more likely for long polyQ chains, as folding of the chain would more effectively constrain the ribbon width to within the fibril widths that have been observed by electron microscopy.[57] In addition, the β-sheet structure within the fibrils often possesses a left-handed twist, due to the left-handed chirality of the constituent amino acid residues, resulting in fibrils with a twisted or helical ribbon structure.[32,33,49]

(a) (b)

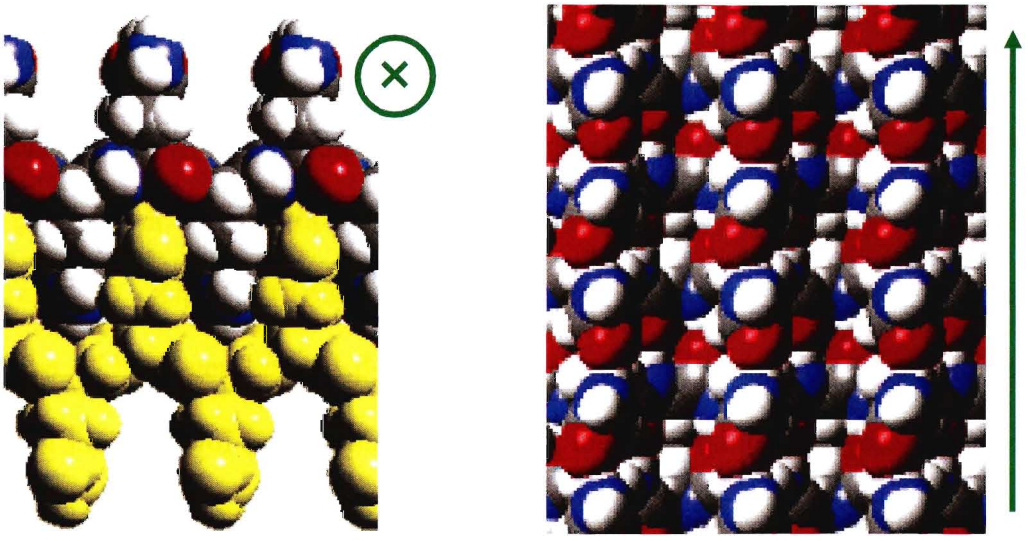


Figure 4: Ribbon structure model of polyglutamine aggregates, composed of stacked β-sheets formed by polyglutamine chains.[35] A segment of a fibril composed of two stacked β-sheets is shown: (a) end view of fibril; (b) top view of fibril. For clarity, the direction of the fibril axis is indicated by a green arrow (illustrated pointing into the page as a circled "X", and pointing along the page as a standard arrow). Also, one β-sheet is highlighted in yellow to distinguish the two sheets from each other. The other sheet is illustrated using the standard atomic colour scheme: gray = carbon; white = hydrogen; blue = nitrogen; red = oxygen. Note the interdigitation of side chains from adjacent β-sheets, which is clearly visible in the fibril end view, and the ridges of hydrogen-bonded side chains (characteristic of a polar zipper), which are clearly visible in the fibril top view.

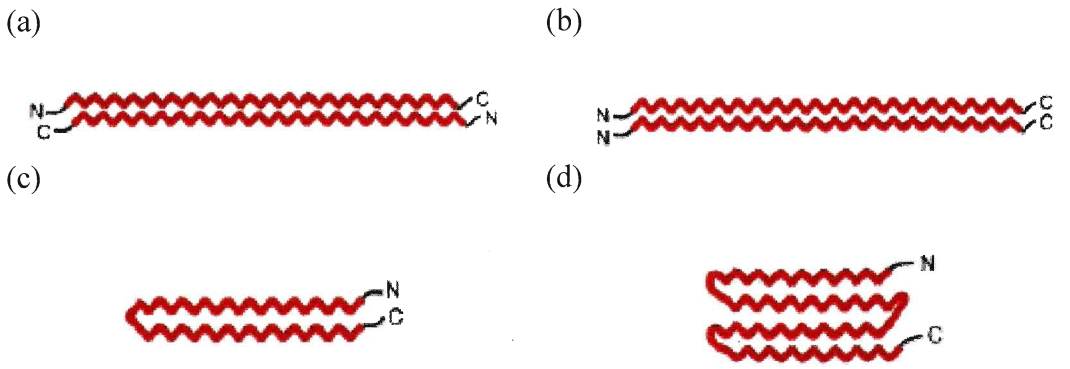


Figure 5: Examples of possible underlying β-sheet structures within ribbon-like fibrils of polyglutamine: (a) antiparallel or (b) parallel arrangements of extended β-strands; (c) β-hairpins; (d) compact β-sheets.[54] The amino-terminus and carboxy-terminus are indicated by "N" and "C", respectively.

Particularly notable evidence for ribbon-like fibril structure comes from a number of recent X-ray diffraction studies, including those performed by Sikorski and Atkins (2005),[35] Sharma *et al.* (2005)[36] and Sawaya *et al.* (2007).[37] In these studies, the observed polyQ X-ray diffraction patterns exhibited a strong 4.8 Å meridional reflection indicative of hydrogen-bonded β-strands, as well as an appreciable 8-11 Å equatorial reflection indicative of β-sheet stacking. Thus, the underlying aggregate structure was determined to be a cross-β structure composed of stacked β-sheets: a ribbon-like fibril structure.

In addition, a number of computational studies on this polyQ model have suggested that this may indeed be a plausible structure for polyQ aggregates. In particular, Bellesia and Shea (2007)[31] and Marchut and Hall (2007)[11] observed the spontaneous formation of β-sheet structures in coarse-grained MD simulations of systems of polyQ chains. Marchut and Hall (2007) even observed the formation of a number of possible β-hairpin and compact β-sheet structures, with longer polyQ chains exhibiting a tendency towards compact β-sheets, as expected from experiment. Therefore, a ribbon-

like fibril structure is a valid model for polyQ aggregates, with the extended polyQ chains likely forming compact β-sheet subunits.

1.3.2. The Perutz Model of Polyglutamine: β-Helix-Based Fibrils

Another notable model for huntingtin polyQ has been a β-helical structure, which was originally proposed by Perutz *et al.* (2002),[34] and supported by a number of other studies.[16,58,59] In the Perutz model for polyQ, sufficiently long polyQ chains fold into a β-helix structure that forms the underlying subunit of polyQ fibrils.[34,48] This structure consists of a chain of amino acid residues in β-strand-like conformations, coiled to form a wide left-handed helix in which adjacent helical turns are hydrogen-bonded to each other, in a configuration similar to that of a parallel β-sheet (see Figure 6). The resulting helix has 20 residues per turn, a pitch of approximately 4.8 Å,[60] and amino acid side chains that alternate between facing the interior and facing the exterior of the helix.

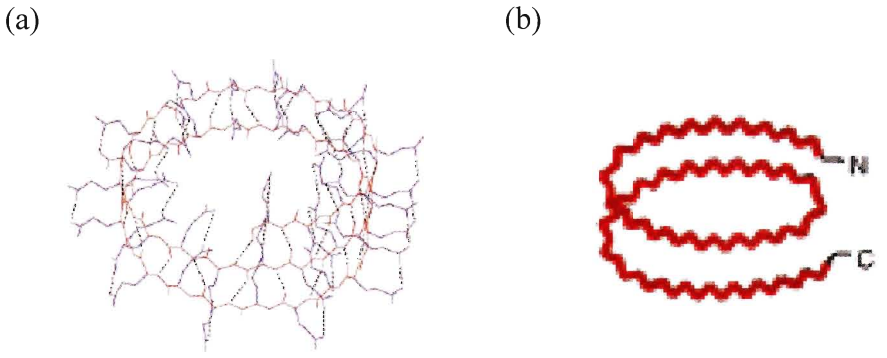


Figure 6: The β-helical structure model of polyglutamine, as deduced by Perutz *et al.* (2002).[34,54] (a) Stick model representation, with the peptide backbone shown in orange, the glutamine side chains in gray, and hydrogen bonds as dotted lines. (b) Ribbon model representation, with the amino-terminus and carboxy-terminus indicated by "N" and "C", respectively.

To deduce the structure of polyQ, Perutz *et al.* (2002)[34] examined X-ray diffraction, electron microscopy, and other experimental data for free poly-L-glutamine

and the exon-1 peptide of human huntingtin. They determined that the only possible structures which could account for all of the data, as well as intramolecular steric considerations, were helical β-sheet structures consisting of parallel β-strands wound around a central axis, with 20 residues per turn. In particular, the X-ray diffraction data exhibited the strong 4.8 Å meridional reflection typical of a cross-β structure (described in Section 1.2), but the 8-11 Å equatorial reflection that would otherwise indicate the presence of stacked β-sheets was absent or greatly diminished. Therefore, Perutz concluded that rather than forming stacked β-sheets, the polyQ must adopt a coiled β-sheet structure, with its β-strands nearly perpendicular to the coil axis: a β-helix.

The Perutz model also provides a possible explanation for the observed polyQ critical chain length of 37-40 residues that is required for aggregation. According to Perutz *et al.* (2002), the first sufficiently stable structure for a β-helical polyQ aggregate nucleus would occur when the chain reached 37-40 residues in length. At this length, the polyQ would form approximately two helical turns, thus permitting stabilization by polar-zipper hydrogen bonding around the entire circumference of the helix. However, if the chain was only 20 or 30 residues in length, the lesser extent of hydrogen bonding would be insufficient to properly stabilize the β-helical structure around its entire circumference, thus discouraging its formation. Once the nucleus has formed, outward growth of the developing fibril would occur primarily at its ends, with polyQ monomers stacking onto each other to form a fibril with continuous β-helix structure along its entire length. The fibrils could then associate with each other over time, via a combination of hydrophobic and electrostatic interactions, to give rise to larger aggregates.

Indeed, by performing molecular dynamics simulations of several proposed structures for polyQ, including several β-hairpin and β-sheet structures, Elliot *et al.* (2006)[13] suggested that the most reasonable structure for polyQ may be a Perutz β-helix, based on the fact that only the Perutz helix demonstrated a systematic increase in stability with polyQ chain length, as expected from experiment. In addition, Marchut and Hall (2007)[11] occasionally observed the spontaneous formation of β-helix structures in their coarse-grained MD simulations of 48-residue polyQ chains. Finally, root-mean-square deviations (RMSDs) from the Perutz helix structure during the polyQ simulations performed by Ogawa *et al.* (2007)[12] have demonstrated an increase in helical structure stability with addition of polyQ chains, which is consistent with the idea that the aggregation mechanism is cooperative in nature.

It should be noted, though, that simulation studies by Khare *et al.* (2005)[41] and Ogawa *et al.* (2007)[12] have suggested that a more stable configuration of β-sheet hydrogen bonds would be formed within the β-helix if it had 18.5 residues per turn, rather than 20 residues per turn. Also, the findings of Marchut and Hall (2007),[11] Ogawa *et al.* (2007)[12] and Elliot *et al.* (2006)[13] seem to suggest that the polyQ monomer would actually adopt a less regular helix structure than that originally proposed by Perutz *et al.* (2002).[34]

1.4. Proposed Strand-Turn-Strand Structural Motif and Implications for Polyglutamine Structure

A recent mutation study of polyQ peptides by Thakur and Wetzel (2002)[26] has provided notable insight into the underlying β-sheet structure of polyQ aggregates. The authors hypothesized that polyQ monomers form antiparallel β-sheets, which in turn

make up the underlying structure of polyQ aggregates. Accordingly, they examined the aggregation kinetics of a synthetic polyQ peptide containing 45 glutamine residues, as well as similar-sized peptides containing four intervals of 7-10 glutamine residues separated by proline-glycine pairs, which are known to promote β-turn formation. Measurement of the concentration of soluble peptide over the course of several hundred hours *in vitro* revealed that the proline-glycine mutants containing intervals of nine (PGQ_9 mutant) or ten (PGQ_{10} mutant) glutamines aggregated at a rate similar to that of the 45-glutamine peptide (see Figure 7). However, the mutant peptides containing intervals of only seven (PGQ_7 mutant) or eight (PGQ_8 mutant) glutamines aggregated much more slowly than the 45-glutamine peptide. In addition, a modified PGQ_9 peptide containing D-proline-L-glycine pairs, which induce β-turns even more strongly than normal L-proline-L-glycine pairs, aggregated more efficiently than did PGQ_9. Finally, antibody binding and electron microscopy assays of the resulting aggregates revealed that they all assumed very similar fibril structures, and heterologous seeding assays revealed that the peptides efficiently seeded each other's aggregation, further supporting the idea of compatible structures. Therefore, the authors determined that the individual chains within polyQ aggregates must adopt a structure composed of alternating β-turn and extended β-strand elements, since the turn-inducing proline-glycine pairs were so readily accommodated upon aggregation, and that uninterrupted intervals of at least nine glutamine residues are required for efficient formation of β-sheet structure.

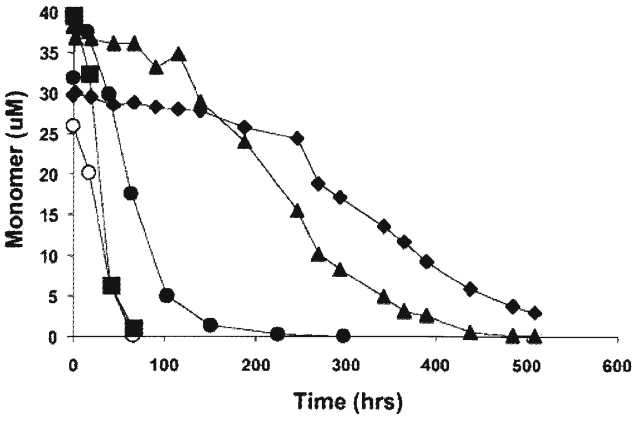


Figure 7: Aggregation kinetics of Q_{45} (○), PGQ_7 (♦), PGQ_8 (▲), PGQ_9 (●), and PGQ_{10} (■).[26] Plots of soluble peptide concentration versus time are shown, and a drop in peptide concentration over time is indicative of aggregation.

Based on this finding, the authors concluded that the polyQ chains formed four-stranded antiparallel β-sheet structures (see Figure 8a), which would form the basis of polyQ aggregates containing ribbon-like fibrils (see Section 1.3.1). Accordingly, they suggested that the mutant peptides folded in such a way that the proline-glycine pairs were incorporated into the β-turns, with which these residue pairs would be structurally compatible, while the four glutamine intervals formed the extended β-strands (see Figure 8b). Indeed, discontinuous molecular dynamics (DMD) simulations of PGQ_9 performed by Khare *et al.* (2005)[41] indicated a preference of this peptide for antiparallel β-sheet structure over β-helix structure, thus supporting Thakur and Wetzel's conclusion about the aggregate structure of the mutant peptides.

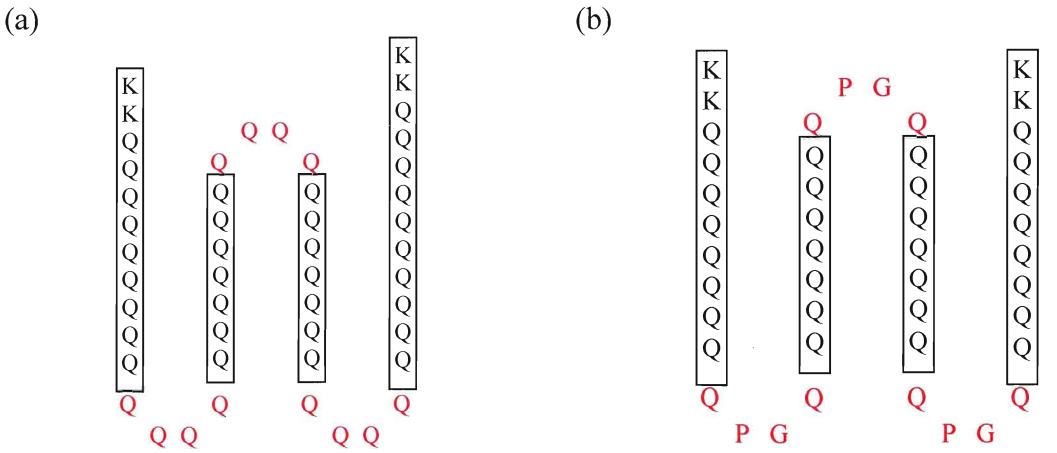


Figure 8: Schematic diagrams of antiparallel β-sheet structures deduced by Thakur and Wetzel (2002) for (a) a 45-glutamine peptide, and (b) the PGQ_9 mutant peptide.[26] The four-residue β-turns are highlighted in red, and the intervening extended β-strands are highlighted in black. All peptides that were examined contained terminal pairs of lysine residues, to ensure monomer solubility in the aqueous experimental solutions.

Additionally, Thakur and Wetzel (2002) examined the aggregation behaviour of a number of mutant variants of the PGQ_9 peptide (see Table 1), in which the central residue of one or two of the four polyglutamine segments was substituted with an additional proline residue. This further analysis was done to confirm whether the four polyglutamine segments of PGQ_9 actually form extended β-strands, based on the logic that proline insertions into the middle of the β-strand-forming elements of PGQ_9 should inhibit its aggregation. Indeed, it was found that such mutations completely abolished aggregation (see Figure 9). The only exception was the PGQ_9 variant containing a single additional proline in the first glutamine interval (*i.e.* $PGQ_9(P^1)$), for which aggregation was slowed rather than completely abolished. Based on these results, the authors confirmed that the four polyglutamine segments of PGQ_9 form extended β-strands, and determined that proline insertions into a polyQ chain effectively inhibit aggregation when

the prolines are placed into regions of the polyQ that are predisposed by existing mutations to form extended β-strand elements of the aggregate structure.[26]

Table 1: PGQ_9-Based Mutant Peptides Examined by Thakur and Wetzel (2002)[26]

Mutant Peptide	Amino Acid Sequence[a]
PGQ_9	K_2-**Q_9**-PG-**Q_9**-PG-**Q_9**-PG-**Q_9**-K_2
$PGQ_9(P^1)$	K_2-**Q_4PQ_4**-PG-**Q_9**-PG-**Q_9**-PG-**Q_9**-K_2
$PGQ_9(P^2)$	K_2-**Q_9**-PG-**Q_4PQ_4**-PG-**Q_9**-PG-**Q_9**-K_2
$PGQ_9(P^{1,2})$	K_2-**Q_4PQ_4**-PG-**Q_4PQ_4**-PG-**Q_9**-PG-**Q_9**-K_2
$PGQ_9(P^{2,4})$	K_2-**Q_9**-PG-**Q_4PQ_4**-PG-**Q_9**-PG-**Q_4PQ_4**-K_2
$PGQ_9(P^{2,3})$	K_2-**Q_9**-PG-**Q_4PQ_4**-PG-**Q_4PQ_4**-PG-**Q_9**-K_2
$PGQ_9(P^{3,4})$	K_2-**Q_9**-PG-**Q_9**-PG-**Q_4PQ_4**-PG-**Q_4PQ_4**-K_2

[a]All peptides that were examined contained terminal pairs of lysine residues, to ensure monomer solubility in the aqueous experimental solutions. The four polyglutamine segments of each peptide are highlighted in bold.

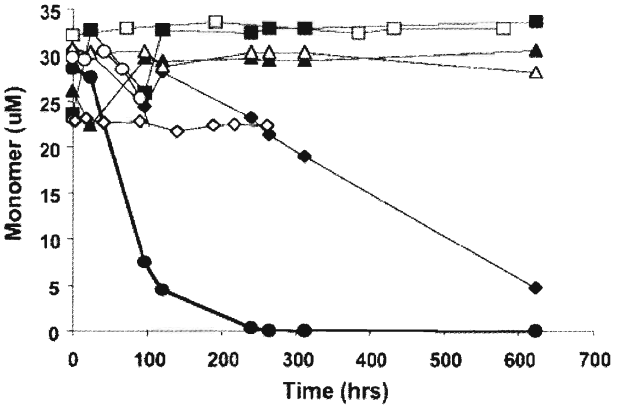


Figure 9: Aggregation kinetics of PGQ_9 (●), $PGQ_9(P^1)$ (♦), $PGQ_9(P^2)$ (◊), $PGQ_9(P^{1,2})$ (■), $PGQ_9(P^{2,4})$ (Δ), $PGQ_9(P^{2,3})$ (▲), and $PGQ_9(P^{3,4})$ (□).[26] Plots of soluble peptide concentration versus time are shown, and a drop in peptide concentration over time is indicative of aggregation.

1.5. Structural Propensity of Polyglutamine: Research Objective

In light of the results of Thakur and Wetzel (2002),[26] and the idea that a single polyQ chain serves as the critical nucleus for aggregation (see Section 1.2), it was likely

that the inhibitory effects of Thakur and Wetzel's proline mutations on aggregation of the variants of PGQ_9 (Figure 9) are due to effects on the structural tendency of the monomer, which disrupt its β-sheet structural propensity. However, Thakur and Wetzel did not make a detailed interpretation regarding the nature of such β-sheet structure disruption. Therefore, in the current study I will be examining, via computer simulations, the monomer structural propensities of a subset of Thakur and Wetzel's mutant peptides, in order to investigate the effect of Thakur and Wetzel's proline mutations.

I hypothesize that the additional proline mutations within contiguous Q_9 stretches of the variants of the PGQ_9 peptide exert localized conformational restraint on the peptide backbone, confining the backbone to an existing tendency for polyproline type II (PPII) and/or turn/bend structure, for which both glutamine and proline residues have demonstrated a propensity.[61,62] Such a localized restraint would in turn tend to confine the peptides to a native global preference for coil structure over β-sheet structure, discouraging the already-unfavourable transition towards the β-sheet structure required for aggregate nucleation/growth, and so would provide a structural explanation for the inhibition of aggregation observed for these peptides. Indeed, a recent computational study by Vitalis *et al.* (2008)[63] proposed that a key factor in polyQ chain association is spontaneous conformational fluctuations within the polyQ monomer, and additional computational studies [64, 65] have also suggested that spontaneous conformational fluctuation is involved in the transition toward β-sheet-like structure.

To investigate the above hypothesis, I will be examining the structural propensities of selected mutant peptides using simulated annealing molecular dynamics (SA-MD). This procedure was chosen in order to bias the molecular structure sampling

towards low-energy structures, as such structures would be most likely to be explored during folding, and thus exert the strongest overall influence on monomer properties. In this technique, an MD simulation is first equilibrated at a high temperature, and then the temperature is slowly reduced to a low final temperature, allowing the simulated polyQ chain to gradually fold into its energetically-optimal structure(s).[66] A number of SA-MD simulations will be performed on the PGQ_9 mutant, which aggregated with similar efficiency to that of non-mutated polyQ, and was proven to have similar aggregate structure (see Section 1.4).[26] In addition, the $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ mutants will be simulated, as sample cases of the PGQ_9 variants with additional proline insertions (see Table 1). $PGQ_9(P^1)$ was chosen as the one peptide with an additional proline in only a terminal polyQ segment of the PGQ_9 sequence, and slowed, but not abolished, aggregation. Meanwhile, $PGQ_9(P^{2,3})$ was chosen as a peptide with additional prolines in only central polyQ segments of the PGQ_9 sequence, and completely abolished aggregation. Finally, the three selected peptides were chosen with the consideration that they all possess the same number of amino acid residues, and the same turn-inducing proline-glycine pairs, thus ensuring consistency among the peptides apart from the additional prolines of $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$.

In all simulations, the carboxy-terminal end of each polyQ peptide will be constrained to a β-turn-β-strand structure, in accordance with the turn-strand structure deduced previously for PGQ_9 (see Section 1.4).[26,41] Such a constraint was chosen to simulate a situation in which β-structure formation has initiated due to interaction with a seed or a growing oligomer/aggregate, thus permitting a sampling of structures that are possible in such a situation. At the same time, however, the chosen constraint was held

away from the sites of the additional glutamine→proline mutations of $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ (see Table 1) – where imposition of β-strand constraints would be problematic – thus allowing these parts of the peptide to form whatever structure types would be energetically-accessible during the simulations. Although this may seem to be a crude representation of polyQ monomer-assembly association, such representation would permit an assessment of structural effects that occur within the monomer, rather than due to interactions between monomers.

Structures generated by the SA-MD simulations will then be subjected to a number of structural analyses, including backbone dihedral calculations, local secondary structure assessment with the Dictionary of Protein Secondary Structure (DSSP),[67] automated histogram filtering (AHF) cluster analysis,[68,69] and assessment of global β-sheet structure formation by counting of "in-register contacts."[64] In addition, the structures will be subjected to energy calculations by molecular mechanics and quantum mechanics methods, in order to assess whether appreciable non-bonded perturbations of peptide structure are present in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ – relative to PGQ_9 – or if any perturbations are mainly local, geometry-based ones as hypothesized. Based on the results, the effect of Thakur and Wetzel's mutations on the structural propensities of the polyQ peptides will be assessed, and the hypothesis of conformational restraint will be addressed. If such alteration to structural propensity is observed, then the results will suggest that the mutations indeed discourage the formation of β-sheet structure *via* effective entrapment in native coil structure, thus resulting in the inhibition of aggregation observed by Thakur and Wetzel. Such information would indicate conformational restraint to be a possible means of inhibiting polyQ aggregation, which

could one day lead to the development of effective rational inhibitors of polyQ aggregation that are based on such a mechanism, thus paving the way for effective future treatments of HD and other polyQ-associated diseases.

2. METHODS

2.1. Simulated Annealing Molecular Dynamics (SA-MD)

2.1.1. Some Simulated Annealing Background

Simulated annealing is a technique for molecular structure optimization that was originally developed by Kirkpatrick *et al.* (1983).[70] Simulated annealing was originally used to find the global-optimum state for simple atomic/molecular systems, but has also been applied to peptides, and to other multidimensional optimization problems such as the travelling salesman calculation.[66,70]

In the technique of SA-MD, an MD simulation is first equilibrated at a high temperature (500-1000 K), which allows the simulated peptide to efficiently explore conformational space by passing over barriers in the peptide folding potential energy landscape. Then, the temperature is slowly reduced to a low final temperature (usually 50 K), during which the simulated peptide gradually folds into an energetically-optimum structure(s). Although in principle, simulated annealing can locate the global minimum with an ideal cooling schedule, in practice it is difficult, if not impossible, to find the global-optimum structure of a peptide due to the complexity of its potential energy function. Still, the technique has proven very useful for efficient sampling of low-energy structures. Therefore, the technique was chosen in order to fulfill the goal of sampling

low-energy structures for the peptides in question, and thus permit examination of the properties of such structures.

2.1.2. Preparation of Initial Polyglutamine Peptide Structures

Structures of Thakur and Wetzel's PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ mutant peptides (see Table 2), each in an extended conformation, were constructed using CHARMm 31.1 molecular modeling software,[71] and saved as protein databank (PDB) coordinate files. The terminal lysine residues of all peptides were replaced with neutral amino-terminal acetyl and carboxy-terminal N-methylamine capping groups, in order to exclude any structural effects that may otherwise arise from electrostatic interactions between charged ends. Accompanying protein structure parameter files for use in the simulations were also constructed using CHARMm 31.1, implementing parameters from the CHARMm22 force field.[72]

Table 2: Polyglutamine Peptides Under Examination

Polyglutamine Peptide	Amino Acid Sequence[a]
PGQ_9	Ace-**Q_9**-PG-**Q_9**-PG-**Q_9**-PG-**Q_9**-Nme
$PGQ_9(P^1)$	Ace-**Q_4PQ_4**-PG-**Q_9**-PG-**Q_9**-PG-**Q_9**-Nme
$PGQ_9(P^{2,3})$	Ace-**Q_9**-PG-**Q_4PQ_4**-PG-**Q_4PQ_4**-PG-**Q_9**-Nme

[a]Ace = amino-terminal acetyl capping group; Nme = carboxy-terminal N-methylamine capping group. The four polyglutamine segments of each peptide are highlighted in bold.

Initial local energy minimizations of the extended structures were then performed using NAMD 2.6 molecular dynamics simulation software,[73] in order to remove atomic clashes within the structures. Minimization was performed using the conjugate gradient algorithm, and was executed *in vacuo* for 2000 steps without constraints, using the same CHARMm22 energy parameters implemented in the subsequent SA-MD simulations.

All minimizations were executed on a 2.2 GHz quad-core Opteron computer system at SHARCNET,[74] using four CPUs per run. Before continuing the structure preparation process, the presence of β-strand structure after minimization was confirmed by examination with VMD 1.8.4 molecular viewing software.[75]

Next, in order to permit application of the desired constraints, a β-turn was generated between the third and fourth polyQ segments of each peptide. To achieve this, the intervening proline-glycine pair was set to backbone dihedral angles necessary for a Type II β-turn. Indeed, proline-glycine pairs are known to be readily-compatible with this type of turn,[76] and the PGQ_9 β-sheet structures reported in the literature demonstrated strand 1-strand 2 and strand 3-strand 4 turns resembling Type II β-turn geometry.[41,54] Using the Torsion Monitor tool of ArgusLab 4.0.1 molecular modeling software,[77] the proline and glycine dihedral angles were set to ($\varphi = -60°$, $\psi = 120°$) and ($\varphi = 90°$, $\psi = 0°$), respectively, in accordance with dihedrals previously reported for Type II β-turns.[76] Then, conjugate gradient energy minimization was repeated for 200 steps to remove any remaining atomic clashes (with a final energy gradient of <0.1 kcal/mol·Å), and the resulting structure was examined with VMD 1.8.4 to confirm the presence of a β-turn.

2.1.3. SA-MD Simulations

As with the energy minimizations, all SA-MD simulations were performed using NAMD 2.6 molecular dynamics simulation software,[73] and executed on a 2.2 GHz quad-core Opteron computer system at SHARCNET.[74] Eight CPUs were used for each simulation, as this number of processors had been determined to yield the fastest run completion times, according to a series of test runs (data not shown). In order to account for the previously-reported native preference of polyQ for non-β-sheet structure (see

Section 1.2), a non-β-structure-biased force field – the CHARMm22 force field[72] – was implemented throughout the simulations. In addition, solvent effects were represented by attenuating the electrostatic component of the force field using a fixed dielectric constant. This method of solvent effect representation was chosen to ensure good computational efficiency, and the dielectric constant was set to a value of 10.0, in accordance with the solvent-excluded regions of a polar uncharged peptide's structure.[78]

A timestep of 1.0 fs was implemented throughout, with non-bonded interactions evaluated every 2.0 fs using a RESPA multiple timestep integrator algorithm.[79] All covalent bonds to hydrogen atoms were constrained using the SHAKE algorithm.[80] The non-bonded cutoff distance was set to a value of 500.0 Å – a greatly exaggerated value compared to the fully-extended length of the peptides – in order to perform simulations that effectively have no cutoff for the non-bonded interactions, thus eliminating cutoff-related bias.

In all simulations, the carboxy-terminal end of each polyQ peptide was constrained to a β-turn-β-strand structure comprising amino acid residues 31-42 – *i.e.* the carboxy-terminal fourth polyQ segment and preceding four-residue turn (see Section 1.5; Figure 10). Indeed, test simulations of PGQ_9 in which the fourth polyQ segment was constrained to β-strand structure without a constrained β-turn (data not shown) indicated the partial formation of β-hairpin structure between the third and fourth polyQ segments, thus validating the inclusion of a turn as part of the constraints. Harmonic atom position constraints were applied to the α-carbon atoms of the constrained residues, the carbon and nitrogen atoms of the intervening peptide bonds, and the peptide bond nitrogen atom of residue 31, based on the positions of the atoms after completion of the initial structure

preparation (Section 2.1.2). Harmonic constraints were chosen in order to maintain the desired structure while still permitting some structural flexibility during the simulations. All constraints were applied using NAMD's intrinsic constraint function, with a relatively large force constant of 7.50 kcal/mol·Å^2 to ensure proper maintenance of the constraints at the high temperatures of the SA-MD simulations.

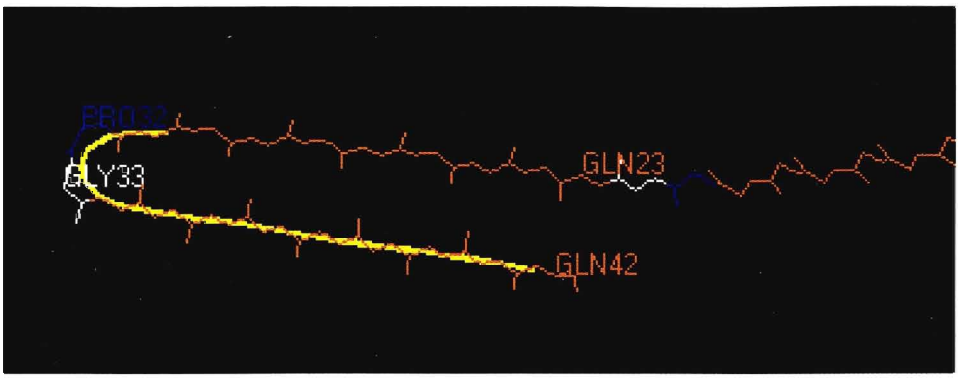


Figure 10: Stick model illustration of the β-turn-β-strand constraints applied to the carboxy-terminal end of the simulated peptides. The PGQ_9 peptide is illustrated as an example (image generated using Swiss-PdbViewer 3.7[81]), and for clarity, amino acid side chains have been omitted. The polyglutamine segments are highlighted in orange, and the proline-glycine amino acid residue pairs are highlighted in blue and white (blue=proline, white=glycine). Residues at the ends of the third and fourth polyglutamine segments are labeled according to their numbers in the amino acid sequence, and the constrained amino acid residues are highlighted by a yellow ribbon.

For each peptide, a series of 25 independent, high-temperature MD simulations were performed, starting from the prepared β-hairpin-containing initial structure. In these simulations, the peptides were heated linearly from 0 K to 900 K over 20 ps, and then simulated at 900 K (NVE ensemble) for another 11 ns. Throughout the simulations, the temperature was maintained at 900 K using the velocity rescaling protocol. The structure of the peptide was saved every 500 ps, thus ensuring that the saved structures would not be appreciably correlated with one another. However, all structures prior to 1.5 ns were ignored in order to permit 1.5 ns of initial equilibration, in accordance with the required

1.5 ns equilibration period observed previously for similar-sized peptides.[82] In total, a set of 500 high-temperature structures (20 structures from each high-temperature simulation) were generated for subsequent use in SA cooling simulations.

Next, each of the 500 high-temperature structures was subjected to a SA cooling simulation. After initialization of the intramolecular forces and energies, the structure was cooled exponentially from 900 K to a final temperature of 50 K over 10 ns. This was accomplished using an iterative cooling protocol, in which the temperature was reduced every 100 fs to a large fraction ("f") of the preceding temperature, in accordance with the exponential cooling protocol of Barakat and Dean (1990):[83]

$$T(t) = T_0 \cdot e^{-At} = T_0 \cdot f^t \qquad (1)$$

where T(t) is the temperature at time t; T_0 is the initial temperature; and A and f are positive constants whose values are pre-calculated to fit the exponential temperature trend to the desired schedule (in this case, $A = 2.89 \times 10^{-4}$ ps^{-1} and $f = 0.9997$). Finally, the peptide was equilibrated at 50 K for 100 ps, and the final structure was saved for subsequent analysis. Throughout the simulations, the temperature was maintained along the desired trend of cooling and final equilibration using the velocity rescaling protocol.

In total, 500 final structures were generated for each of the three peptides. This number of structures was comparable to the numbers of structures generated previously by SA-MD for other peptides,[84] and was therefore likely to represent a sufficiently large sample of structures for analysis.

2.2. Peptide Comparisons: Local Structure Analysis

The local secondary structure tendencies exhibited by the structures generated for each of the three peptides were characterized by a combination of two methods: calculation of peptide backbone dihedrals (φ and ψ), and analysis using the Dictionary of Protein Secondary Structure (DSSP).[67] The backbone dihedrals for each peptide were computed using the Torsion Trace tool of the Quanta 2005 molecular modeling software package.[85] In order to permit efficient analysis, the structures for each peptide were collected into a single trajectory file (DCD-format) prior to input into Quanta. The calculated dihedrals were then plotted as two-dimensional (φ,ψ) dihedral frequency histograms (see Figures 24-27) using Origin 7.5 graphing software.[86] The analysis was performed on the first three polyQ segments, while ignoring the fourth segment to avoid structural bias resulting from the constraints (see Figure 10) applied to this segment during the SA-MD simulations. In particular, the central seven residues of each polyQ segment were examined, since it is these residues that would give rise to the β-strands of the antiparallel β-sheet aggregate structure deduced previously for PGQ_9 (see Section 1.4).[26,41]

The DSSP analysis of the peptides was performed using the Dictionary of Protein Secondary Structure software package.[67] This software identifies secondary structure types adopted by each amino acid residue, based on a combination of backbone hydrogen bond and geometry criteria (see Appendix C for details). In total, seven structure types were examined: 3_{10}-helix; α-helix; π-helix; hydrogen-bonded β-structure; hydrogen-bonded turns; non-hydrogen-bonded bends; and non-hydrogen-bonded loop/irregular structure (which includes PPII structure). Creation of the necessary PDB-formatted input

structure files for each peptide, and compilation of the analysis results, was carried out using computer programs written in Fortran77. The number of residues per peptide structure that exhibited each secondary structure type was then plotted for each of the three peptides as frequency histograms, using Origin 7.5 (see Figure 28). The analysis was performed on the central seven residues of each of the first three polyQ segments, while ignoring the fourth segment, for reasons similar to those of the dihedral analysis (see above).

2.3. Peptide Comparisons: Potential Energy Analysis

In addition to structural analyses of the three peptides, potential energies of the peptide structures were also examined, using a combination of CHARMm molecular mechanics energy calculations, and quantum mechanics energy calculations based on fragment molecular orbital (FMO) methodology.[87] This was done to determine whether the additional glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ exerted appreciable non-bonded perturbations of peptide structure – *i.e.* beyond the inherent loss of hydrogen-bonding functional groups incurred by the glutamine→proline mutations – or if the effects of these mutations are primarily local, geometry-based ones, in accordance with the local structure analysis results.

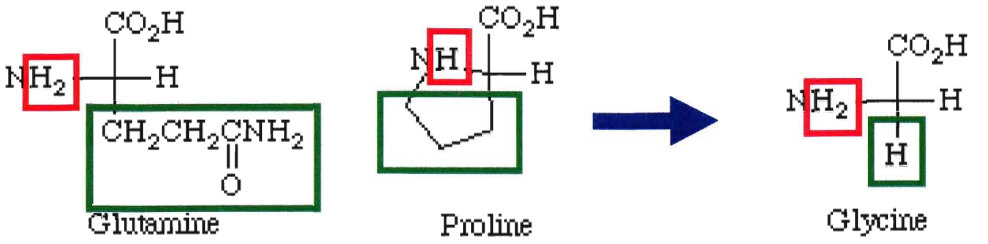
Prior to performing any potential energy calculations, the structures of each peptide were energy-minimized for 100 steps using NAMD 2.6 (with a final energy gradient of <0.1 kcal/mol·Å).[73] This step was necessary because due to the nature of the molecular mechanics potential energy functions (see Appendix B), even small structural perturbations can make a big difference to the calculated potential energies. Thus, the final minimization was necessary in order to remove energetic artifacts from the

structures, even though the minimization would have made negligible structural difference to peptide structures generated at a final simulation temperature as low as 50 K (as was done in the current study). Minimization was performed using the conjugate gradient algorithm, and implemented the same energy calculation parameters used in the SA-MD simulations – including the same harmonic constraints (see Section 2.1.3). All minimizations were executed on a 2.2 GHz quad-core Opteron computer system at SHARCNET,[74] using 8 CPUs per run.

2.3.1. Analysis of Molecular Mechanics Backbone Potential Energies

In order to permit examination and comparison of the peptide backbone potential energies, without the energy contributions of the side chains, polyglycine versions of the energy-minimized peptide structures were constructed. To accomplish this objective, modifications to the atoms of all glutamine and proline residues were made within the respective DCD trajectory files, and input PDB files, using programs written in Fortran77. These modifications included replacement of all glutamine and proline residue side chains with hydrogen atoms – the side chain of glycine – as well as addition of a hydrogen atom to the peptide bond nitrogen of each proline residue – which has one fewer hydrogen atom than the peptide bond nitrogens of other amino acid residues – while retaining the terminal capping groups (see Figure 11). Accompanying protein structure parameter files were then constructed using CHARMm 31.1,[71] once again implementing parameters from the CHARMm22 force field.[72] Finally, the positions of the new hydrogen atoms in the structure were optimized by 100 steps of conjugate gradient energy minimization using NAMD, with fixed-atom constraints applied to the rest of the peptide to prevent structure alteration during the energy minimization.

(a)



(b)

Ace-Q_9-PG-Q_9-PG-Q_9-PG-Q_9-Nme
Ace-Q_4PQ_4-PG-Q_9-PG-Q_9-PG-Q_9-Nme
Ace-Q_9-PG-Q_4PQ_4-PG-Q_4PQ_4-PG-Q_9-Nme

Ace-G_{42}-Nme

Figure 11: (a) Structural modifications made to the glutamine and proline residues of PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ during construction of the corresponding polyglycine structures. Green box = side chains; red box = peptide bond hydrogen atoms. (b) Amino acid sequence transformations achieved by the modifications. Ace = amino-terminal acetyl capping group; Nme = carboxy-terminal N-methylamine capping group.

The CHARMm22 molecular mechanics potential energies of the polyglycine structures for all three peptides were then computed using NAMD. All calculations were performed using the same energy calculation parameters implemented in the SA-MD simulations (see Section 2.1.3), and using programs written in Fortran77, the total bonded (geometric) and non-bonded potential energies were computed from the NAMD output as follows:

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} + E_{improper} \qquad (2)$$

$$E_{non\text{-}bonded} = E_{VDW} + E_{electrostatic} \qquad (3)$$

where E_{bond}, E_{angle}, $E_{dihedral}$, $E_{improper}$, E_{VDW} and $E_{electrostatic}$ are the total computed CHARMm22 bond-stretching, bond-angle-bending, dihedral (bond rotation), improper dihedral, van der Waals and electrostatic energies, respectively (see Appendix B for details). The resulting bonded and non-bonded backbone potential energies were then

plotted as frequency histograms using Origin 7.5 graphing software,[86] with the results for all three peptides plotted on a single set of graphs for comparison (Figure 29).

2.3.2. Analysis of Molecular Mechanics Total Potential Energies

In addition to backbone potential energies, CHARMm22 molecular mechanics potential energies including the amino acid side chains were also computed. However, the inherent differences between the peptides, and thus between their molecular mechanics potential energy functions, necessitated the exclusion of functional groups that were not common to all three peptides in order to achieve a meaningful comparison of the peptides' energies. In order to accomplish this, the energy calculations were performed only on functional groups that were common to all three peptides, which were specified during the calculation using NAMD's intrinsic Pair Interaction tool.

Two sets of CHARMm22 molecular mechanics potential energies including side chains were calculated for all three peptides. The first set of energies was calculated across the entire peptide structure, except for the central residues of the first, second and third polyQ segments – residues 5, 16 and 27 – where the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ occur (see Table 2, Section 2.1.2). The second set of energies was calculated across the entire peptide structure, except for the δ-carbon functional groups and peptide bond hydrogen atoms of residues 5, 16 and 27, which are the specific sites of the functional group substitutions associated with the glutamine→proline mutations (Figure 12). All calculations were performed using the same energy calculation parameters implemented in the SA-MD simulations, and using programs written in Fortran77, the total bonded (geometric) and non-bonded potential energies were computed from the NAMD output (Equations 2 and 3). The resulting

bonded and non-bonded potential energies were then plotted as frequency histograms using Origin 7.5 graphing software, with the results for all three peptides plotted on a single set of graphs for comparison (Figures 30 and 31).

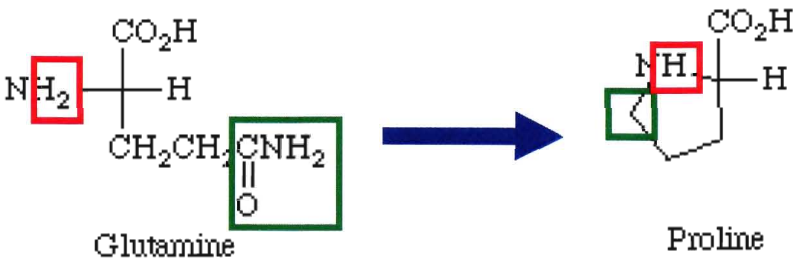


Figure 12: Functional group substitutions associated with the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$. Green box = δ-carbon functional groups; red box = peptide bond hydrogen atoms.

2.3.3. Analysis of Quantum Mechanical FMO Potential Energies

In order to confirm whether the occurrence of potential energy differences between the peptides, as determined by molecular mechanics, is consistent with what quantum mechanics would indicate, quantum-mechanical potential energy calculations were performed on the three peptides. Such examination of energies by an alternative type of methodology was deemed necessary due to the possibility of non-corresponding energy terms for the structures of the three peptides – a possibility that is inherently associated with CHARMm22 molecular mechanics potential energy comparisons between different peptides. The quantum-mechanical potential energy calculations were performed using a fragment molecular orbital (FMO) methodology developed specially for examination of proteins and peptides.[87] In this method, a peptide/protein is divided into amino acid monomers ("fragments"; see Figure 13), and quantum-mechanical energy calculations are performed to obtain the energies of all amino acid monomers, and the

energies of interaction between all pairs of monomers. Once these energies are obtained, the total local conformational (FMO geometric) and long-range non-bonded potential energies can then be computed for the peptide/protein structure in question (as described below in Equations 5 and 6).

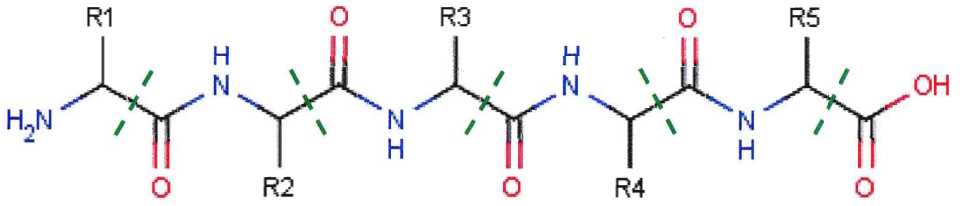


Figure 13: Peptide fragmentation scheme employed by FMO.[87] All fragmentations between amino acid monomers occur across backbone C_α-C bonds (as indicated by green dotted lines).

Before performing the quantum-mechanical energy calculations, representative structures needed to be selected for analysis from among the structures of each peptide. The complete sets of structures could not be analyzed, due to computational expense issues with the computer system on which the analysis software was run. To select representative structures, automated histogram filtering (AHF) cluster analysis[68,69] was performed on the structures of each peptide.

AHF is a cluster analysis method capable of sorting molecular structures into distinct groups, based on geometric similarity. The method rests on performing a principal component analysis on each set of structures, with the inter-α-carbon distances as variables. Then, frequency histograms of the variable factor scores – *i.e.* values of projections of the variables onto each principal component – are computed and scanned for peaks, each of which corresponds to a cluster of geometrically-related structures (see Figure 14 for example histogram). When a division into clusters occurs, the whole

procedure is then repeated on each of the clusters that are obtained, until they cannot be further sub-divided into smaller clusters. Throughout the process, the number of peaks that are identified is regulated by a parameter called the "peak distinction," in order to filter out random noise in the data, which would otherwise result in identification of an excessive number of clusters (see Figure 14).

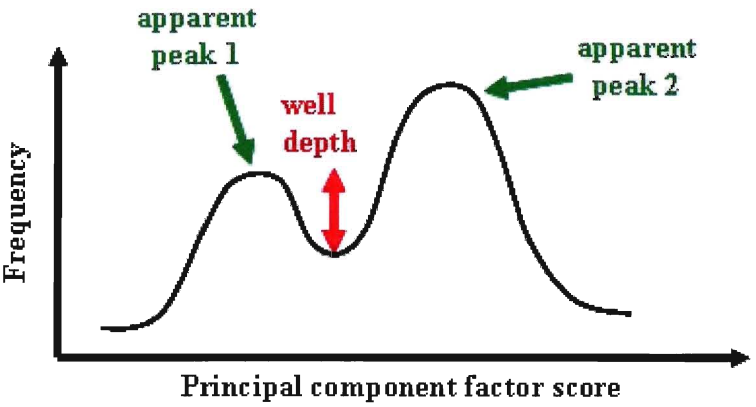


Figure 14: Scheme for identification of peaks in the AHF factor score histograms.[68,69] A simplified histogram is shown, with the factor scores for the corresponding principal component plotted along the horizontal axis, and the number of molecular structures exhibiting each factor score plotted along the vertical axis. The well depth is defined as the distance from the bottom of the well between two apparent peaks, to the top of the shorter peak. If this distance is greater or equal in magnitude to the peak distinction parameter, then the two apparent peaks are accepted as actual peaks (each corresponding to a distinct cluster), rather than random noise in the histogram. Therefore, as the peak distinction value is reduced, the chance of apparent clusters being accepted as distinct clusters increases.

The AHF method has an advantage in that unlike conventional RMSD-based clustering, AHF examines collective patterns of structural variation in a molecular structure sample as a whole, rather than differences for discrete pairs of structures within the sample. As a result, AHF can more effectively assess true patterns of variation among the sample structures than RMSD-based clustering, and can account for variation patterns that RMSD-based clustering may not detect. In addition, by focusing the

analysis only on key atoms of the peptide backbone (*i.e.* α-carbon atoms), AHF is less likely to be affected by variational noise in the data than are RMSD-based methods which examine all atoms or all backbone atoms.

The AHF analysis was carried out using a Fortran-based software package developed by members of the Rothstein research group,[68,69] in which principal component calculations were performed with SHARCNET's AMD ScaLAPACK library.[74,88] The peak distinction parameter was set to a value of 5, which was found to yield maximum numbers of representative structures (see below for selection criteria) for the three peptides, without introducing noise into the results (as evidenced by very large numbers of clusters; data not shown). The data input parameters for AHF clustering were adjusted to the necessary values for the structures of each peptide (*i.e.* 500 structures, each with 42 α-carbons and 42*41/2 inter-α-carbon distances), and the CPU memory allotment parameters were adjusted accordingly. The analysis was run on 4 CPUs of a 2.2 GHz quad-core Opteron computer system at SHARCNET. Once clusters were obtained, the most representative structure from each cluster was then identified, based on the criterion that the structure must exhibit minimum total mean-square deviation (MSD) in inter-α-carbon distances from all other structures in the cluster:[68,69]

$$MSD_{total\,(i)} = \sum_{j \neq i}^{all\ strucs} \frac{\sum_{k}^{all\ dists} (dist_{j,k} - dist_{i,k})^2}{N_{dists}} = minimum \qquad (4)$$

where dist is the k^{th} inter-α-carbon distance of the i^{th}/j^{th} peptide structure, and N_{dists} is the number of inter-α-carbon distances per structure. In total, eight representative structures

for PGQ_9, 12 for $PGQ_9(P^1)$, and 14 for $PGQ_9(P^{2,3})$ were selected for subsequent FMO analysis (see Appendix A).

Analysis of the representative structures by FMO was carried out using the ABINIT-MP software package.[87,89] Using programs written in Fortran77, each of the structures was compiled into a PDB-formatted structure file for input into ABINIT-MP. During the calculations, the MP2 methodology was employed in order to properly account for both electrostatic and electron correlation effects.[90] In addition, the 6-31G basis set was used for the molecular orbitals, since this basis set has proven very effective in previous studies of proteins.[91] All calculations were executed on 16 CPUs of a 2.0 GHz, dual-core Opteron computer system at Kobe University, and using programs written in Fortran77, the total local conformational (FMO geometric; $E_{local\ conf}$) and long-range non-bonded ($E_{long\text{-}range\ NB}$) potential energies for each structure were computed from the ABINIT-MP output as follows:

$$E_{local\ conf} = \sum_i E_i + \sum_i E_{i,i+1} \qquad (5)$$

$$E_{long-range\ NB} = \sum_i \sum_{j \geq i+2} E_{i,j} \qquad (6)$$

where E_i is the energy of amino acid monomer i; $E_{i,i+1}$ is the energy of interaction between monomers i and i+1, which are adjacent to each other in the amino acid sequence; $E_{i,j}$ is the energy of interaction between monomers i and j, which are not adjacent to each other. In order to achieve a meaningful comparison of the peptides' energies, the monomers corresponding to the central amino acid residues of the first, second and third polyQ segments – where the glutamine→proline mutations introduced

in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ occur (see Table 2) – were excluded from the total energy calculations.

2.3.4. Comparison of Quantum Mechanical FMO Potential Energies With Molecular Mechanics Potential Energies

In order to permit an effective comparison between the quantum mechanics and molecular mechanics potential energy results, another set of molecular mechanics potential energies was computed, following energy definitions analogous to those of the FMO method.[87] The results of the quantum mechanics and molecular mechanics calculations were then compared for signs of consistency in the presence/absence of appreciable potential energy differences among the peptides.

For each peptide structure, CHARMm22 molecular mechanics potential energies were computed using NAMD for the entire peptide, for each amino acid monomer, and for each pair of adjacent monomers. Amino acid monomers were delineated based on the same scheme used in the FMO procedure (Figure 13), and during each energy calculation, the portion(s) of the peptide to be analyzed were specified using NAMD's intrinsic Pair Interaction tool. All calculations were performed using the same energy calculation parameters implemented in the SA-MD simulations (see Section 2.1.3). Using programs written in Fortran77, the total CHARMm22 molecular mechanics local conformational ($E_{local\ conf}$) and long-range non-bonded ($E_{long\text{-}range\ NB}$) potential energies of all peptide structures were then computed from the NAMD output as follows:

$$E_{monomer(i)} = E_{bonded(i)} + E_{non\text{-}bonded(i)} \tag{7}$$

$$E_{monomer\ pair(i\ and\ i+1)} = E_{bonded\ monomer\ pair(i\ and\ i+1)} + E_{non\text{-}bonded\ monomer\ pair(i\ and\ i+1)} \tag{8}$$

$$E_{dimer(i,i+1)} = E_{monomer\ pair(i\ and\ i+1)} - E_{monomer(i)} - E_{monomer(i+1)} \tag{9}$$

$$E_{total\ peptide} = E_{bonded\ peptide} + E_{non\text{-}bonded\ peptide} \quad (10)$$

$$E_{local\ conf} = \sum_i E_{monomer(i)} + \sum_i E_{dimer(i,i+1)} \quad (11)$$

$$E_{long\text{-}range\ NB} = E_{total\ peptide} - E_{local\ conf} \quad (12)$$

where $E_{monomer(i)}$ is the total energy of monomer i; $E_{monomer\ pair(i\ and\ i+1)}$ is the total energy of the dimer composed of monomers i and i+1, which are adjacent to each other in the amino acid sequence; $E_{dimer(i,i+1)}$ is the energy of interaction between monomers i and i+1; $E_{total\ peptide}$ is the total energy of the peptide; and the E_{bonded} and $E_{non\text{-}bonded}$ terms are the respective total bonded and non-bonded energies, computed according to Equations 2 and 3.

As with the calculations of the FMO total energies (see Section 2.3.3 above), the amino acid monomers corresponding to the central amino acid residues of the first, second and third polyQ segments were excluded from the calculations. The resulting CHARMm22 molecular mechanics local conformational and long-range non-bonded potential energies were then plotted as frequency histograms using Origin 7.5 graphing software,[86] with the results for all three peptides plotted on a single set of graphs for comparison (Figure 34).

2.3.5. Molecular Mechanics Total Potential Energies: Further Analysis

As a final assessment of peptide potential energies, to further examine the localization of the differences in bonded/local conformational energies that were observed among the three peptides (Sections 3.3.1 and 3.3.2), the CHARMm22 molecular mechanics and FMO analyses of total potential energy (Sections 2.3.2, 2.3.3 and 2.3.4) were repeated. This time, however, the calculations excluded the central amino acid residues of the first, second and third polyQ segments, and the residues

immediately amino-terminal to the central residues – which were also affected by the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, as observed from the peptide backbone dihedrals analysis (see Section 3.2.1).

2.3.6. Statistical Evaluation of Differences in Potential Energy

Due to the small numbers of peptide structures analyzed in each set of quantum-mechanical energy calculations, similarities/differences would have been very difficult to detect graphically as with the full complement of conformations obtained by SA-MD. Therefore, in order to check for significant differences among the three peptides, the respective sets of local conformational and long-range non-bonded energies were compared using a Jonckheere k-sample test (results shown in Table 4).[92] In addition, all sets of CHARMm22 molecular mechanics energies in which a difference was observed graphically were subjected to the k-sample test, to confirm whether or not the difference was significant (results shown in Table 3).

As prescribed by Jonckheere,[92] the normal approximation was applied to the large samples, and the Student's t approximation was applied to the smaller samples (for test details, see Appendix D). In order to ensure maximum confidence in the test results, all tests were performed at the 1% significance (99% confidence) level. The tests were performed in two-tailed format, and all calculations were executed using programs written in Fortran77.

2.4. Peptide Comparisons: Global Structure Analysis

To conclude the comparative analysis of the three peptides, the effects of the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ on global structural tendency were assessed. This was accomplished using a combination of automated histogram filtering (AHF) cluster analysis[68,69] to assess patterns of global structure variation, as well as counting of "in-register contacts"[64] to assess the extent of global β-sheet structure tendency.

2.4.1. Automated Histogram Filtering (AHF) Cluster Analysis

To assess global structure variation for each of the three peptides, automated histogram filtering (AHF) cluster analysis[68,69] was performed on the structures of each peptide. This analysis was carried out with the same Fortran-based software that was used to select peptide structures for FMO analysis (see Section 2.3.3). Multiple AHF calculations were performed on the structures of each peptide, during which the value of the peak distinction parameter was varied incrementally, and the number of clusters identified at each peak distinction value was determined. All other parameters were set to the necessary values (as outlined in Section 2.3.3), and held constant throughout the analysis. The results for each peptide were then plotted as a graph of the number of identified clusters versus the peak distinction value (Figures 37 and 38), and compared for differences in the rate of increase in the number of clusters with decreasing peak distinction. Such an increase in the number of clusters can be expected because as the peak distinction value is reduced, the chance of possible clusters being accepted as distinct clusters increases (see Figure 14).

In this analysis, a greater rate of increase in the number of clusters for a given peptide, compared to the others, would be consistent with an internal fragmentation of the affected peptide's structural ensemble into smaller clusters. In this situation, while the structural ensembles of the non-affected peptides would have stronger tendency toward accounting for a continuous region of global structure space, the structural ensemble of the affected peptide would tend toward accounting for a more fragmented region of global structure space, with the overall diversity of structures remaining unaltered (see Figure 15). This would be consistent with the affected peptide lacking energetically-favourable access to portions of global structure space that are readily accessible to the non-affected peptides. Therefore, the affected peptide would likely have access to a reduced proportion of global structure space – compared to the non-affected peptides – which would in turn be consistent with increased global structural rigidity.

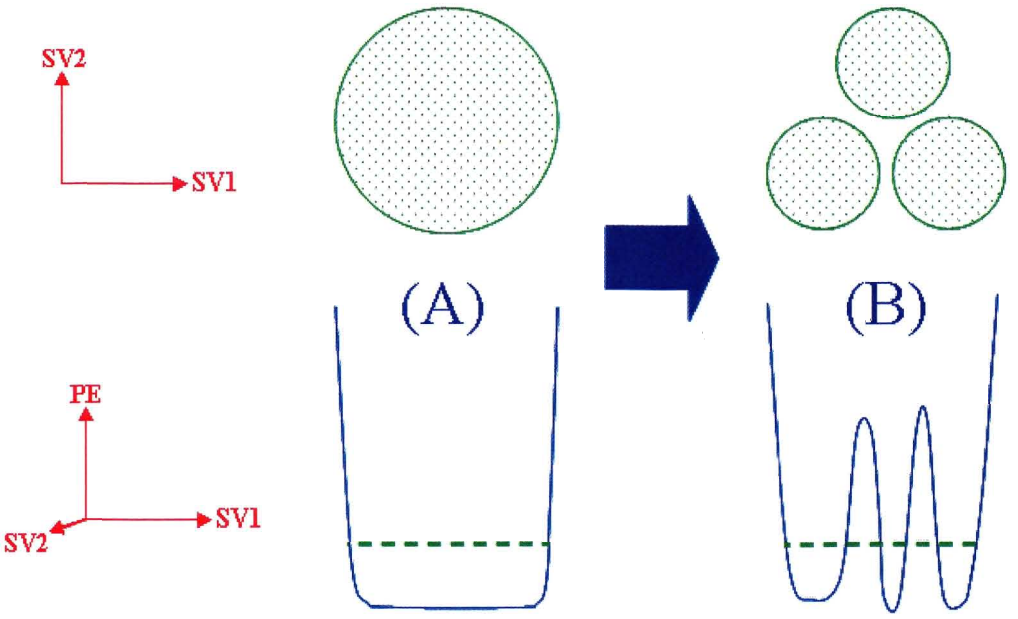


Figure 15: A possible situation consistent with a greater rate of increase in the number of identified clusters with decreasing peak distinction value, but unchanged overall structural diversity. A lower rate of increase corresponds with stronger tendency toward scenario (A), where low-energy structures constitute a continuous range of global structure space. A greater rate of increase, on the other hand, corresponds with stronger tendency toward scenario (B), where low-energy structures constitute a fragmented (*i.e.* discontinuous) range of global structure space. For each scenario, a rough representation of a corresponding peptide potential energy landscape is illustrated (bottom; side view shown), along with the top view of a horizontal cross-section (top) taken through a low-energy region of each landscape (as indicated by a green dotted line). In addition, potential energy landscape variable axes for the side and top views are shown: SV1, SV2 = variables describing peptide structure; PE = peptide potential energy. Although in reality peptide structure would be described by more than two variables, for simplicity only two structure variables are shown.

However, the situation outlined above relies on the assumption that the structural ensembles for all three peptides are of comparable overall structure diversity, according to the inter-α-carbon distance variables implemented in the AHF analysis. Thus, in order to confirm whether the peptide structural ensembles were of comparable diversity, root-mean-square deviations (RMSDs) were computed among all pairs of structures for each

peptide using Fortran77. As with the AHF calculations, all RMSDs were calculated based on inter-α-carbon distances within the structures:

$$RMSD_{i,j} = \sqrt{\frac{\sum_{k}^{all\ dists}(dist_{j,k} - dist_{i,k})^2}{N_{dists}}} \qquad (13)$$

where dist is the k^{th} inter-α-carbon distance of the i^{th}/j^{th} peptide structure, and N_{dists} is the number of inter-α-carbon distances per structure. The calculated RMSDs were then plotted as frequency histograms using Origin 7.5 graphing software (Figure 39),[86] to permit examination of the distributions of RMSD values covered by each peptide. In the event that the RMSD distributions overlap, the peptide structure ensembles in question could be deemed as possessing comparable overall diversity, which would be consistent with the situation outlined above (Figure 15).

Two sets of AHF and RMSD analyses were performed on each of the three peptides. In one set of analyses, the entire peptide was included in the calculations, to permit examination of flexibility across the entire structure of the peptide. In the second set of analyses, however, the carboxy-terminal end of the peptide – which was structurally-constrained during the SA-MD simulations (see Figure 10) – was excluded from the calculations, to permit examination of flexibility across only the portion of the peptide whose structure was free to vary during the simulations.

2.4.2. In-Register Contacts: Assessment of Global β-Sheet Tendency

Finally, the extent of global β-sheet structure tendency was assessed for all three peptides. After attempting a number of methods for assessing global β-sheet structure, including counting of β-structure hydrogen bonds and imposition of a cutoff on DSSP β-

structure residue counts, counting of "in-register" contacts[64] was selected as the quantitative method that could distinguish most effectively between the peptides. Numbers of "in-register" contacts were calculated for the structures of each peptide, in accordance with the antiparallel β-sheet aggregate structure deduced previously for PGQ_9 (see Section 1.4).[26,41] During this analysis, a contact between two amino acid residues was identified if the α-carbon atoms of the two residues were within a distance of 5 Å from each other. A cutoff of 5 Å was chosen because this distance is in line with the inter-β-strand distance of 4.8 Å reported experimentally for polyQ peptides (see Section 1.2), while still allowing a small amount of deviation from perfect β-sheet structure. If identified, a contact between two amino acid residues was accepted as an "in-register" contact if it was one of the potential inter-residue contacts expected within the ideal antiparallel β-sheet structure (see Figure 16).

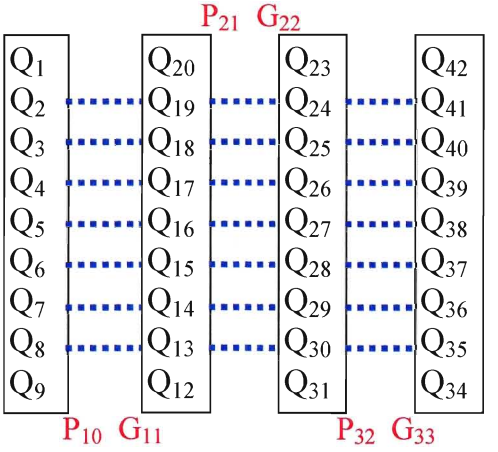


Figure 16: Potential in-register contacts within the ideal antiparallel β-sheet structure of PGQ_9.[26,41] The proline-glycine residue pairs are highlighted in red, the polyQ segments are highlighted in black, and the in-register residue-residue contacts are indicated by blue dotted lines between the two residues involved in each contact.

All calculations were carried out using programs written in Fortran77. The total numbers of in-register contacts per peptide structure were then plotted for each of the three peptides as frequency histograms (Figure 40), using Origin 7.5.[86]

2.5. Control Studies of Non-Constrained Peptide Structures and Antiparallel β-Sheet Structures: Evaluation of Constrained SA-MD Protocol Success

In order to assess the effectiveness of the SA-MD protocol with applied structural constraints, two sets of control analyses were performed on PGQ_9. The PGQ_9 peptide was chosen for these analyses because out of the three peptides examined in the current study, the experimental aggregation behaviour of PGQ_9 most closely resembled that of non-mutated polyQ,[26] and so this peptide would presumably form β-sheet structure the most readily.

2.5.1. Examination of Non-Constrained Peptide Structures

The first of the two sets of control analyses was performed on a series of 500 structures for PGQ_9, for which no structural constraints were applied to the peptide. Structures of PGQ_9 were generated using the same SA-MD protocol as was used previously (Section 2.1), except that no harmonic constraints were applied during the simulations. The resulting structures were then subjected to a DSSP-based local secondary structure analysis (as outlined in Section 2.2), and an analysis of in-register contacts (following the protocol outlined in Section 2.4.2), to assess the extent of β-sheet structure tendency and how it compared to that of the constrained SA-MD structures.

In addition, potential energies of the non-constrained SA-MD structures of PGQ_9 were computed using CHARMm22 molecular mechanics energy calculations, and

subsequently compared to the energies of the constrained SA-MD structures. As with analysis of the constrained SA-MD structures (see Section 2.3), the non-constrained structures were first energy-minimized for 100 steps using NAMD 2.6 (with a final energy gradient of <0.1 kcal/mol·Å).[73] Minimization was performed using the conjugate gradient algorithm, and implemented the same energy calculation parameters used in the other SA-MD simulations – except for the harmonic constraints, which were excluded this time. All minimizations were executed on a 2.2 GHz quad-core Opteron computer system at SHARCNET,[74] using 8 CPUs per run.

After minimization, CHARMm22 molecular mechanics potential energies were first computed across the entire peptide structure, using NAMD. Calculations were performed using the same energy calculation parameters implemented during minimization, and using Fortran77, the total bonded (geometric) and non-bonded energies were computed from the NAMD output according to Equations 2 and 3 (see Section 2.3.1).

In addition, two additional sets of CHARMm22 molecular mechanics potential energies were computed. The first of these sets of energies was calculated across the entire peptide structure, except for the residues responsible for forming the three four-residue turns of the PGQ_9 antiparallel β-sheet structure (*i.e.* residues 9-12, 20-23 and 31-34; see Section 1.4).[26] The final set of energies was calculated across the peptide structure, excluding the residues responsible for forming the three four-residue turns of the β-sheet structure, as well as the side-chain atoms of the remaining amino acid residues – *i.e.* focusing only on the backbone atoms of the β-strand-forming segments of

PGQ_9. In both cases, the peptide functional groups to be analyzed were specified during the calculations using NAMD's intrinsic Pair Interaction tool.

2.5.2. Antiparallel β-Sheet Structures: Structure Generation

The second of the two sets of control analyses was performed on a number of four-stranded antiparallel β-sheet structures generated for PGQ_9. To generate these structures, an energy-minimized extended structure of PGQ_9 was first prepared, following the same protocol outlined in Section 2.1.2. Then, three β-hairpin turns were generated within the structure, producing an initial structure for subsequent energy minimization. In order to generate the turns between the first and second polyQ segments of PGQ_9, and between the third and fourth polyQ segments, the respective proline-glycine residue pairs were each set to backbone dihedral angles necessary for a Type II β-turn. Indeed, the PGQ_9 β-sheet structures reported in the literature demonstrated strand 1-strand 2 and strand 3-strand 4 turns resembling Type II β-turn geometry.[41,54] Using the Torsion Monitor tool of ArgusLab 4.0.1 molecular modeling software,[77] the proline and glycine dihedral angles were set to ($\varphi = -60^\circ$, $\psi = 120^\circ$) and ($\varphi = 90^\circ$, $\psi = 0^\circ$), respectively, in accordance with dihedrals previously reported for Type II β-turns.[76]

Generation of the turn between the second and third polyQ segments, however, could not be performed as described above, as the PGQ_9 β-sheet structures reported in the literature clearly demonstrated a strand 2-strand 3 turn with non-standard geometry.[41,54] In order to establish a best estimate of the backbone dihedrals necessary for this turn, the dihedrals of the proline-glycine pair, as well as those of the flanking glutamine residues (which should also participate in turn formation, according to Thakur and Wetzel,

2002[26]), were adjusted by eye using the Torsion tool of Swiss-PdbViewer 3.7[81] until a suitable β-hairpin alignment of the second and third polyQ segments was achieved, with a turn geometry visibly comparable to that reported in the literature.[41,54] From this assessment, it was determined that the following best-estimate dihedrals would be necessary to pre-set the turn: (φ = -46°, ψ = -67°) for glutamine 20; (φ = -80°, ψ = 161°) for proline 21; (φ = 36°, ψ = 113°) for glycine 22; and (φ = -79°, ψ = 155°) for glutamine 23. These dihedrals were then set using the Torsion Monitor tool of ArgusLab 4.0.1[77] (as the Swiss-PdbViewer software was unable to process the structural data for the peptide's terminal capping groups).

Once the three β-hairpin turns had been set, the resulting preliminary β-sheet was then subjected to an initial energy minimization process using CHARMm 31.1 to remove atomic clashes, and complete the generation of the initial β-sheet structure. As with the SA-MD simulations (Section 2.1), the CHARMm22 force field[72] was implemented, and solvent effects were represented by attenuating the electrostatic component of the force field using a fixed dielectric constant of 10.0. In addition, the non-bonded cutoff distance was set to a value of 999.0 Å – a greatly exaggerated value compared to the fully-extended length of the peptide – in order to create a situation where there is effectively no cutoff for the non-bonded interactions. However, energy minimization was now done using CHARMm 31.1 molecular modeling software[71] instead of NAMD 2.6, because the structural constraints that needed to be implemented (see below) were unavailable in NAMD. Minimization was executed on an 800 MHz SGI Tezro computer at Brock University.

The first part of the minimization process required optimizing the alignment of the four β-strands with one another. To do so, one-way harmonic distance constraints were imposed between pairs of peptide bonds expected to form the backbone-backbone hydrogen bonds within the antiparallel β-sheet structure, with particular attention paid to the central seven residues of each polyQ segment – *i.e.* the residues predicted to give rise to the β-strands of PGQ_9.[26] For each expected hydrogen bond, a distance constraint was imposed between the interacting amide hydrogen and amide oxygen (using the RESDistance tool of CHARMm), and was set up such that a potential energy penalty would only be applied if the distance between the interacting atoms exceeded an allowed cutoff value. A large force constant of 10.0 kcal/mol·Å^2, and a generous cutoff of 3.0 Å, were applied such that the constraint would guide and hold the interacting atoms close to one another, while final assembly of the hydrogen bond (which is expected to occur at hydrogen-oxygen distances <3.0 Å[93]) would be carried out solely by the normal van der Waals and electrostatic potential energies of the CHARMm force field. Minimization was then performed for 1000 steps using CHARMm's adopted-basis Newton-Raphson (ABNR) algorithm.

The second part of the minimization process required aligning the glutamine side chains of the four β-strands in order to establish the side chain-side chain hydrogen bonds implicated in the β-sheet polar zipper (see Section 1.2). Amide hydrogen-amide oxygen distance constraints were imposed between side-chain amide groups from neighbouring strands following the protocol described above, with particular attention once again paid to the central seven residues of each polyQ segment. Then, a further 1000 steps of ABNR minimization were performed, resulting in a final energy gradient of <0.1

kcal/mol·Å. Finally, the presence of β-sheet structure after minimization (see Figure 17) was confirmed by examination with VMD 1.8.4 molecular viewing software,[75] prior to continuing with simulation of the structure.

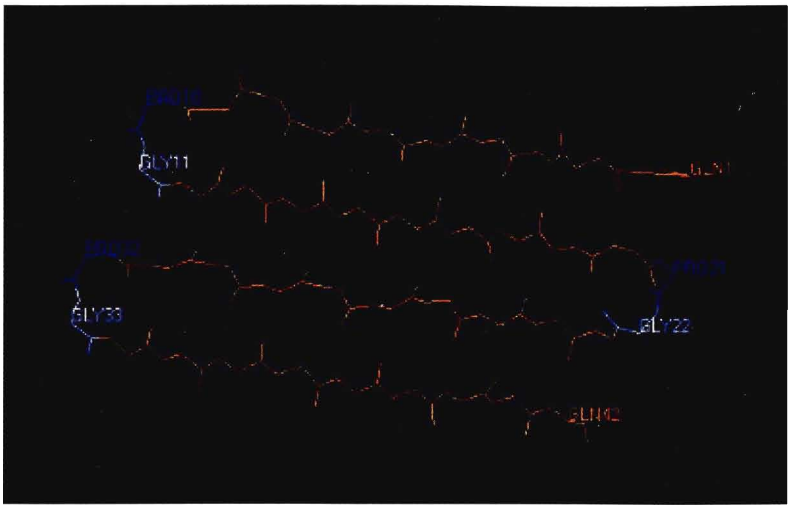


Figure 17: Stick model illustration of the generated antiparallel β-sheet structure of PGQ_9, after completion of initial energy minimization (image generated using Swiss-PdbViewer 3.7[94]). The polyglutamine segments are highlighted in orange, and the proline-glycine amino acid residue pairs are highlighted in blue and white (blue=proline, white=glycine). For clarity, the amino acid side chains have been omitted, and residues at the ends of each polyglutamine segment are labeled according to their numbers in the amino acid sequence.

Upon completion of the minimization process, the constraints were released, and the antiparallel β-sheet structure was then subjected to a low-temperature MD simulation – using the same low temperature implemented during SA-MD (see Section 2.1.3) – in order to generate an ensemble of related structures for examination. As with the minimization, the simulation was performed using CHARMm 31.1,[71] and executed on an 800 MHz SGI Tezro computer at Brock University. The CHARMm22 force field[72] was implemented, with a non-bonded cutoff distance of 999.0 Å, and solvent effects were represented by attenuating the electrostatic component of the force field using a fixed dielectric constant of 10.0. A timestep of 1.0 fs was implemented throughout the

simulation, and all covalent bonds to hydrogen atoms were constrained using the SHAKE algorithm.[80] During the simulation, the minimized β-sheet structure was heated linearly from 0 K to 50 K over 5.0 ps (velocity reassignment protocol), and then simulated at 50 K (NVE ensemble; velocity rescaling protocol) for 50 ps. During constant-temperature simulation, the structure of the peptide was saved every 1.0 ps, thus yielding a set of 50 antiparallel β-sheet structures for subsequent analysis. However, before proceeding with the subsequent analysis, the presence of β-sheet structure within the saved structures was confirmed by examination with VMD 1.8.4.[75]

2.5.3. Antiparallel β-Sheet Structures: Structure Analysis

The generated antiparallel β-sheet structures of PGQ_9 were subjected to a DSSP-based local secondary structure analysis, an analysis of in-register contacts, and analysis of potential energies *via* CHARMm22 molecular mechanics energy calculations. The analyses were done following the same protocols implemented for the non-constrained SA-MD structures of PGQ_9 (see Section 2.5.1). It should be noted that although the antiparallel β-sheet structures generated by CHARMm 31.1 were produced using the same CHARMm22 force field[72] implemented in NAMD,[73] the final energy minimization and calculations were performed using the conjugate gradient algorithm of NAMD to ensure consistency with the final calculated energies of the SA-MD-generated structures.

Finally, the structural and potential energy results obtained for the antiparallel β-sheet structures of PGQ_9 were plotted as frequency histograms, alongside the respective results for both sets of PGQ_9 structures that were generated by SA-MD (Sections 2.1 and 2.5.1). All graphs were generated using Origin 7.5 graphing software,[86] and the results

for the three sets of structures were plotted together to permit comparison of the three structure sets (see Figures 18-22).

3. RESULTS AND DISCUSSION

3.1. Comparison of PGQ_9 Constrained SA-MD Structures with Non-Constrained SA-MD Structure and Antiparallel β-Sheet Structure Controls

3.1.1. Comparison of Molecular Mechanics Potential Energies

Upon examination of the computed CHARMm22 molecular mechanics potential energies for the three sets of PGQ_9 structures (see Figure 18), it was found that the bonded and non-bonded potential energies of the structures generated by SA-MD simulations with constraints (Section 2.1) were crowded considerably closer to those of the antiparallel β-sheet structures (Section 2.5.2) than were the energies of the structures generated by SA-MD simulations without constraints (Section 2.5.1). Therefore, from an energetic standpoint, the constraints appear to have pushed the PGQ_9 peptide closer to antiparallel β-sheet structure, thereby fulfilling the expected effect of enhancing tendency toward β-structure. Still, the energies of the constrained SA-MD structures were generally lower in value than those of the antiparallel β-sheet structures. Therefore, assuming that the generated antiparallel β-sheet structures are reasonably close to those that would be formed upon aggregation, the SA-MD simulations with constraints located PGQ_9 structures lower in energy than β-sheet structure, and thus appear to have fulfilled the purpose of generating low-energy structures of the peptide.

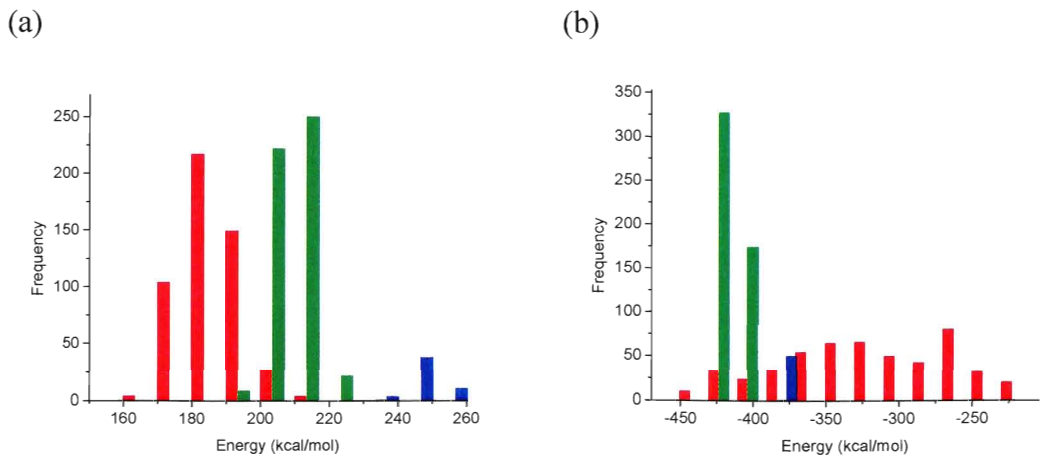


Figure 18: Frequency histograms of CHARMm22 molecular mechanics total potential energies of the three structure sets generated for PGQ_9. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of each structure set) exhibiting each energy value is plotted along the vertical axis. Red = non-constrained SA-MD structures of PGQ_9 (Section 2.5.1); green = constrained SA-MD structures of PGQ_9 (Section 2.1); blue = antiparallel β-sheet structures of PGQ_9 (Section 2.5.2). (a) Bonded energies (Equation 2, Section 2.3.1); (b) non-bonded energies (Equation 3, Section 2.3.1).

More notably, by shifting the potential energy tendencies in stronger favour of antiparallel β-sheet structure, an overall shift of bonded (geometric) potential energies toward higher values and non-bonded potential energies toward lower values was incurred, compared to the SA-MD structures generated without constraints (see Figure 18). Therefore, assuming that the generated antiparallel β-sheet structures are reasonably close to those that would be formed upon aggregation, conformational effects appear likely to be a major limiting factor in the unfavourable transition toward β-sheet structure during aggregation. Indeed, a closer examination of the major components of bonded energy (see Equation 2) for the three structure sets indicated that the vast majority of the bonded energy deviation between the structure sets was due to differences in dihedral energy (see Figure 19). This observation fits the trend of energy difference magnitudes that would be expected from the CHARMm22 potential energy function in the case of

physically-relevant conformational differences between structure sets for a single peptide, and is thus consistent with a conformational difference between the three structure sets.

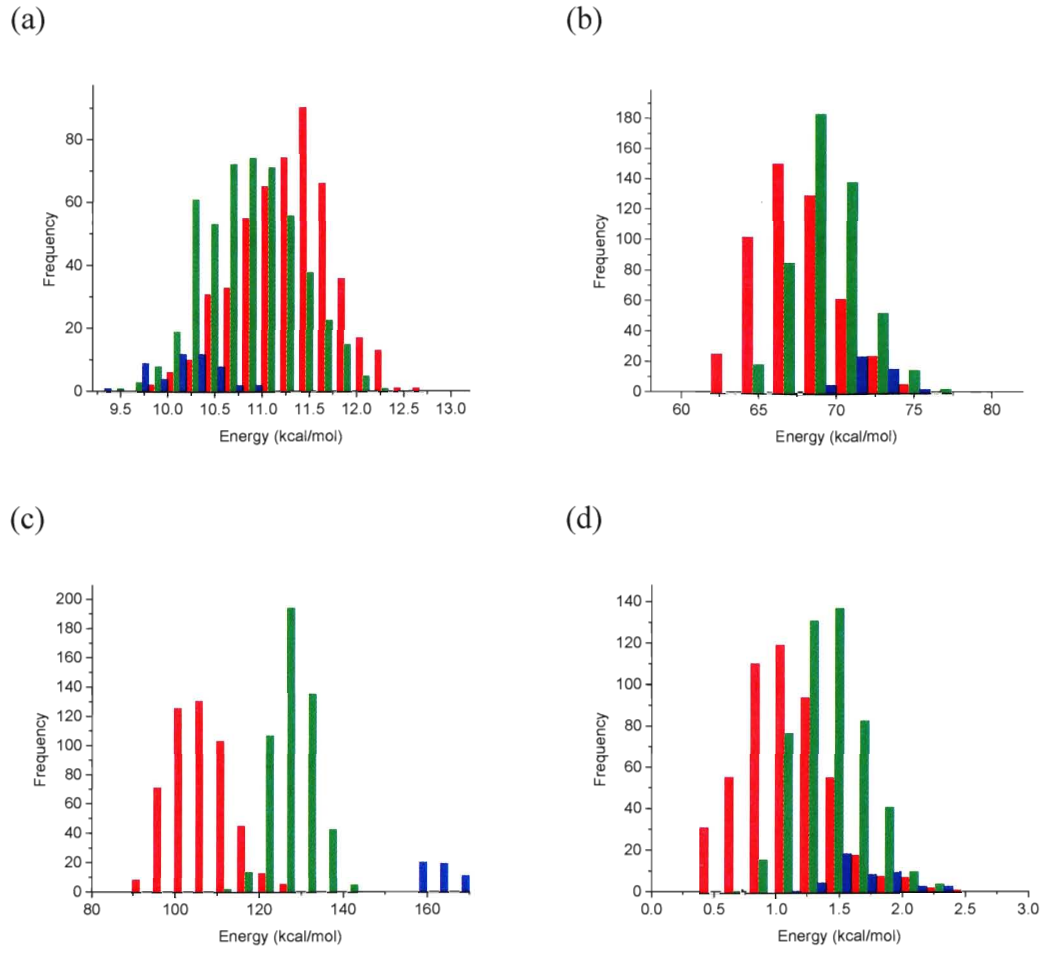


Figure 19: Frequency histograms of CHARMm22 molecular mechanics bonded energy components (see Equation 2) of the three structure sets generated for PGQ_9. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of each structure set) exhibiting each energy value is plotted along the vertical axis. Red = non-constrained SA-MD structures of PGQ_9 (Section 2.5.1); green = constrained SA-MD structures of PGQ_9 (Section 2.1); blue = antiparallel β-sheet structures of PGQ_9 (Section 2.5.2). (a) Bond-stretching energies; (b) bond-angle-bending energies; (c) dihedral energies; (d) improper dihedral energies.

In addition, a calculation of dihedral energies excluding the three turn-forming regions of PGQ_9 (*i.e.* residues 9-12, 20-23 and 31-34) indicated that the observed differences in dihedral energy were due entirely to differences within the four β-strand-

forming segments of PGQ_9 (see Figure 20a), and a further calculation on only the backbone atoms of the β-strand-forming segments indicated that the energy differences were localized to the peptide backbones of these segments (see Figure 20b). Therefore, the large difference in dihedral energy exhibited by the antiparallel β-sheet structures is not due to introduction of the non-standard turn between the two central strands, and so such a turn must indeed be energetically compatible with the peptide. Also, assuming that the generated antiparallel β-sheet structures are reasonably close to those that would be formed upon aggregation, it would appear that β-strand structure is conformationally-disfavoured in the case of polyQ peptides – an observation that is indeed compatible with the experimentally-observed thermodynamic penalty associated with polyQ aggregate nucleation (as described in Section 1.2).[47] Thus, conformational perturbations favouring a native preference for coil structure should further discourage the β-sheet transition as was hypothesized, thereby inhibiting aggregation of the peptide.

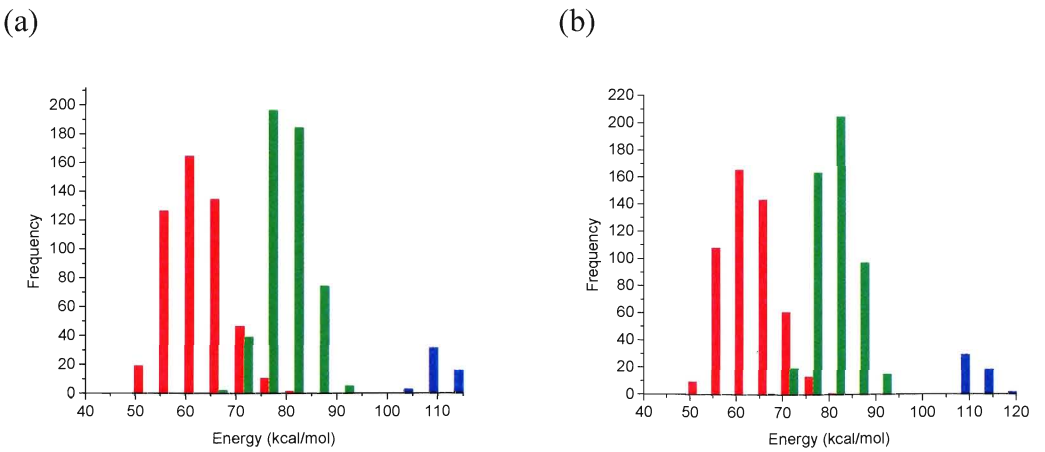


Figure 20: Frequency histograms of CHARMm22 molecular mechanics dihedral potential energies of the three structure sets generated for PGQ_9. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of each structure set) exhibiting each energy value is plotted along the vertical axis. Red = non-constrained SA-MD structures of PGQ_9 (Section 2.5.1); green = constrained SA-MD structures of PGQ_9 (Section 2.1); blue = antiparallel β-sheet structures of PGQ_9 (Section 2.5.2). (a) Dihedral energies excluding the three turn-forming regions of the peptide; (b) dihedral energies excluding the three turn-forming regions of the peptide, and only including the backbone atoms of the remaining peptide segments.

3.1.2. Comparison of β-Sheet Structural Tendency

Upon examination of the in-register contact counts obtained for the three sets of PGQ_9 structures (see Figure 21), it became clear that the presence of the applied structural constraints resulted in a preference for formation of a number of in-register contacts – and thus, an extent of global β-sheet structure tendency – closer to that of the antiparallel β-sheet structures than when no constraints were used. In addition, a DSSP analysis demonstrated a shift in the extent of local β-structure formation towards that of the antiparallel β-sheet structures in the presence of the applied constraints (see Figure 22). Therefore, the potential energy perturbation observed upon application of the constraints (see Section 3.1.1) was indeed accompanied by a visible shift in structural tendency favouring a greater extent of β-sheet structure formation.

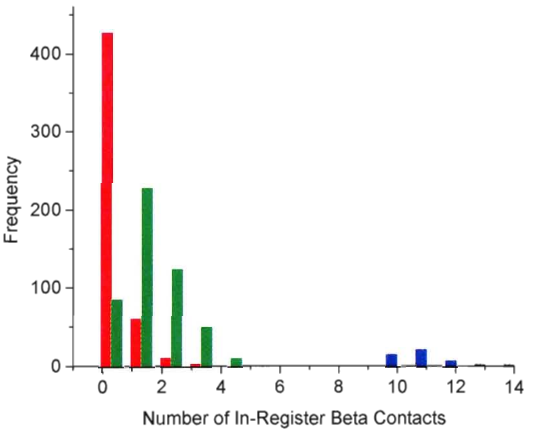


Figure 21: Frequency histogram of the occurrence of in-register β-sheet contacts within the three structure sets generated for PGQ_9. The number of contacts within the peptide is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting that number of contacts is plotted along the vertical axis. The greater the value along the horizontal axis at which the histogram reaches a maximum value along the vertical axis, the stronger the measured tendency is for global β-sheet structure formation. Total in-register contact counts across all three pairs of interacting polyglutamine segments (outlined in Figure 16) are shown: red = non-constrained SA-MD structures of PGQ_9 (Section 2.5.1); green = constrained SA-MD structures of PGQ_9 (Section 2.1); blue = antiparallel β-sheet structures of PGQ_9 (Section 2.5.2).

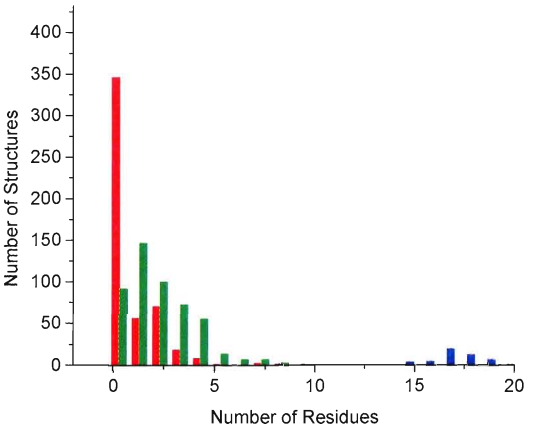


Figure 22: Frequency histogram of the occurrence of hydrogen-bonded β-structure within the first, second and third polyQ segments of PGQ_9, as determined by DSSP analysis. The number of residues in the segments in question that exhibited hydrogen-bonded β-structure is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting that number of residues is plotted along the vertical axis. The greater the value along the horizontal axis at which the histogram reaches a maximum value along the vertical axis, the stronger the measured tendency is for hydrogen-bonded β-structure. Total residue counts across all three segments are shown: red = non-constrained SA-MD structures of PGQ_9 (Section 2.5.1); green = constrained SA-MD structures of PGQ_9 (Section 2.1); blue = antiparallel β-sheet structures of PGQ_9 (Section 2.5.2).

More notably, the non-constrained structures of PGQ_9 exhibited a strong preference for formation of no β-structure at all. Therefore, if the analyzed structures of PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ had been generated without the use of structural constraints, it was likely that very little (if any) tendency for β-structure would have been observed for any of the peptides. As a result, any differences in β-structure tendency between the peptides would have been more difficult to resolve if the constraints had not been used. In addition, it made sense to perform the other analyses of interest on the constrained structures as well, in order to ensure consistency in the structures throughout the study. Thus, the application of structural constraints to the three peptides during SA-MD was necessary for the purpose of the current study.

Still, the extent of β-structure formation observed from the SA-MD simulations with constraints was rather low, and so any evidence of differences in β-structure tendency between the peptides would have inherently been rather weak. However, if β-sheet structure is energetically disfavoured – as has indeed been suggested both experimentally,[47] and computationally with β-sheet-biased force fields[63] – then the low-energy-structure bias of SA-MD could be expected to yield a low occurrence of β-sheet structure (as was indeed observed). Nevertheless, examination of low-energy structures was still likely to yield information that would have been unresolved or overlooked with conventional simulation methods, and so application of the method was continued with the remaining peptides of interest (producing the results described in the following sections).

3.2. Peptide Comparisons: Local Structure Analysis

3.2.1. Peptide Backbone Dihedrals

Examination of the dihedrals exhibited by the three non-constrained polyQ segments of PGQ_9 (Figure 24), and their constituent individual glutamine residues (Figure 25), indicated a strong preference of these residues for PPII-like and α-helix-like dihedrals (see Figure 23 for outline of dihedral angle ranges). There was also a tendency for other dihedrals, including those consistent with β-strand structure. However, such dihedrals occurred much less frequently than the preferred PPII-like and α-helix-like dihedrals, indicating a rather weak inherent tendency for these other dihedrals.

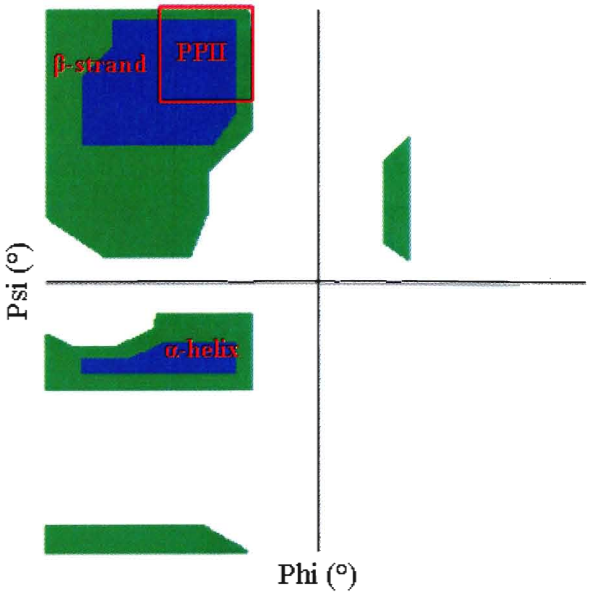


Figure 23: Outline of the major dihedral angle ranges examined during backbone dihedral analysis of the three peptides: β-strand-like dihedrals, PPII-like dihedrals, and α-helix-like dihedrals. The value of the φ-dihedral is plotted along the horizontal axis, and the value of the ψ-dihedral is plotted along the vertical axis. For clarity, the boundaries of the PPII-like dihedrals range ($-105° \leq \varphi \leq -45°$, $120° \leq \psi \leq 180°$)[84,95] are indicated by a red box.

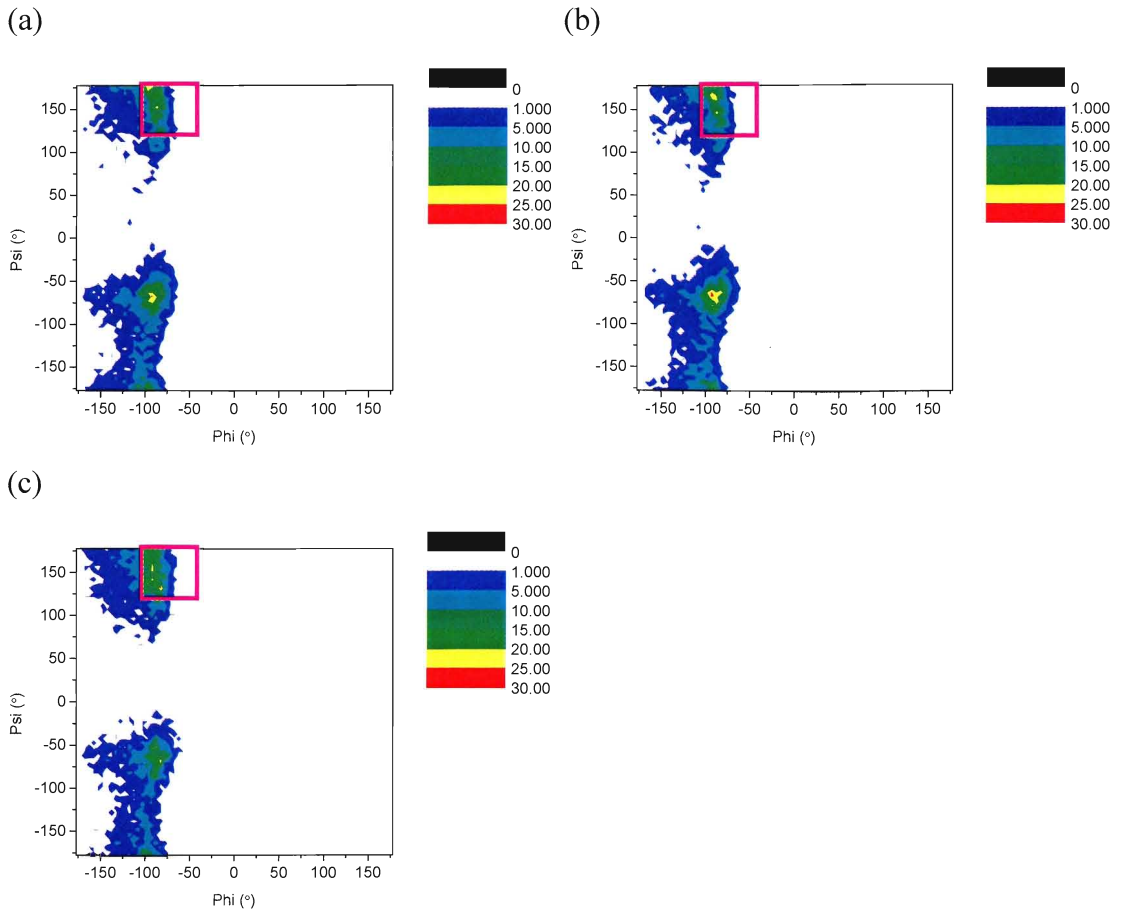


Figure 24: Two-dimensional backbone dihedral frequency histograms for the first, second and third polyQ segments of PGQ_9. The value of the φ-dihedral is plotted along the horizontal axis; the value of the ψ-dihedral is plotted along the vertical axis; and the colours plotted on the histogram indicate the frequency of occurrence of each pair of (φ,ψ) values in the peptide segment in question. For clarity, the boundaries of the PPII-like dihedrals range ($-105° \leq \varphi \leq -45°$, $120° \leq \psi \leq 180°$)[84,95] are indicated by a pink box. (a) First segment; (b) second segment; (c) third segment.

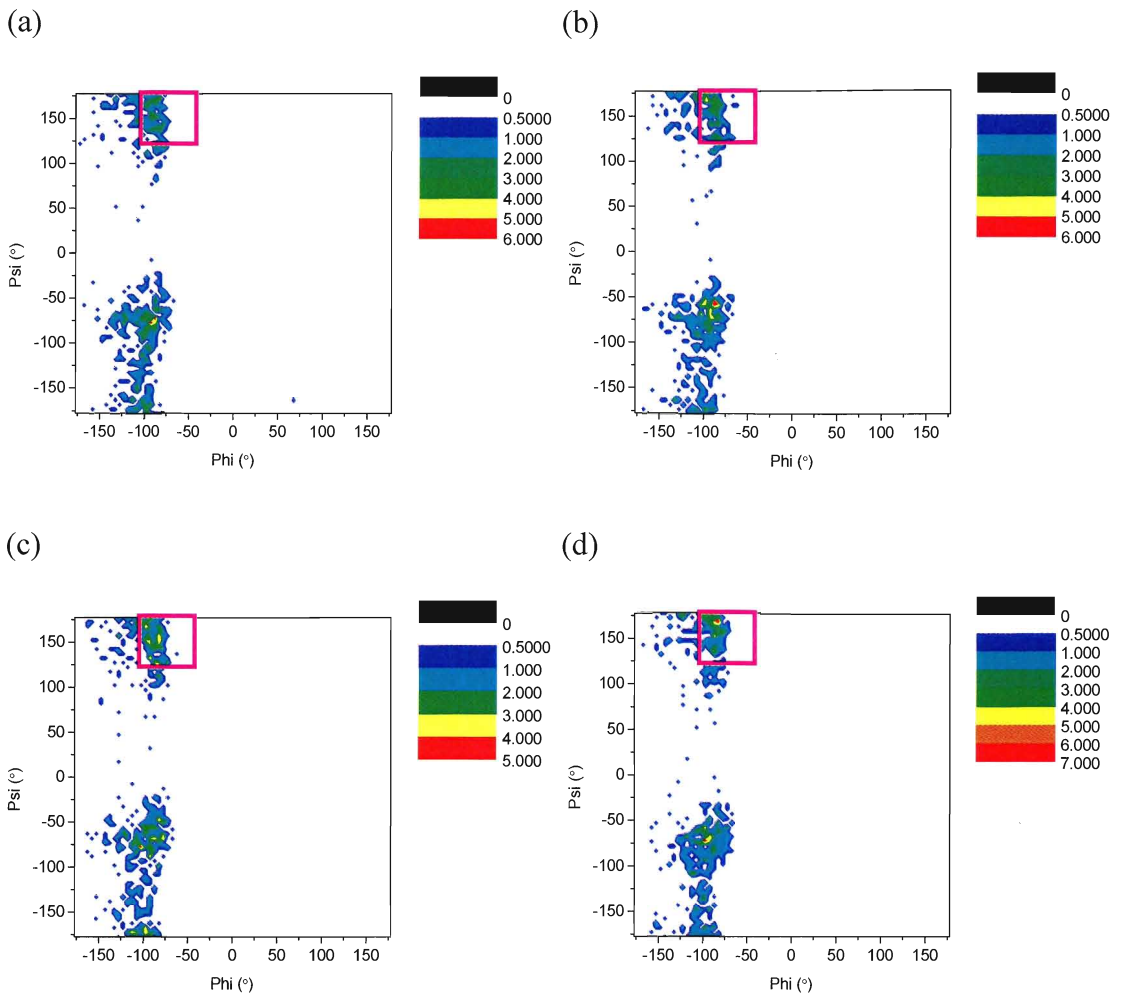


Figure 25: Two-dimensional backbone dihedral frequency histograms for some individual glutamine residues from the first, second and third polyQ segments of PGQ_9. The value of the φ-dihedral is plotted along the horizontal axis; the value of the ψ-dihedral is plotted along the vertical axis; and the colours plotted on the histogram indicate the frequency of occurrence of each pair of (φ,ψ) values in the residue in question. For clarity, the boundaries of the PPII-like dihedrals range ($-105° \leq \varphi \leq -45°$, $120° \leq \psi \leq 180°$)[84,95] are indicated by a pink box. Residues of the first segment: (a) glutamine 3; (b) glutamine 4; (c) glutamine 5 (mutated to proline in $PGQ_9(P^1)$); (d) glutamine 6. Residues of the second segment: (e) glutamine 14; (f) glutamine 15; (g) glutamine 16 (mutated to proline in $PGQ_9(P^{2,3})$); (h) glutamine 17. Residues of the third segment: (i) glutamine 25; (j) glutamine 26; (k) glutamine 27 (mutated to proline in $PGQ_9(P^{2,3})$); (l) glutamine 28.

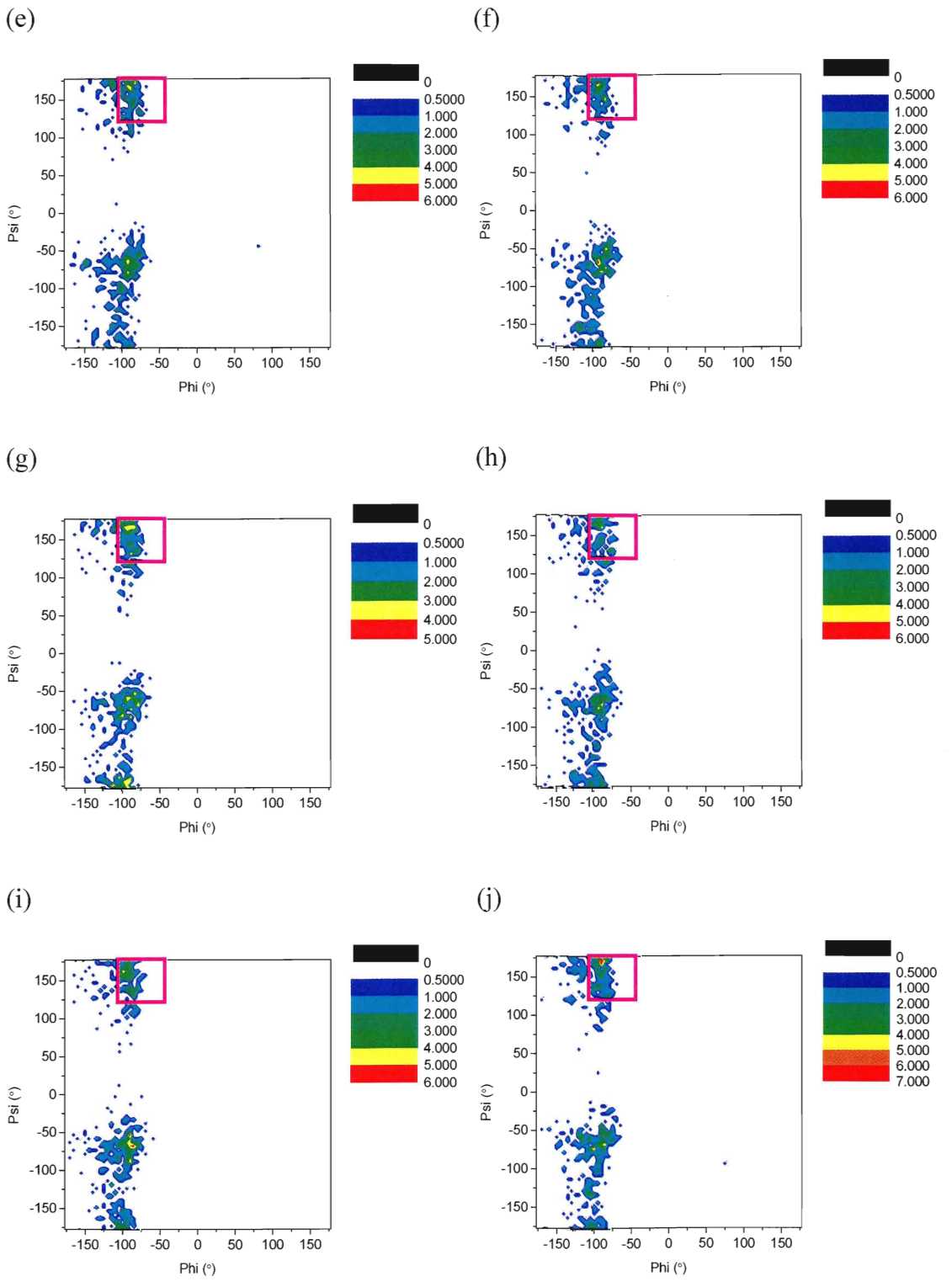


Figure 25 continued.

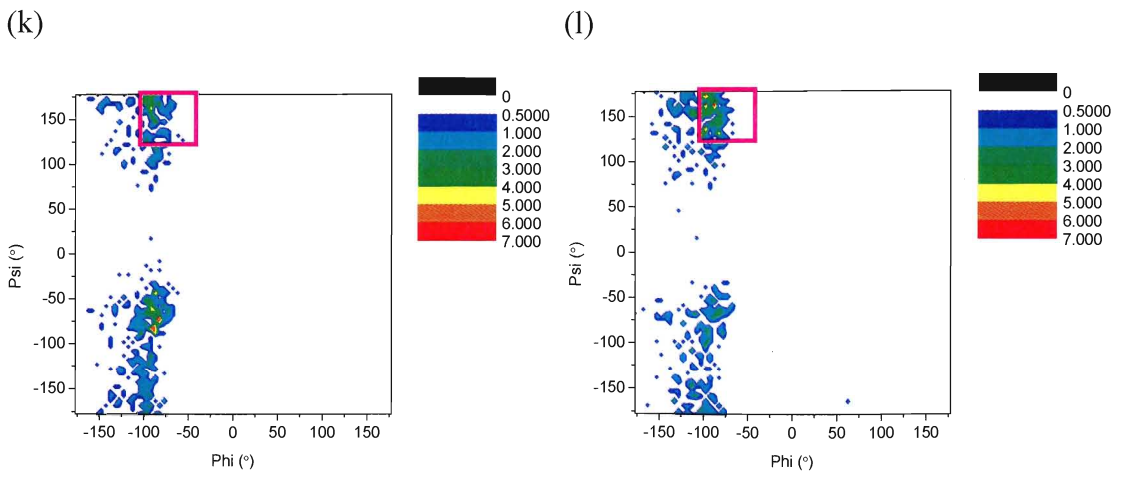


Figure 25 continued.

When the central glutamine residue of the first polyQ segment (in the case of $PGQ_9(P^1)$; residue 5), or the second and third polyQ segments (in the case of $PGQ_9(P^{2,3})$; residues 16 and 27), was mutated to proline, the range of dihedrals exhibited by these residues became considerably more confined (see Figures 26c and 27c,g; compare to Figure 25c,g,k). Now, the dihedrals resided almost entirely within two narrow ranges of values – PPII-like dihedrals, and turn-like dihedrals close to the α-helix range – with little appreciable tendency for β-strand-like or other dihedrals. Also, the glutamine residues immediately amino-terminal to the mutated residues exhibited confinement in the range of values for their ψ-dihedrals, which now resided mostly near values compatible with PPII-like structure, although the range of values for the φ-dihedrals appeared to remain the same as in PGQ_9 (see Figures 26b and 27b,f; compare to Figure 25b,f,j). Therefore, as expected, the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ induce conformational confinement within the peptide backbone. However, the ranges of dihedrals exhibited by all other residues remained the same as in PGQ_9, indicating that the conformational confinement was highly localized. In addition, the confined ranges of

dihedrals exhibited a great degree of similarity with the preferred dihedrals observed in PGQ_9, thus suggesting that the observed structural confinement is based largely on local structural tendencies that are already present prior to introduction of the glutamine→proline mutations.

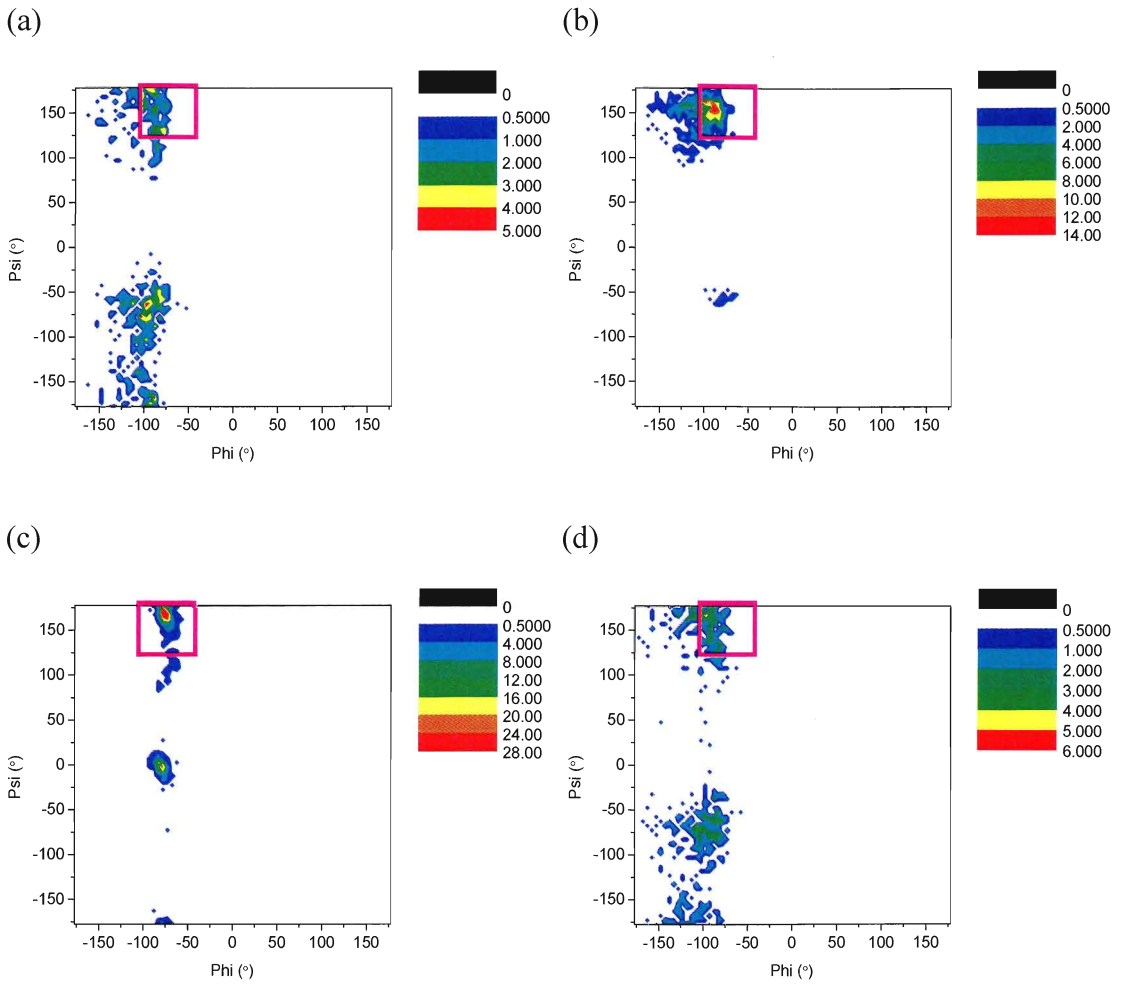


Figure 26: Two-dimensional backbone dihedral frequency histograms for some individual glutamine/proline residues from the first polyQ segment of $PGQ_9(P^1)$. The value of the φ-dihedral is plotted along the horizontal axis; the value of the ψ-dihedral is plotted along the vertical axis; and the colours plotted on the histogram indicate the frequency of occurrence of each pair of (φ,ψ) values in the residue in question. For clarity, the boundaries of the PPII-like dihedrals range ($-105° \leq \varphi \leq -45°$, $120° \leq \psi \leq 180°$)[84,95] are indicated by a pink box. (a) Glutamine 3; (b) glutamine 4; (c) proline 5 (mutated from glutamine to proline); (d) glutamine 6. All other residues exhibit trends similar to those in PGQ_9.

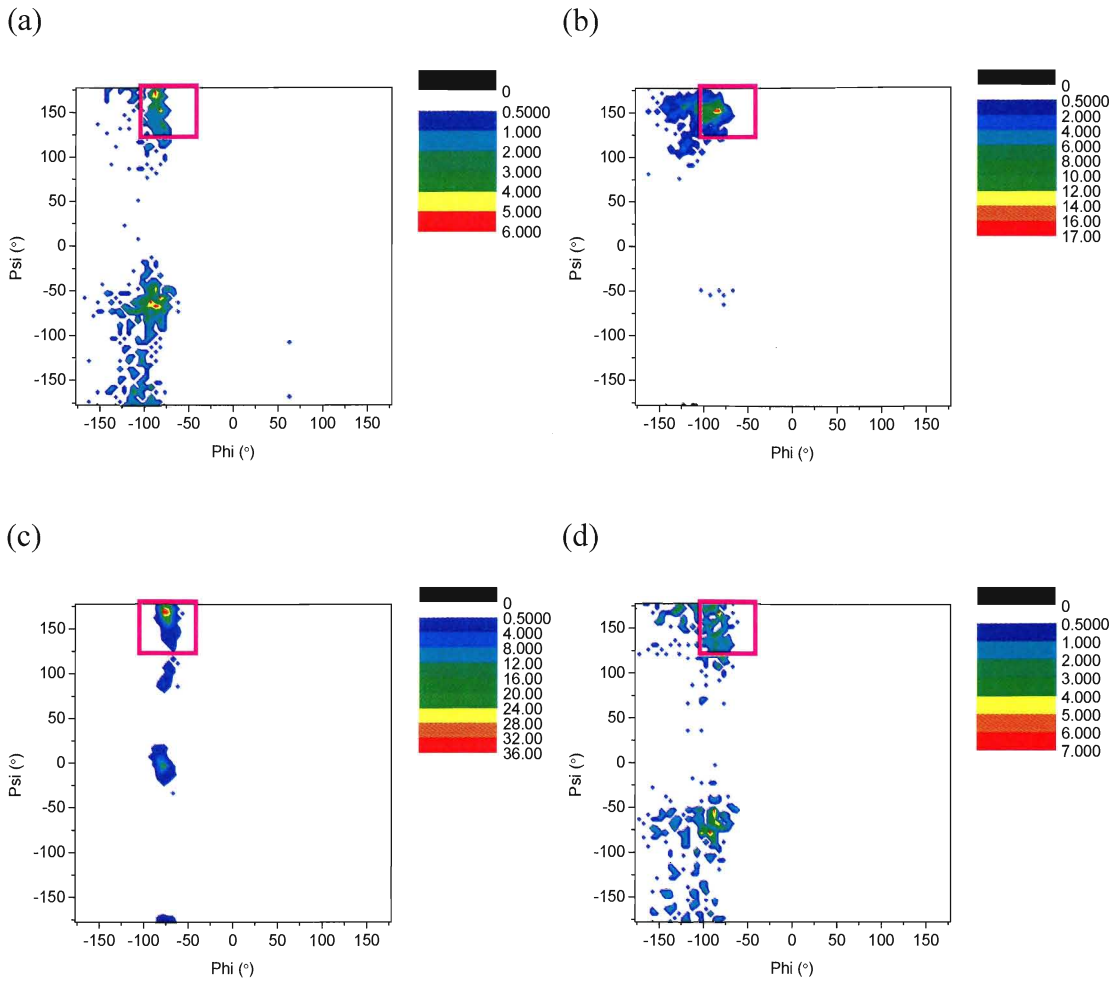


Figure 27: Two-dimensional backbone dihedral frequency histograms for some individual glutamine/proline residues from the second and third polyQ segments of $PGQ_9(P^{2,3})$. The value of the φ-dihedral is plotted along the horizontal axis; the value of the ψ-dihedral is plotted along the vertical axis; and the colours plotted on the histogram indicate the frequency of occurrence of each pair of (φ,ψ) values in the residue in question. For clarity, the boundaries of the PPII-like dihedrals range ($-105° \leq \varphi \leq -45°$, $120° \leq \psi \leq 180°$)[84,95] are indicated by a pink box. Residues of the second segment: (a) glutamine 14; (b) glutamine 15; (c) proline 16 (mutated from glutamine to proline); (d) glutamine 17. Residues of the third segment: (e) glutamine 25; (f) glutamine 26; (g) proline 27 (mutated from glutamine to proline); (h) glutamine 28. All other residues exhibit trends similar to those in PGQ_9.

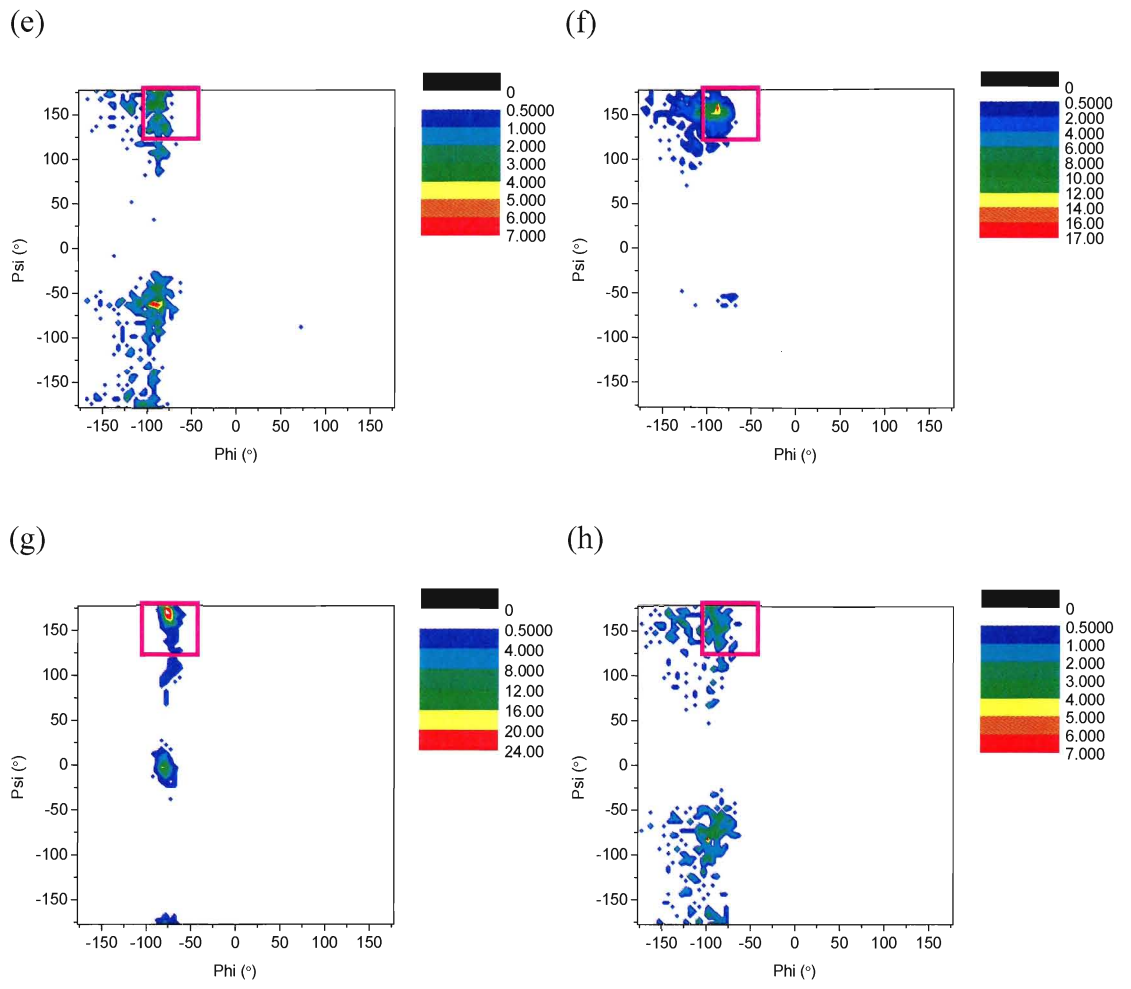


Figure 27 continued.

3.2.2. Dictionary of Protein Secondary Structure (DSSP) Analysis

As with the backbone dihedral analysis, very similar local secondary structure preferences were observed from the DSSP analysis of all three peptides, thus further suggesting that the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ act upon peptide structure based largely on existing local structural tendencies. The observed secondary structure tendencies (see Figure 28) included a strong preference for the non-hydrogen-bonded secondary structure types – bends and loop/irregular structure – over other structure types. In addition, there was a weaker, but appreciable tendency

for hydrogen-bonded β-structure, and very little tendency for hydrogen-bonded turns. Finally, there was no visible tendency for 3_{10}-helix, α-helix or π-helix structure, indicating that the strong observed tendency for α-helix-like dihedrals was not due to α-helix-like structures.

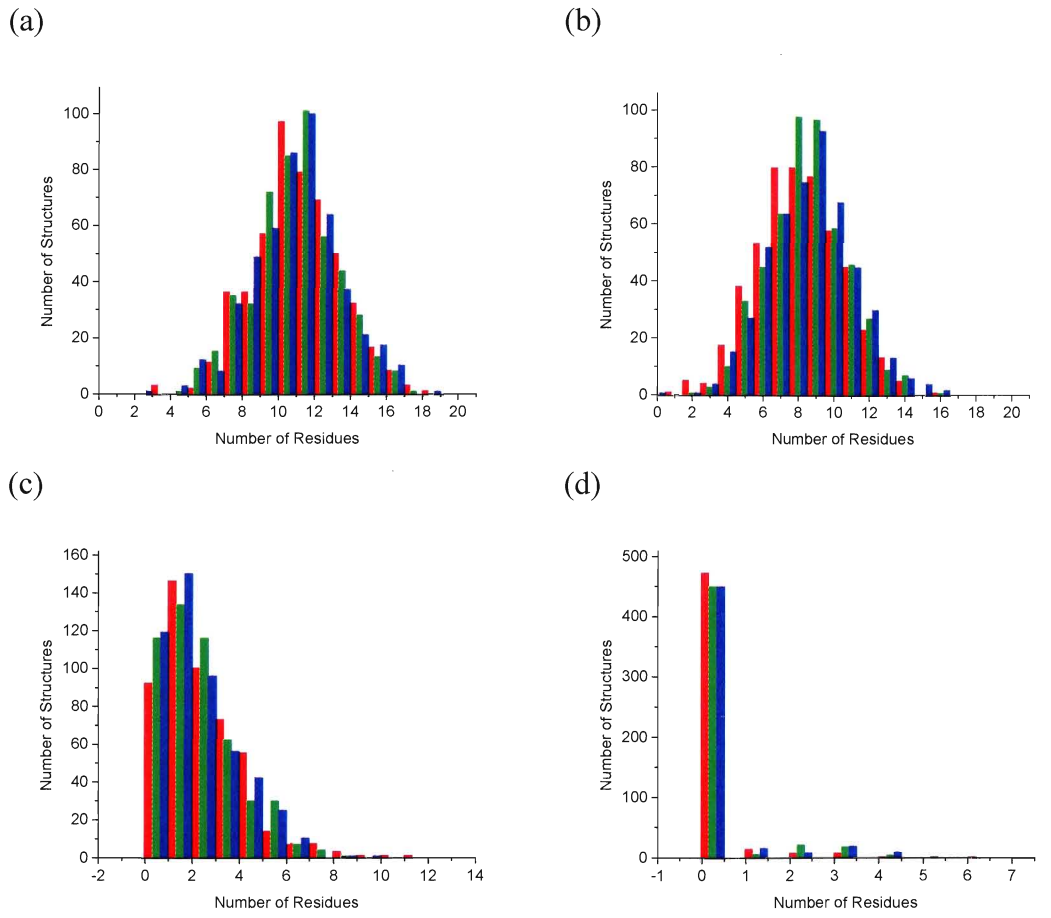


Figure 28: Frequency histograms of the occurrence of seven types of secondary structure within the first, second and third polyQ segments of PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, as determined by DSSP analysis. The number of residues in the segments in question that exhibited each secondary structure type is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting that number of residues is plotted along the vertical axis. The greater the value along the horizontal axis at which the histogram reaches a maximum value along the vertical axis, the stronger the measured tendency is for the respective secondary structure type. Total residue counts across all three segments are shown: (a) non-hydrogen-bonded bends; (b) non-hydrogen-bonded loop/irregular structure (including PPII); (c) hydrogen-bonded β-structure; (d) hydrogen-bonded turns. No tendency for 3_{10}-helix, α-helix or π-helix was observed (data not shown). Colour key: red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$.

3.2.3. Summary of Local Structure Results

Very similar local structure preferences were visible for all three peptides from the backbone dihedral and DSSP analyses. First of all, the strong loop/irregular structure tendency observed from the DSSP analysis, and the observed tendency for PPII-like dihedrals, both indicate a prominent occurrence of PPII-like local structure. In addition, the strong bend (and weak hydrogen-bonded turn) tendency observed from the DSSP analysis, and the observed tendency for α-helix-like dihedrals, are consistent with a prominent occurrence of bend/turn local structure – certain types of which possess dihedrals within or near the α-helix range.[84] Finally, the dihedral and DSSP analyses both indicated a weak occurrence of local β-structure. Therefore, all three peptides appear to possess a strong preference for PPII and bend/turn local structures, with a weaker tendency for β-structure.

This local structure tendency is significant for two reasons. First of all, such tendency is compatible with a global preference for coil structure, and a less favourable tendency for β-structure. Thus, the results are compatible with the normal structural behaviour of polyQ peptides as reported from experiment (see Section 1.2). Secondly, PPII and bend/turn structure types are strongly favoured by proline, and are thus compatible with the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$.

These results, coupled with the localized conformational confinement observed from the backbone dihedral analysis, suggest that as hypothesized, the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ influence peptide structure by means of localized conformational restraint. This restraint confines the

peptide backbone to existing local tendencies for aggregation-incompetent PPII and/or turn/bend structure, which would in turn tend to increase the global preference for coil structure over β-sheet structure.

3.3. Peptide Comparisons: Potential Energy Analysis

3.3.1. Analysis of Molecular Mechanics Potential Energies

Upon examination of the computed CHARMm22 molecular mechanics bonded and non-bonded potential energies for the three peptides, it was found that the non-bonded energy distributions for all three peptides consistently overlapped. This was true in the case of the peptide backbone potential energies (excluding side chains; Figure 29b), the potential energies excluding the three mutated residues (Figure 30b), and even for the potential energies excluding as little as only the functional groups of the mutated residues that are altered by mutation (see Section 2.3.2 for description; Figure 31b). Thus, it appears that any differences in non-bonded energy are limited strictly to the hydrogen-bonding functional groups that are lost as a direct result of the glutamine→proline mutations of $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ – *i.e.* the δ-carbon amide groups and peptide bond hydrogen atoms of the mutated glutamine residues (Figure 12). Therefore, the conformational samples could not be discriminated from one another on the basis of non-bonded interactions, and so non-bonded perturbations of peptide structure do not appear to be an appreciable effect of the glutamine→proline mutations.

(a) (b)

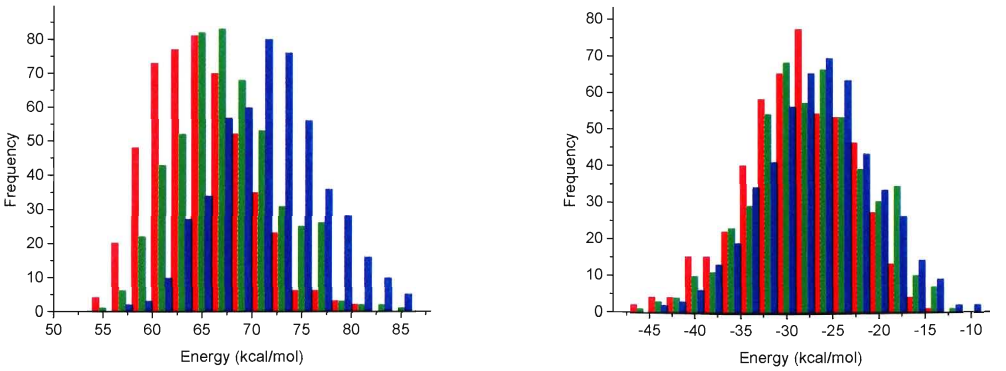


Figure 29: Frequency histograms of CHARMm22 molecular mechanics peptide backbone potential energies for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$. These energies were computed from polyglycine versions of the structures for each peptide, in order to exclude energy contributions from the amino acid side chains. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting each energy value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$. (a) Bonded energies (Equation 2, Section 2.3.1); (b) non-bonded energies (Equation 3, Section 2.3.1).

(a) (b)

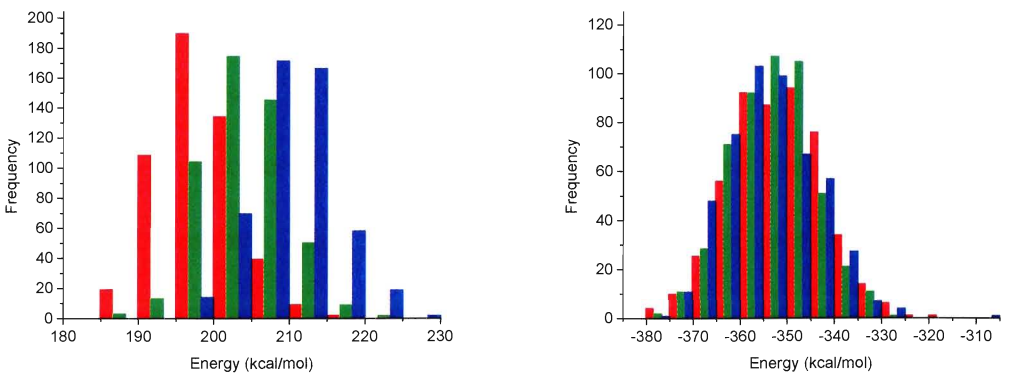


Figure 30: Frequency histograms of CHARMm22 molecular mechanics potential energies for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, excluding the central residues of the first, second and third polyglutamine segments. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting each energy value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$. (a) Bonded energies (Equation 2, Section 2.3.1); (b) non-bonded energies (Equation 3, Section 2.3.1).

(a) (b)

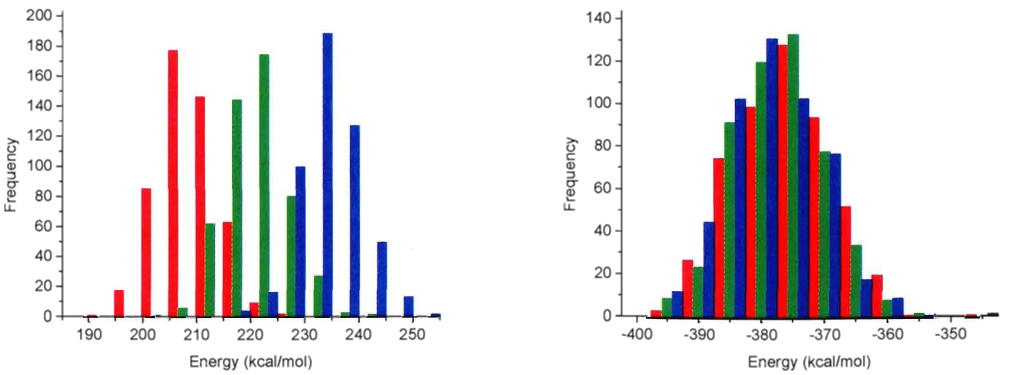


Figure 31: Frequency histograms of CHARMm22 molecular mechanics potential energies for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, excluding the δ-carbon functional groups and peptide bond hydrogen atoms (see Section 2.3.2) of the central residues of the first, second and third polyglutamine segments. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting each energy value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$. (a) Bonded energies (Equation 2, Section 2.3.1); (b) non-bonded energies (Equation 3, Section 2.3.1).

The bonded energy distributions for all three peptides, on the other hand, consistently deviated from one another, with a clear increase in energy from PGQ_9 to $PGQ_9(P^1)$, to $PGQ_9(P^{2,3})$ (see Figures 29a, 30a and 31a). Indeed, a subsequent Jonckheere k-sample test (Section 2.3.6) clearly confirmed the presence of significant energy differences (see Table 3). The observed ordering of energies correlated with the number of glutamine→proline mutations introduced into the polyQ segments of the peptides – none for PGQ_9, one mutation for $PGQ_9(P^1)$, and two mutations for $PGQ_9(P^{2,3})$ (see Table 2) – and was thus not surprising. Since the bonded energies consistently deviated when the non-bonded energies did not, it is clear that bonded energy differences between the peptides are more prominent than are any non-bonded energy differences. In addition, examination of the major components of bonded energy (see Equation 2) revealed that although the first three components all exhibited deviations in value

between the peptides, dihedral energy was the major contributor to the observed bonded energy differences, with a smaller contribution from bond-angle-bending energy and only minimal contribution from bond-stretching energy (see Figure 32, for example). This observation fits the trend of energy difference magnitudes that would be expected from the CHARMm22 potential energy function in the case of physically-relevant conformational differences between peptides, and is thus consistent with a conformational difference between the structural ensembles for the three peptides.

Table 3: Results of Jonckheere k-Sample Test for Molecular Mechanics Potential Energy Data

Potential Energy Data Set	$S_{calculated}$	$Z_{calculated}$[a]	$Z_{critical}$[a]	**Significant Difference Among the Three Peptides?**[b]
Backbone Bonded (Figure 29a)	332142	18.19	2.58	Yes
Total Bonded, Without Mutated Residues (Figure 30a)	505064	27.65	2.58	Yes
Total Bonded, Without Mutated Functional Groups (Figure 31a)	692510	37.92	2.58	Yes
Total Local Conformational, Without Mutated Residues (Figure 34a)	247742	13.56	2.58	Yes

[a]Since the data sets for each peptide were so large, the normal approximation was utilized for each test (for test details, see Appendix D).[92]

[b]A significant difference exists between the three samples if $-Z_{critical} \leq Z_{calculated} \leq Z_{critical}$.

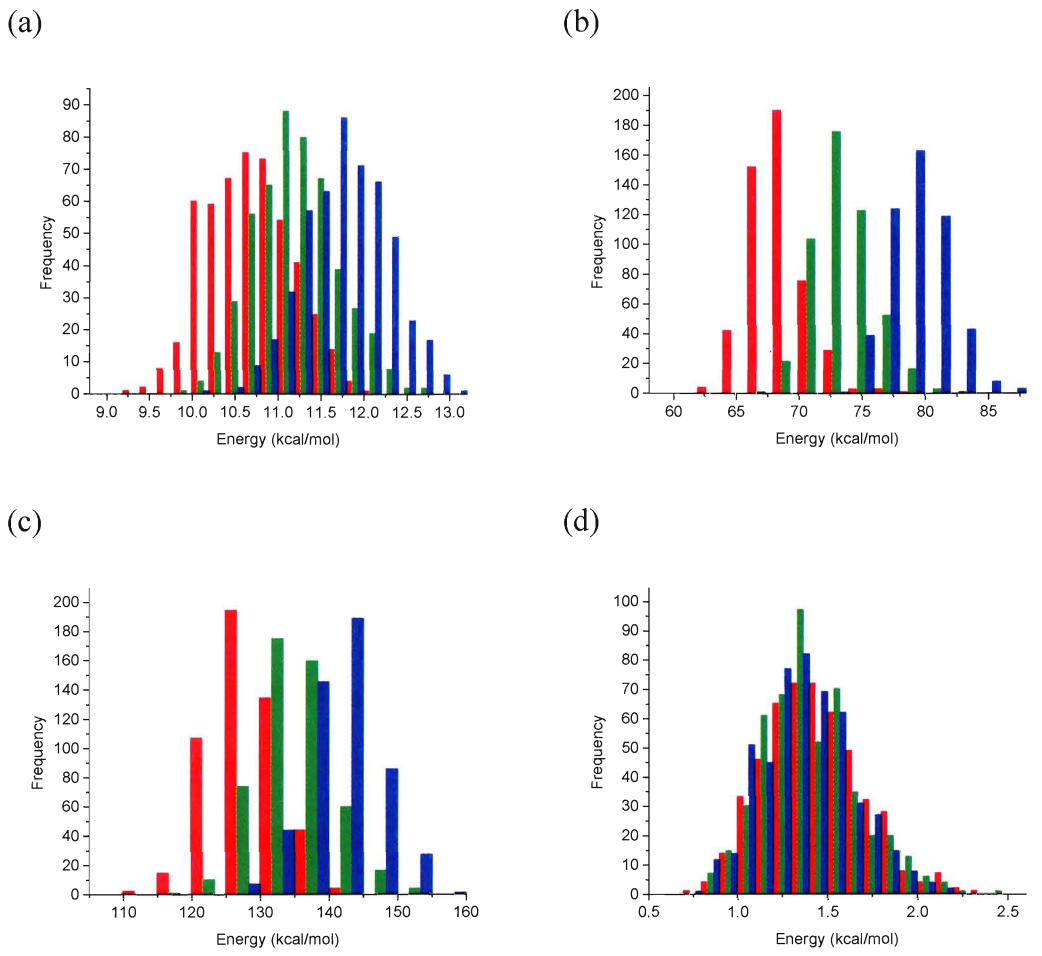


Figure 32: Frequency histograms of CHARMm22 molecular mechanics bonded energy components (see Equation 2) for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, excluding the δ-carbon functional groups and peptide bond hydrogen atoms (see Section 2.3.2) of the central residues of the first, second and third polyglutamine segments. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting each energy value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$. (a) Bond-stretching energies; (b) bond-angle-bending energies; (c) dihedral energies; (d) improper dihedral energies. Here, the following deviations in energy value are visible between the PGQ_9 and $PGQ_9(P^{2,3})$ distributions: bond-stretching energy = ~1 kcal/mol; bond-angle-bending energy = ~10 kcal/mol; dihedral energy = ~15 kcal/mol; improper dihedral energy = no visible deviation.

These results suggest that in accordance with the hypothesis, the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ perturb peptide structure primarily by means of local, geometry-based effects, with little appreciable contribution from non-bonded effects.

3.3.2. Analysis of Quantum Mechanical FMO Potential Energies, and Comparison With Molecular Mechanics

Initial examination of the quantum-mechanical potential energies computed based on the representative structures of each peptide (Section 2.3.3) seemed to suggest a trend in energy values from $PGQ_9 \rightarrow PGQ_9(P^1) \rightarrow PGQ_9(P^{2,3})$ (as seen in Figure 33). However, due to the small numbers of peptide structures examined in the quantum-mechanical energy calculations (see Section 2.3.3 for details), a Jonckheere k-sample test (Section 2.3.6) was performed on the calculated energies of the three peptides to check for statistically-significant differences among them. Upon doing so it was found that the local conformational energies (Equation 5) of the three peptides differed significantly in value, while the long-range non-bonded energies (Equation 6) of the peptides were not significantly different after all (see Table 4).

Table 4: Results of Jonckheere k-Sample Test for Quantum-Mechanical FMO Potential Energy Data

Potential Energy Data Set	$S_{calculated}$	$t_{calculated}$[a]	$t_{critical}$[a]	Significant Difference Among the Three Peptides?[b]
Total Local Conformational, Without Mutated Residues (Figure 33a)	212	3.66	2.65	Yes
Total Long-Range Non-Bonded, without Mutated Residues (Figure 33b)	130	2.12	2.65	No
Total Local Conformational, Without Mutated and Neighbouring Residues (Section 3.3.3)	102	1.64	2.65	No

[a]Since the data sets for each peptide were small in size (8 structures for PGQ_9, 12 structures for $PGQ_9(P^1)$, and 14 structures for $PGQ_9(P^{2,3})$), the Student's t approximation was utilized for each test (for test details, see Appendix D).[92]

[b]A significant difference exists between the three samples if $-t_{critical} \leq t_{calculated} \leq t_{critical}$.

(a) (b)

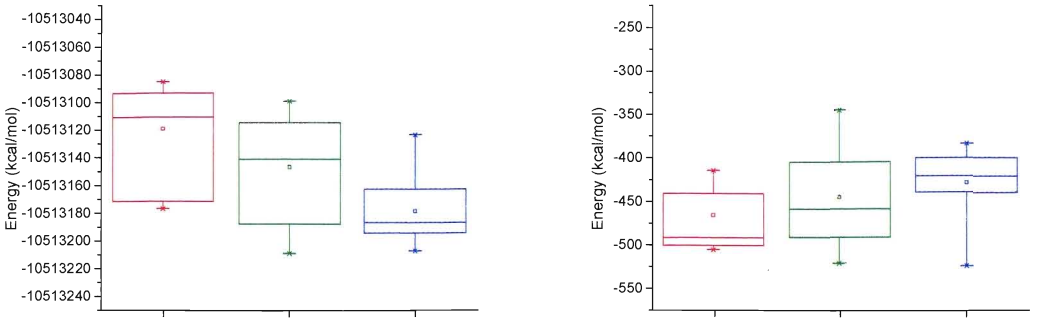


Figure 33: Statistical boxplots of total quantum-mechanical FMO potential energies for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, excluding the central residues of the first, second and third polyglutamine segments (plots constructed using Origin 7.5 graphing software[86]). The potential energy value (in kcal/mol) is plotted along the vertical axis. The range of energy values covered by each peptide is indicated by a box with flanking vertical lines; and the 25^{th} percentile, median and 75^{th} percentile energy values are indicated by the three horizontal lines of the box. Red = PGQ_9 (8 structures); green = $PGQ_9(P^1)$ (12 structures); blue = $PGQ_9(P^{2,3})$ (14 structures). (a) Local conformational energies; (b) long-range non-bonded energies.

Upon examination of the analogous molecular mechanics potential energies (Section 2.3.4), it was found that in agreement with the quantum-mechanical energy results, the long-range non-bonded energy distributions for all three peptides overlapped (see Figure 34b). The molecular mechanics local conformational energies, on the other hand, deviated significantly from one another, following a trend similar to that observed for the molecular mechanics bonded energies (see Figure 34a and Table 3). Therefore, the quantum mechanical FMO and molecular mechanics potential energy calculations both indicate that the energy differences between the peptides are primarily local-conformational in nature, with little appreciable long-range non-bonded contribution. Thus, the occurrence of potential energy differences between the peptides as determined by molecular mechanics is consistent with what quantum mechanics would indicate,

confirming that the conclusions drawn from the molecular mechanics potential energy results must indeed be valid.

(a) (b)

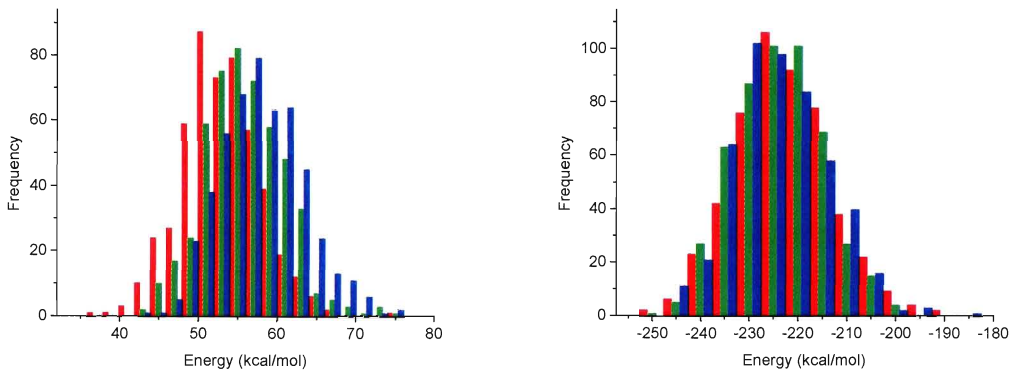


Figure 34: Frequency histograms of total molecular mechanics FMO-analogue potential energies (Section 2.3.4) for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, excluding the central residues of the first, second and third polyglutamine segments. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting each energy value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$. (a) Local conformational energies; (b) long-range non-bonded energies.

3.3.3. Molecular Mechanics Total Potential Energies: Further Analysis

Upon examination of the final sets of computed molecular mechanics and quantum mechanical FMO energies, it was found that the bonded and local conformational energy distributions for all three peptides overlapped (see Figures 35 and 36, and Table 4) – in contrast to the results from the preceding energy calculations. Therefore, while the differences between the peptides have proven to be due primarily to local, geometry-based effects, as reflected by the previously-observed inter-peptide differences in bonded/local conformational potential energies (Sections 3.3.1 and 3.3.2), the effects are still quite localized in nature. In particular, the final energy calculations confirm that the differences in bonded/local conformational potential energies between

the peptides are confined to the central amino acid residues of the first, second and third polyQ segments, and the residues immediately amino-terminal to the central residues – the same residues whose local structure tendencies were affected by the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, as observed from the peptide backbone dihedrals analysis (see Section 3.2.1). Therefore, the local structure analysis and potential energy calculations both confirm that the glutamine→proline mutations of $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ exert a localized conformational effect on the structures of the peptides.

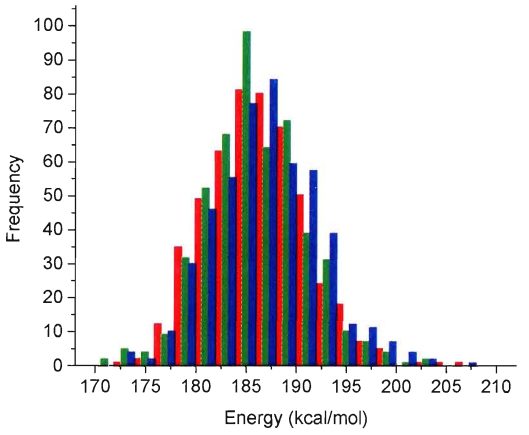


Figure 35: Frequency histograms of molecular mechanics total bonded energies for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, excluding the central residues of the first, second and third polyglutamine segments, and the residues immediately amino-terminal to the central residues. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting each energy value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$.

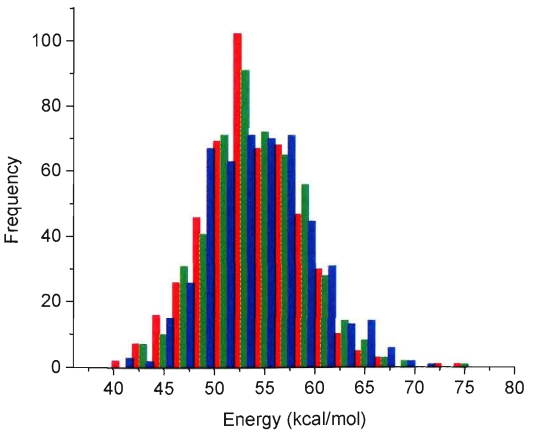


Figure 36: Frequency histograms of molecular mechanics total local conformational energies for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$, excluding the central residues of the first, second and third polyglutamine segments, and the residues immediately amino-terminal to the central residues. The potential energy value (in kcal/mol) is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting each energy value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$.

3.4. Peptide Comparisons: Global Structure Analysis

3.4.1. Automated Histogram Filtering (AHF) Cluster Analysis

Upon examination of the trends in number of identified clusters for the AHF analysis of inter-α-carbon distances including the carboxy-terminus (Section 2.4.1), it was found that the three peptides demonstrated somewhat different rates of increase in the number of identified clusters with decreasing peak distinction value (see Figure 37). The number of clusters identified for PGQ_9 did not begin to rise sharply until a peak distinction value of 3 had been reached, while the numbers of clusters identified for $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ began to rise sharply somewhat sooner than PGQ_9, at a peak distinction value of 4. In addition, the number of clusters identified for $PGQ_9(P^{2,3})$ began to increase above one at a peak distinction value approximately 50% higher than did

PGQ_9 or $PGQ_9(P^1)$. Therefore, the global structure behaviours of the three peptides could be distinguished from one another in terms of structural variation.

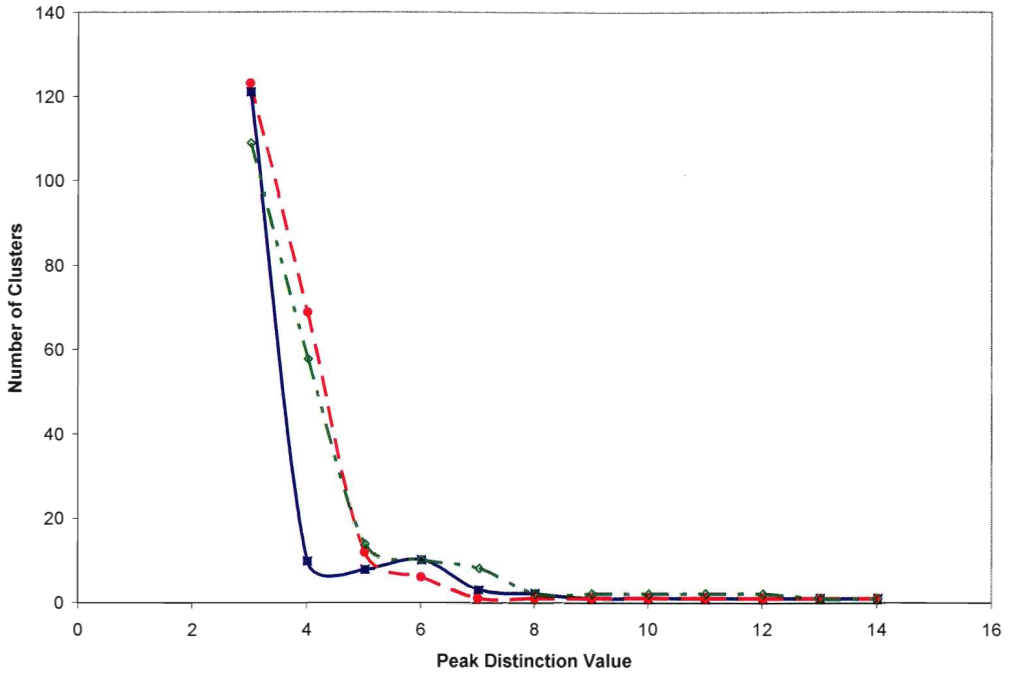


Figure 37: Plots of the number of clusters of related structures identified by AHF analysis, versus the numerical value of the peak distinction parameter employed in the analysis. All amino acid residues of each peptide were included in the analysis. Blue line = PGQ_9; red line = $PGQ_9(P^1)$; green line = $PGQ_9(P^{2,3})$.

Upon examination of the trends in number of identified clusters for the AHF analysis of inter-α-carbon distances excluding the carboxy-terminus, it was found that the PGQ_9 and $PGQ_9(P^1)$ peptides only ever yielded one cluster of structures each (see Figure 38). The $PGQ_9(P^{2,3})$ peptide, on the other hand, demonstrated a rapid increase in the number of identified clusters with decreasing peak distinction value, starting at a peak distinction value of 6. Therefore, the global structure behaviour of $PGQ_9(P^{2,3})$ could be greatly distinguished from that of PGQ_9 and $PGQ_9(P^1)$ in terms of structural variation – an extent of difference that was hinted at by the fact that when the carboxy-terminus was

included, the number of clusters identified for $PGQ_9(P^{2,3})$ began to increase above one at a peak distinction value appreciably higher than for PGQ_9 or $PGQ_9(P^1)$. Thus, the global structural variation of $PGQ_9(P^{2,3})$ proved to be very different from that of the other two peptides, and only minor differences appeared to exist between PGQ_9 and $PGQ_9(P^1)$.

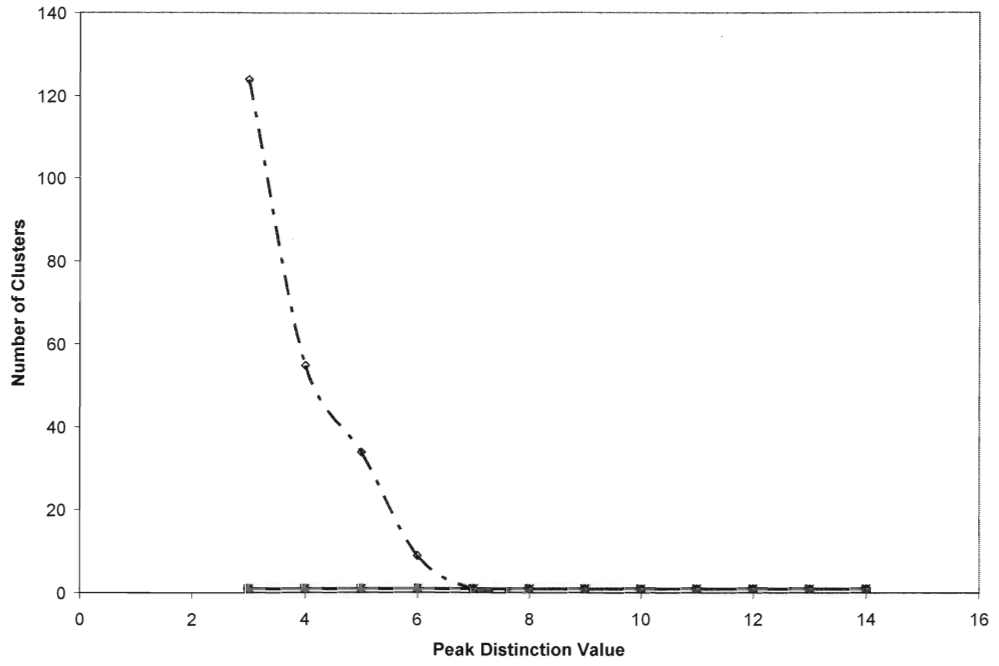


Figure 38: Plots of the number of clusters of related structures identified by AHF analysis, versus the numerical value of the peak distinction parameter employed in the analysis. The amino acid residues of the carboxy-terminal end of each peptide – which was structurally-constrained during the SA-MD simulations – were excluded from the analysis, to permit examination of only the portion of the peptide whose structure was free to vary during the simulations. Blue line = PGQ_9; red line = $PGQ_9(P^1)$; green line = $PGQ_9(P^{2,3})$.

Examination of the RMSD results for all three peptides (Section 2.4.1) revealed that the three peptides gave nearly identical distributions of RMSD values between structures (see Figure 39). Therefore, the differences in structural variation observed among the peptides (as described above) appear to be attributable to the structure clustering situation outlined in Section 2.4.1 and Figure 15 – *i.e.* a greater rate of increase

in the number of clusters with decreasing peak distinction being attributable to greater discontinuity in the range of possible global structures.

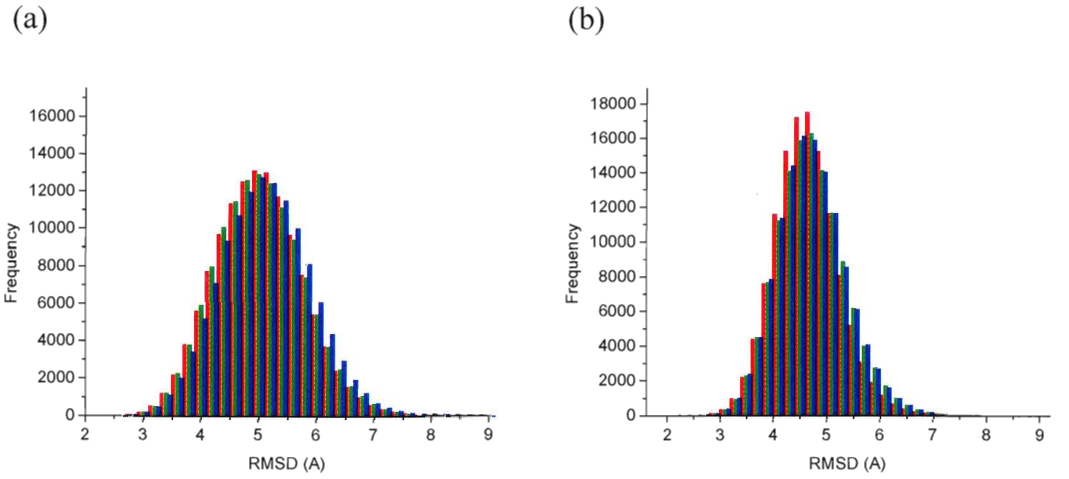


Figure 39: Frequency histograms of inter-α-carbon distance RMSDs computed for PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$. The RMSD value (in Å) is plotted along the horizontal axis, and the number of peptide structure pairs exhibiting each RMSD value is plotted along the vertical axis. Red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$. (a) Calculated RMSD values across all amino acid residues; (b) calculated RMSD values excluding the amino acid residues of each peptide's carboxy-terminal end.

According to the above results, $PGQ_9(P^{2,3})$ demonstrates a substantially reduced range of possible global structures compared to the other two peptides, while $PGQ_9(P^1)$ demonstrates only a marginally reduced range of possible global structures compared to PGQ_9. Therefore, it is reasonable to infer that $PGQ_9(P^{2,3})$ exhibits substantially greater global structural rigidity than the other two peptides, while $PGQ_9(P^1)$ exhibits only marginally greater global structural rigidity than PGQ_9.

3.4.2. In-Register Contacts: Assessment of Global β-Sheet Tendency

Upon examination of the in-register contact counts for the three peptides, $PGQ_9(P^{2,3})$ demonstrated evidence of a significant shift toward formation of fewer in-register contacts – and thus, a lesser extent of global β-sheet structure tendency –

compared to PGQ_9 and $PGQ_9(P^1)$ (see Figure 40 and Table 5). Indeed, visual examination of the structures of each peptide (see Appendix A, for example) revealed structures resembling two- or three-stranded sheets in the case of PGQ_9 and $PGQ_9(P^1)$, but poor tendency for even two-stranded structures in the case of $PGQ_9(P^{2,3})$. Thus, the visually-observed extent of difference between the peptides was, to a certain extent, in agreement with the extent of difference observed from the in-register contacts analysis – thereby validating the use of the method.

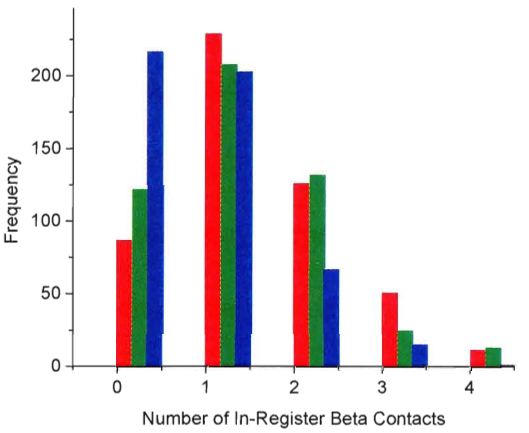


Figure 40: Frequency histograms of the occurrence of in-register β-sheet contacts within PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$. The number of contacts within the peptide is plotted along the horizontal axis, and the number of peptide structures (out of the set of 500 structures) exhibiting that number of contacts is plotted along the vertical axis. The greater the value along the horizontal axis at which the histogram reaches a maximum value along the vertical axis, the stronger the measured tendency is for global β-sheet structure formation. Total in-register contact counts across all three pairs of interacting polyglutamine segments (outlined in Figure 16) are shown: red = PGQ_9; green = $PGQ_9(P^1)$; blue = $PGQ_9(P^{2,3})$.

Table 5: Results of Two-Sample Z-Tests for Differences in Numbers of In-Register β-Sheet Contacts

Peptide Pair	$Z_{calculated}$	$Z_{critical}$[a]	Significant Difference Between The Two Peptides?[b]
PGQ_9 and $PGQ_9(P^1)$	1.90	2.58	No
PGQ_9 and $PGQ_9(P^{2,3})$	10.37	2.58	Yes
$PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$	7.95	2.58	Yes

[a]The test was performed at the 1% significance (99% confidence) level, in two-tailed format.
[b]A significant difference exists if $-Z_{critical} \leq Z_{calculated} \leq Z_{critical}$.

These results are interesting for two reasons. First of all, $PGQ_9(P^{2,3})$ was the peptide that failed to aggregate experimentally, while PGQ_9 and $PGQ_9(P^1)$ both successfully aggregated – although at different rates.[26] Thus, the observed differences in global β-sheet structure tendency appear to be consistent with the differences in aggregation behaviour of the three peptides, further supporting the idea that disruption of monomeric β-sheet structure propensity may be responsible for the observed inhibitory effect on aggregation (as discussed in Section 1.5). Secondly, $PGQ_9(P^{2,3})$ demonstrated what appeared to be a likely tendency toward greater global structure rigidity than the other two peptides (see Section 3.4.1). Therefore, it is reasonable to infer from the results that increased global structure rigidity may result in a decreased global β-sheet structure propensity, thereby triggering the experimentally-observed inhibition of peptide aggregation.

However, very few in-register contacts were generally identified in the structures obtained by SA-MD, and so the quantitative evidence of differences in β-sheet propensity was still rather weak (as predicted for the SA-MD methodology; see Section 3.1.2). Therefore, further simulation work would be warranted in the future to more thoroughly

evaluate the β-sheet tendencies of the three peptides, and the extent to which these tendencies differ among the peptides.

4. SUMMARY AND CONCLUSIONS

4.1. Concluding Remarks

Regarding the mutated variants of Thakur and Wetzel's aggregation-prone PGQ_9 peptide, it was hypothesized that the additional proline mutations within the PGQ_9 variants (see Table 1) exert localized conformational restraint on the peptide backbone, confining the backbone to an existing conformational tendency for polyproline type II (PPII) and/or turn/bend structure (see Section 1.5). It was also hypothesized that such restraint would in turn tend to confine the peptides to a native global preference for coil structure over β-sheet structure, discouraging the already-unfavourable transition towards the β-sheet structure required for aggregate nucleation/growth. To investigate this hypothesis, the structural propensities of three of Thakur and Wetzel's peptides – PGQ_9, $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ – were examined, with emphasis on low-energy structures of the peptides.

Control analyses performed on the PGQ_9 peptide suggested that, in agreement with experimental studies of polyglutamine aggregation, transition of the peptide monomer toward β-sheet structure is conformationally-disfavoured. In particular, formation of β-strand structure within the peptide backbone of the four polyQ segments of PGQ_9 was shown to be the source of the observed conformational destabilization. Therefore, the structural data utilized in the current study appears to be in line with the experimentally-determined idea of an unfavourable transition (from native coil structure

to abnormal β-sheet structure) during polyQ aggregate nucleation. Thus, conformational perturbations favouring a native preference for coil structure should indeed discourage the β-sheet transition as was hypothesized, thereby inhibiting aggregation of the peptide.

Examination of the local secondary structure tendencies exhibited by structures of all three selected peptides revealed evidence that the glutamine→proline mutations introduced in $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ exert localized conformational restraint that confines the peptide backbone to existing local tendencies for aggregation-incompetent PPII and/or turn/bend structure – which would in turn tend to increase the global preference for coil structure over β-sheet structure. In addition, the potential energy calculations confirmed that the glutamine→proline mutations exert a localized conformational effect on the structures of the peptides, and ruled out the alternative possibility of appreciable non-bonded perturbations – *i.e.* beyond the inherent loss of hydrogen-bonding functional groups incurred by the glutamine→proline mutations – which is also possible with proline mutations. Therefore, as hypothesized, the glutamine→proline mutations of $PGQ_9(P^1)$ and $PGQ_9(P^{2,3})$ exert a localized primary perturbation of peptide structure based on conformational confinement of the peptide backbone to an existing tendency for PPII and/or turn/bend structure, which would in turn tend to increase the global preference for coil structure over β-sheet structure.

Subsequent analysis of the global structural tendencies of the peptides indicated a difference in the range of possible global structures, with $PGQ_9(P^{2,3})$ exhibiting a substantially reduced range of possible global structures compared to the other two peptides. In addition, $PGQ_9(P^{2,3})$ demonstrated weak evidence of a lesser extent of global β-sheet structure tendency than PGQ_9 or $PGQ_9(P^1)$ – a result that coincided with the

reduced range of global structures observed for $PGQ_9(P^{2,3})$, as well as the failure of $PGQ_9(P^{2,3})$ to aggregate experimentally when PGQ_9 and $PGQ_9(P^1)$ both aggregated.[26] Therefore, the localized primary perturbation of peptide structure exerted by the glutamine→proline mutations appears to trigger a secondary, entropic perturbation of global peptide structure, discouraging random conformational fluctuations which may be involved in the unfavourable transition towards β-sheet structure,[47,64,65] and so confining the global structure to an existing preference for the coil state. Indeed, introduction of such mutations into the polyQ segments of PGQ_9 would result in mutant polyQ sequences containing glutamine/proline-rich regions, in which intervals of only four or five residues separate consecutive prolines (see Table 2). Thus, the affected portion of the peptide would become too rigid to make the transition from coil structure to β-sheet structure, and as a result, the glutamine→proline mutations should inhibit β-sheet assembly while favouring the coil structure state.

It should be noted, however, that the global perturbation relative to PGQ_9 was particularly prominent in the case of $PGQ_9(P^{2,3})$, with little perturbation in the case of $PGQ_9(P^1)$. Thus, the localized primary effect of the glutamine→proline mutations appears to trigger an appreciable secondary effect most strongly if central polyQ segments of the PGQ_9 sequence become affected by the mutations – as is the case with $PGQ_9(P^{2,3})$. This finding is not surprising, as it is consistent with the rigidity introduced by the terminal segment mutation of $PGQ_9(P^1)$ remaining confined to the end of the peptide, with flexibility in the rest of the peptide remaining unaffected. Consequently, the rest of the peptide would remain flexible enough to make the transition to β-sheet structure with minimal hindrance. Also, subsequent monomer-monomer associations

would require structural rearrangement of only a small part of the peptide – *i.e.* the affected end – which could be achieved with little difficulty, thus explaining this peptide's retention of aggregation propensity (see Figure 41). However, this additional folding step would tend to slow down the overall β-sheet assembly process, thus explaining the slower observed aggregation rate for $PGQ_9(P^1)$ compared to PGQ_9 (see Figure 9).[26]

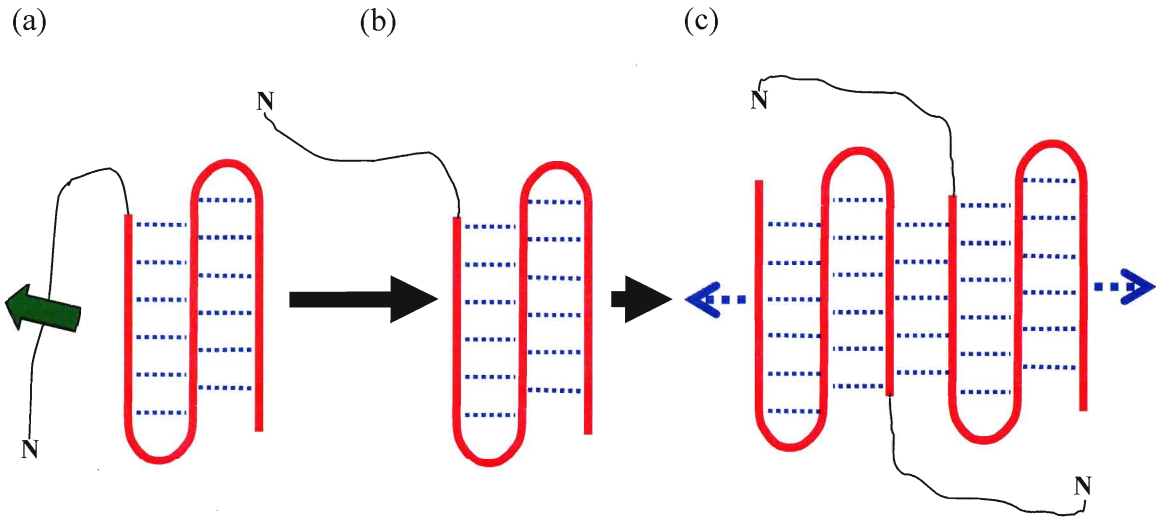


Figure 41: Peptide folding considerations in the aggregation of $PGQ_9(P^1)$. Confinement of added rigidity to the amino-terminal end of the peptide leaves the peptide flexible enough to arrange itself into β-sheet structure (a), and then fold so as to place the affected end of the peptide out of the way of subsequent monomer-monomer associations (b-c), thereby allowing aggregation to proceed despite the added rigidity. Hydrogen bond formation is illustrated by blue dotted lines, and for clarity, the amino terminus of each peptide is marked by an "N".

On the other hand, if the mutations are introduced into the middle of the peptide, then the resulting rigidity will affect a more substantial fraction of the peptide – as is indeed reflected by the observed global structure differences between peptides. As a result, structural rearrangements required to accommodate the peptide's mutations into stable β-sheet structure can no longer be readily achieved, as this would require

substantial unfolding, and thus an unfavourable extent of intramolecular hydrogen bond breakage. Therefore, a stable structure suitable for aggregate nucleation cannot be established, explaining the observed altogether-lack of aggregation for peptides such as $PGQ_9(P^{2,3})$ (see Figure 9).[26]

This would not be the first time that the idea of introducing localized conformational confinement to trigger an inhibition of aggregation has been proposed for an amyloidogenic peptide. Indeed, in a recent experimental study of β-amyloid by Kapurniotu *et al.* (2003),[96] a covalent constraint was applied to the middle of the β-amyloid peptide Aβ(1-28), resulting in the direct confinement of the middle of the peptide to its native α-helix structure. Such localized constraint of the middle of the peptide to its existing structural tendency effectively abolished aggregation, indicating an underlying disruption of propensity for the β-sheet aggregate structure. Therefore, the results of this study follow the same basic premise as those of the current study, reaching the conclusion that local conformational confinement favouring native global structure can inhibit global transition towards abnormal β-sheet structure, inhibiting amyloidogenic aggregation of the peptide in question. Thus, in designing $PGQ_9(P^{2,3})$ and other non-aggregating peptides, Thakur and Wetzel (2002) appear to have incorporated this very idea for inhibiting polyglutamine aggregation. However, designing confinement to coil structure in a rational manner is not as easy as designing confinement to ordered secondary structure (as was done by Kapurniotu *et al.*), and so the fact that such an idea for native coil structure appears to have been previously overlooked is not surprising.

Besides inhibition of the peptide's own aggregation, confinement to coil structure would also permit peptides such as $PGQ_9(P^{2,3})$ to act *in trans* as inhibitors of the growth

of existing polyQ oligomers/fibrils, and thus protect against cytotoxicity of such polyQ species – an inhibition and protection that were indeed observed in later experiments by Thakur *et al.* (2004).[97] In this mechanism, the $PGQ_9(P^{2,3})$ would bind to the end of the growing oligomer/fibril *via* polar-zipper interaction (see Section 1.2) of one of its non-mutated, terminal polyQ segments with the end of the oligomer/fibril β-sheet structure. However, the inherent conformational confinement would trap the remainder of the peptide in a coil conformation, effectively creating a constrained-coil cap on the end of the oligomer/fibril β-sheet structure, and preventing further aggregate growth by blocking β-strand addition to the end of the β-sheet (see Figure 42). Therefore, polyQ aggregation inhibitors based on peptide conformational confinement are a very possible means of inhibiting polyQ aggregation, and development of such rational inhibitors could pave the way for effective future treatments of Huntington's and other diseases associated with polyQ aggregation.

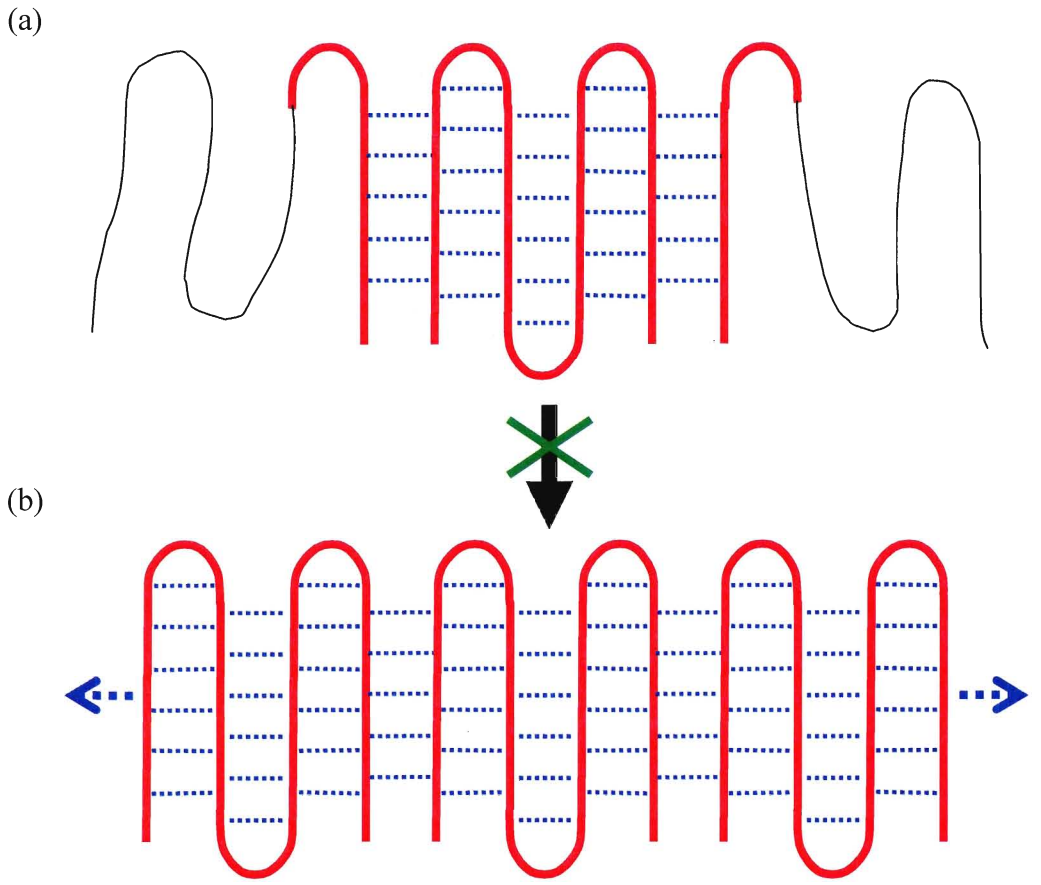


Figure 42: Inhibition of polyglutamine aggregate growth by bound $PGQ_9(P^{2,3})$. Although it could easily interact with either end of the aggregate *via* binding of one of its non-mutated, terminal polyglutamine segments, the remainder of the peptide remains entrapped in the coil structure state (a). Consequently, the peptide cannot fully make the transition to β-sheet structure (b), which would otherwise provide an exposed β-strand as the necessary site for binding of further β-sheet subunits to the aggregate.

4.2. Limitations and Future Work

As with any computational study, the current study has its limitations. For instance, due to the approximation of solvent effects using a fixed dielectric constant, interactions of the peptide with the solvent may not have been optimally represented. Thus, conclusions drawn from the study are necessarily qualitative in nature. In addition, the evidence of differences in β-sheet propensity that was obtained from the current study was rather weak. Therefore, further simulation work would be warranted in the future to

more thoroughly evaluate the β-sheet tendencies of the three peptides, and the extent to which these tendencies differ among the peptides. Finally, the antiparallel β-hairpin turn structures implemented in this study were all selected based on educated estimates, as the precise structure is not known with absolute certainty. Nevertheless, the results have proven to be congruent in several attributes with previously reported data, and so as described above (Section 4.1), the current study highlights some key ideas for consideration in the future design of effective treatments for Huntington's and other polyQ-associated diseases.

In the future, it may be desirable to perform further simulation work to more thoroughly evaluate the structural tendencies of the three peptides, and any differences in these tendencies among the peptides. Also, it would be desirable to perform the analyses of the current study on the other non-aggregating mutant variants of PGQ_9 that were examined by Thakur and Wetzel (2002; see Table 1),[26] in order to confirm whether similar patterns of structural perturbation exist within these peptides as well as $PGQ_9(P^{2,3})$. Finally, it would be interesting to extend the study to analysis of each mutant peptide bound to the end of a preformed antiparallel β-sheet monomer or oligomer (simulating an existing polyQ aggregate), in order to further confirm the proposal regarding inhibition of further aggregate growth by these peptides (as described in Figure 42, and the end of Section 4.1).

References:

[1] 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. and 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.

[2] Rubinsztein, D.C., Leggo, J., Coles, R., Almqvist, E., Biancalana, V., Cassiman, J.J., Chotai, K., Connarty, M., Craufurd, D., Curtis, A., Curtis, D., Davidson, M.J., Differ, A.M., Dode, C., Dodge, A., Frontali, M., Ranen, N.G., Stine, O.C., Sherr, M., Abbott, M.H., Franz, M.L., Graham, C.A., Harper, P.S., Hedreen, J.C., Jackson, A., Kaplan, J.C., Losekoot, M., MacMillan, J.C., Morrison, P., Trottier, Y., Novelletto, A., Simpson, S.A., Theilmann, J., Whittaker, J.L., Folstein, S.E., Ross, C.A. and Hayden, M.R. 1996. "Phenotypic characterization of individuals with 30-40 CAG repeats in the Huntington Disease (HD) gene reveals HD cases with 36 repeats and apparently normal elderly individuals with 36-39 repeats." *American Journal of Human Genetics* **59**: 16-22.

[3] Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G.P., Davies, S.W., Lehrach, H. and Wanker, E.E. 1997. "Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates *in vitro* and *in vivo*." *Cell* **90**: 549-558.

[4] Wacker, J.L., Zareie, M.H., Fong, H., Sarikaya, M. and Muchowski, P.J. 2004. "Hsp70 and Hsp40 attenuate formation of spherical and annular polyglutamine oligomers by partitioning monomer." *Nature Structural and Molecular Biology* **11**: 1215-1222.

[5] The Huntington's Disease Collaborative Research Group. 1993. "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's Disease chromosomes." *Cell* **72**: 971-983.

[6] Lumsden, A.L., Henshall, T.L., Dayan, S., Lardelli, M.T. and Richards, R.I. 2007. "Huntingtin-deficient zebrafish exhibit defects in iron utilization and development." *Human Molecular Genetics* **16**: 1905-1920.

[7] Firdaus, W.J.J., Wyttenbach, A., Giuliano, P., Kretz-Remy, C., Currie, R.W. and Arrigo, A.P. 2006. "Huntingtin inclusion bodies are iron-dependent centers of oxidative events." *FEBS Journal* **273**: 5428-5441.

[8] Caviston, J.P., Ross, J.L., Antony, S.M., Tokito, M. and Holzbaur, E.L.F. 2007. "Huntingtin facilitates dynein/dynactin-mediated vesicle transport." *Proceedings of the National Academy of Science USA* **104**: 10045-10050.

[9] Strehlow, A.N.T., Li, J.Z. and Myers, R.M. 2007. "Wild-type huntingtin participates in protein trafficking between the Golgi and the extracellular space." *Human Molecular Genetics* **16**: 391-409.

[10] 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. and MacDonald, M.E. 1995. "Inactivation of the mouse Huntington's disease gene homolog Hdh." *Science* **269**: 407-410.

[11] Marchut, A.J. and Hall, C.K. 2007. "Effects of chain length on the aggregation of model polyglutamine peptides: molecular dynamics simulations." *Proteins: Structure, Function, and Bioinformatics* **66**: 96-109.
[12] Ogawa, H., Nakano, M., Watanabe, H., Starikov, E.B., Rothstein, S.M. and Tanaka, S. 2008. "Molecular dynamics simulation study on the structural stabilities of polyglutamine peptides." *Computational Biology and Chemistry* **32**: 102-110.
[13] Elliot, J., Starikov, E.B., Crawshaw, J., Claiden, P., Nilsson, L. and Windle, A. "Nucleation of polyglutamine amyloid fibres modeling using molecular dynamics." In *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, pp. 211-226. Elsevier B.V.: 2006.
[14] Sugaya, K., Matsubara, S., Kagamihara, Y., Kawata, A. and Hayashi, H. 2007. "Polyglutamine expansion mutation yields a pathological epitope linked to nucleation of protein aggregate: determinant of Huntington's Disease onset." *PLoS ONE* **2**: e635(1-7).
[15] Tanaka, M., Morishima, I., Akagi, T., Hashikawa, T. and Nukina, N. 2001. "Intra- and intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin as a model for polyglutamine diseases." *Journal of Biological Chemistry* **276**: 45470-45475.
[16] Nagai, Y., Inui, T., Popiel, H.A., Fujikake, N., Hasegawa, K., Urade, Y., Goto, Y., Naiki, H. and Toda, T. 2007. "A toxic monomeric conformer of the polyglutamine protein." *Nature Structural and Molecular Biology* **14**: 332-340.
[17] Singer, S.J. and Dewji, N.N. 2006. "Evidence that Perutz's double-β-stranded subunit structure for β-amyloids also applies to their channel-forming structures in membranes." *Proceedings of the National Academy of Science USA* **103**: 1546-1550.
[18] Michalik, A. and Van Broeckhoven, C. 2003. "Pathogenesis of polyglutamine disorders: aggregation revisited." *Human Molecular Genetics* **12**: R173-R186.
[19] Shimohata, T., Onodera, O. and Tsuji, S. 2000. "Interaction of expanded polyglutamine stretches with nuclear transcription factors leads to aberrant transcriptional regulation in polyglutamine diseases." *Neuropathology* **20**: 326-333.
[20] Ross, C.A. and Poirier, M.A. 2004. "Protein aggregation and neurodegenerative disease." *Nature Medicine* **10**: S10-S17.
[21] Chen, S., Berthelier, V., Yang, W. and Wetzel, R. 2001. "Polyglutamine aggregation behaviour *in vitro* supports a recruitment mechanism of cytotoxicity." *Journal of Molecular Biology* **311**: 173-182.
[22] Brusilow, W.S.A. 2006. "Is Huntington's a glutamine storage disease?" *The Neuroscientist* **12**: 300-304.
[23] Rochet, J.C. and Lansbury, P.T. 2000. "Amyloid fibrillogenesis: themes and variations." *Current Opinion in Structural Biology* **10**: 60-68.
[24] Perutz, M.F., Johnson, T., Suzuki, M. and Finch, J.T. 1994. "Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases." *Proceedings of the National Academy of Science USA* **91**: 5355-5358.
[25] Chen, S., Ferrone, F.A. and Wetzel, R. 2002. "Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation." *Proceedings of the National Academy of Science USA* **99**: 11884-11889.

[26] Thakur, A.K. and Wetzel, R. 2002. "Mutational analysis of the structural organization of polyglutamine aggregates." *Proceedings of the National Academy of Science USA* **99**: 17014-17019.
[27] Poirier, M.A., Jiang, H. and Ross, C.A. 2005. "A structure-based analysis of huntingtin mutant polyglutamine aggregation and toxicity: evidence for a compact beta-sheet structure." *Human Molecular Genetics* **14**: 765-774.
[28] Ignatova, Z. and Gierasch, L.M. 2006. "Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein." *Journal of Biological Chemistry* **281**: 12959-12967.
[29] Scherzinger, E., Sittler, A., Schweiger, K., Heiser, V., Lurz, R., Hasenbank, R., Bates, G.P., Lehrach, H. and Wanker, E.E. 1999. "Self-assembly of polyglutamine-containing huntingtin fragments into amyloid-like fibrils: implications for Huntington's disease pathology." *Proceedings of the National Academy of Science USA* **96**: 4604-4609.
[30] Bevivino, A.E. and Loll, P.J. 2001. "An expanded glutamine repeat destabilizes native ataxin-3 structure and mediates formation of parallel β-fibrils." *Proceedings of the National Academy of Science USA* **98**: 11955-11960.
[31] Bellesia, G. and Shea, J.E. 2007. "Self-assembly of β-sheet forming peptides into chiral fibrillar aggregates." *Journal of Chemical Physics* **126**: 245104(1-11).
[32] Serpell, L.C., Sunde, M. and Blake, C.C.F. 1997. "The molecular basis of amyloidosis." *Cellular and Molecular Life Sciences* **53**: 871-887.
[33] Röhrig, U.F., Laio, A., Tantalo, N., Parrinello, M. and Petronzio, R. 2006. "Stability and structure of oligomersof the Alzheimer peptide $A\beta_{16-22}$: from the dimer to the 32-mer." *Biophysical Journal* **91**: 3217-3229.
[34] Perutz, M.F., Finch, J.T., Berriman, J. and Lesk, A. 2002. "Amyloid fibers are water-filled nanotubes." *Proceedings of the National Academy of Science USA* **99**: 5591-5595.
[35] Sikorski, P. and Atkins, E. 2005. "New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils." *Biomacromolecules* **6**: 425-432.
[36] Sharma, D., Shinchuk, L.M., Inouye, H., Wetzel, R. and Kirschner, D.A. 2005. "Polyglutamine homopolymers having 8-45 residues form slab-like β-crystallite assemblies." *Proteins: Structure, Function, and Bioinformatics* **61**: 398-411.
[37] Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J.W., McFarlane, H.T., Madsen, A.O., Riekel, C. and Eisenberg, D. 2007. "Atomic structures of amyloid cross-β spines reveal varied steric zippers." *Nature* **447**: 453-457.
[38] Jin, L.W., Claborn, K.A., Kurimoto, M., Geday, M.A., Maezawa, I., Sohraby, F., Estrada, M., Kaminksy, W. and Kahr, B. 2003. "Imaging linear birefringence and dichroism in cerebral amyloid pathologies." *Proceedings of the National Academy of Science USA* **100**: 15294-15298.
[39] Li, L., Darden, T.A., Bartolotti, L., Kominos, D. and Pedersen, L.G. 1999. ''An atomic model for the pleated β-sheet structure of Aβ amyloid protofilaments.'' *Biophysical Journal* **76**: 2871-2878.
[40] Puchtler, H., Sweat, F. and Levine, M. 1962. "On the binding of Congo red by amyloid." *Journal of Histochemistry and Cytochemistry* **10**: 355-364

[41] Khare, S.D., Ding, F., Gwanmesia, K.N. and Dokholyan, N.V. 2005. "Molecular origin of polyglutamine aggregation in neurodegenerative diseases." *PLoS Computational Biology* **1**: 230-235.
[42] Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R. and Eisenberg, D. 2005. "Structure of the cross-β spine of amyloid-like fibrils." *Nature* **435**: 773-778.
[43] Chan, J.C.C., Oyler, N.A., Yau, W.M. and Tycko, R. 2005. "Parallel β-sheets and polar zippers in amyloid fibrils formed by residues 10-39 of the yeast prion protein Ure2p." *Biochemistry* **44**: 10669-10680.
[44] Tycko, R. 2004. "Progress towards a molecular-level structural understanding of amyloid fibrils." *Current Opinion in Structural Biology* **14**: 96-103.
[45] Thirumalai, D., Klimov, D.K. and Dima, R.I. 2003. "Emerging ideas on the molecular basis of protein and peptide aggregation." *Current Opinion in Structural Biology* **13**: 146-159.
[46] Tobelmann, M., Lee, C., Walters, R. and Murphy, R.M. 2007. "Folding, misfolding, and aggregation of polyglutamine peptides and proteins." *234th ACS National Meeting.*
[47] Bhattacharyya, A.M., Thakur, A.K. and Wetzel, R. 2005. "Polyglutamine aggregation nucleation: thermodynamics of a highly unfavourable protein folding reaction." *Proceedings of the National Academy of Science USA* **102**: 15400-15405.
[48] Lynn, D.G. and Meredith, S.C. 2000. "Review: model peptides and the physicochemical approach to β-amyloids." *Journal of Structural Biology* **130**: 153-173.
[49] Aggeli, A., Nyrkova, I.A., Bell, M., Harding, R., Carrick, L., McLeish, T.C.B., Semenov, A.N. and Boden, N. 2001. "Hierarchical self-assembly of chiral rod-like molecules as a model for peptide β-sheet tapes, ribbons, fibrils, and fibers." *Proceedings of the National Academy of Science USA* **98**: 11857-11862.
[50] Marchut, A.J. and Hall, C.K. 2006. "Spontaneous formation of annular structures observed in molecular dynamics simulations of polyglutamine peptides." *Computational Biology and Chemistry* **30**: 215-218.
[51] Lashuel, H.A., Petre, B.M., Wall, J., Simon, M., Nowak, R.J., Walz, T. and Lansbury, P.T. 2002. "α-Synuclein, especially the Parkinson's disease-associated mutants, forms pore-like annular and tubular protofibrils." *Journal of Molecular Biology* **322**: 1089-1102.
[52] Ding, T.T., Lee, S.J., Rochet, J.C. and Lansbury, P.T. 2002. "Annular α-synuclein protofibrils are produced when spherical protofibrils are incubated in solution or bound to brain-derived membranes." *Biochemistry* **41**: 10209-10217.
[53] Elam, J.S., Taylor, A.B., Strange, R., Antonyuk, S., Doucette, P.A., Rodriguez, J.A., Hasnain, S.S., Hayward, L.J., Valentine, J.S., Yeates, T.O. and Hart, P.J. 2003. "Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS." *Nature Structural Biology* **10**: 461-467.
[54] Ross, C.A., Poirier, M.A., Wanker, E.E. and Amzel, M. 2003. "Polyglutamine fibrillogenesis: the pathway unfolds." *Proceedings of the National Academy of Science USA* **100**: 1-3.
[55] Sunde, M., Serpell, L.C., Bartlam, M., Fraser, P.E., Pepys, M.B. and Blake, C.C.F. 1997. "Common core structure of amyloid fibrils by synchrotron X-ray diffraction." *Journal of Molecular Biology* **273**: 729-739.

[56] Melquiond, A. and Gelly, J.C. 2007. "Probing amyloid fibril formation of the NFGAIL peptide by computer simulations." *Journal of Chemical Physics* **126**: 065101(1-7).
[57] Kajava, A.V., Baxa, U., Wickner, R.B. and Steven A.C. 2004. "A model for Ure2p prion filaments and other amyloids: the parallel superpleated β-structure." *Proceedings of the National Academy of Science USA* **101**: 7885-7890.
[58] Goaverts, C., Wille, H., Prusiner, S.B. and Cohen, F.E. 2004. "Evidence for assembly of prions with left-handed β-helices into trimers." *Proceedings of the National Academy of Science USA* **101**: 8342-8347.
[59] Stork, M., Giese, A., Kretschmar, H.A. and Tavan, P. 2005. "Molecular dynamics simulations indicate a possible role of parallel β-helices in seeded aggregation of poly-Gln." *Biophysical Journal* **88**: 2442-2451.
[60] Kajava, A.V., Cheng, N., Cleaver, R., Kessel, M., Simon, M.N., Willery, E., Jacob-Dubuisson, F., Locht, C. and Steven, A.C. 2001. "Beta-helix model for the filamentous haemagglutinin adhesion of *Bordetella pertussis* and related bacterial secretory proteins." *Molecular Microbiology* **42**: 279-292.
[61] Darnell, G., Orgel, J.P.R.O., Pahl, R. and Meredith, S.C. 2007. "Flanking polyproline sequences inhibit β-sheet structure in polyglutamine segments by inducing PPII-like helix structure." *Journal of Molecular Biology* **374**: 688-704.
[62] Chellgren, B.W., Miller, A.F. and Creamer, T.P. 2006. "Evidence for polyproline II helical structure in short polyglutamine tracts." *Journal of Molecular Biology* **361**: 362-371.
[63] Vitalis, A., Wang, X. and Pappu, R.V. 2008. "Atomistic simulations of the effects of polyglutamine chain length and solvent quality on conformational equilibria and spontaneous homodimerization." *Journal of Molecular Biology* **384**: 279-297.
[64] Chopra, M., Reddy, A.S., Abbott, N.L. and de Pablo, J.J. 2008. "Folding of polyglutamine chains." *Journal of Chemical Physics* **129**: 135102(1-8).
[65] Li, M.S., Klimov, D.K., Straub, J.E. and Thirumalai, D. 2008. "Probing the mechanisms of fibril formation using lattice models." *Journal of Chemical Physics* **129**: 175101(1-10).
[66] Gordon, H.L., Kwan, W.K., Gong, C., Larrass, S. and Rothstein, S.M. 2003. "Efficient generation of low-energy folded states of a model protein." *Journal of Chemical Physics* **118**: 1533-1540.
[67] Kabsch, W. and Sander, C. 1983. "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features." *Biopolymers* **22**: 2577-2637.
[68] Gordon, H.L. and Rothstein, S.M. "Protein structure generation and elucidation: applications of automated histogram filtering cluster analysis." In *Modern Methods for Theoretical Physical Chemistry of Biopolymers*, pp. 325-336. Elsevier B.V.: 2006.
[69] Imada, J., Chapman, P. and Rothstein, S. 2005. "Recognizing patterns in high-dimensional data: automated histogram filtering for protein structure elucidation." *Proceedings of the 19th International Symposium on High Performance Computing Systems and Applications*: 238- 243.
[70] Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. 1983. "Optimization by simulated annealing." *Science* **220**: 671-680.

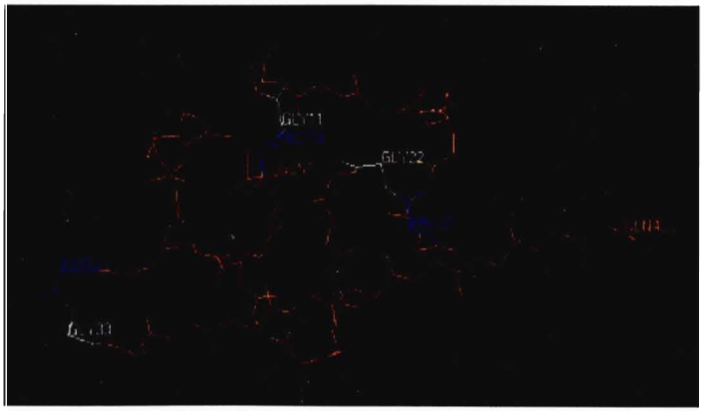
[71] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M. 1983. "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations." *Journal of Computational Chemistry* **4**: 187-217.
[72] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M. 1998. "All-atom empirical potential for molecular modeling and dynamics studies of proteins." *Journal of Physical Chemistry B* **102**: 3586-3616.
[73] Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L. and Schulten, K. 2005. "Scalable molecular dynamics with NAMD." *Journal of Computational Chemistry* **26**: 1781-1802.
[74] The Shared Hierarchical Academic Research Computing Network (SHARCNET): www.sharcnet.ca.
[75] Theoretical Biophysics Group, Beckman Institute. "VMD Version 1.8.4." 2006.
[76] Voet, D. and Voet, J.G. *Biochemistry, 3rd Edition*, p. 230. John Wiley and Sons, Inc.: 2004.
[77] Thompson, M.A. and Planaria Software LLC. "ArgusLab Version 4.0.1." 2004.
[78] Cruzeiro, L. 2005. "Why are proteins with glutamine- and asparagine-rich regions associated with protein misfolding diseases?" *Journal of Physics: Condensed Matter* **17**: 7833-7844.
[79] Tuckerman, M., Berne, B.J., Martyna, G.J. 1992. "Reversible multiple time scale molecular dynamics." *Journal of Chemical Physics* **97**: 1990-2001.
[80] Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. 1977. "Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes." *Journal of Computational Physics* **23**: 327–341.
[81] Guex, N., Peitsch, M., Schwede, T. and Diemand, A. "Swiss-PdbViewer Version 3.7." 2001.
[82] Merlino, A., Esposito, L. and Vitagliano, L. 2006. "Polyglutamine repeats and β-helix structure: molecular dynamics study." *Proteins: Structure, Function, and Bioinformatics* **63**: 918-927.
[83] Barakat, M.T. and Dean, P.M. 1990. "Molecular structure matching by simulated annealing. I. A comparison between different cooling schedules." *Journal of Computer-Aided Molecular Design* **4**: 295-316.
[84] Leitgeb, B., Kerényi, A., Bogár, F., Paragi, G., Penke, B. and Rákhely, G. 2007. "Studying the structural properties of polyalanine and polyglutamine peptides." *Journal of Molecular Modeling* **13**: 1141-1150.
[85] Accelrys Inc. "Quanta 2005." 2005.
[86] OriginLab Corp. "Origin Version 7.5." 2006.
[87] Nakano, T., Kaminuma, T., Sato, T., Akiyama, Y., Uebayasi, M. and Kitaura, K. 2000. "Fragment molecular orbital method: application to polypeptides." *Chemical Physics Letters* **318**: 614-618.
[88] Society for Industrial and Applied Mathematics. "ScaLAPACK Version 3.1.0."

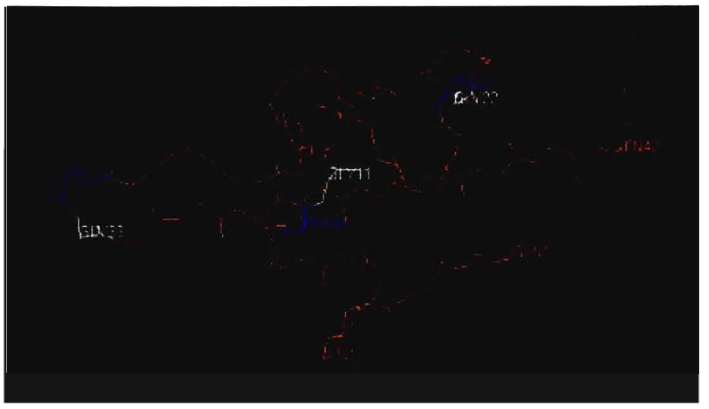
[89] Nakano, T., Kaminuma, T., Sato, T., Fukuzawa, K., Akiyama, Y., Uebayasi, M. and Kitaura, K. 2002. "Fragment molecular orbital method: use of approximate electrostatic potential." *Chemical Physics Letters* **351**: 475-480.
[90] Levine, I.N. *Quantum Chemistry, 5th Edition*, pp. 563-568. Prentice-Hall, Inc.: 2000.
[91] Fukuzawa, K., Komeiji, Y., Mochizuki, Y., Kato, A., Nakano, T. and Tanaka, S. 2006. "Intra- and intermolecular interactions between cyclic-AMP receptor protein and DNA: *Ab initio* fragment molecular orbital study." *Journal of Computational Chemistry* **27**: 948-960.
[92] Jonckheere, A.R. 1954. "A distribution-free k-sample test against ordered alternatives." *Biometrika* **41**: 133-145.
[93] Pimentel, G.C. and McClellan, A.L. *The Hydrogen Bond*, pp. 258-293. Reinhold Publishing Corporation: 1960.
[94] Guex, N., Peitsch, M., Schwede, T. and Diemand, A. "Swiss-PdbViewer Version 3.7." 2001.
[95] Brovchenko, I., Burri, R.R., Krukau, A., Oleinikova, A. and Winter, R. 2008. "Intrinsic thermal expansivity and hydrational properties of amyloid peptide $A\beta_{42}$ in liquid water." *Journal of Chemical Physics* **129**: 195101(1-11).
[96] Kapurniotu, A., Buck, A., Weber, M., Schmauder, A., Hirsch, T., Bernhagen, J. and Tatarek-Nossol, M. 2003. "Conformational restriction via cyclization in β-amyloid peptide Aβ(1-28) leads to an inhibitor of Aβ(1-28) amyloidogenesis and cytotoxicity." *Chemistry & Biology* **10**: 149-159.
[97] Thakur, A.K., Yang, W. and Wetzel, R. 2004. "Inhibition of polyglutamine aggregate cytotoxicity by a structure-based elongation inhibitor." *FASEB Journal* **18**: 923-925.

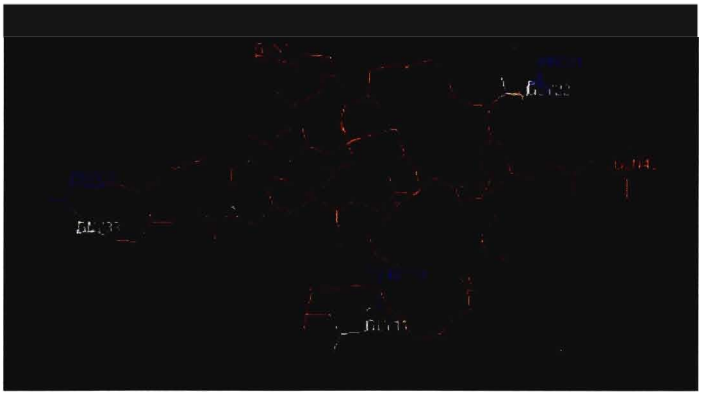
APPENDIX A: Stick Model Diagrams of the Representative Peptide Structures Selected for Analysis by Quantum-Mechanical Fragment Molecular Orbital (FMO) Calculations

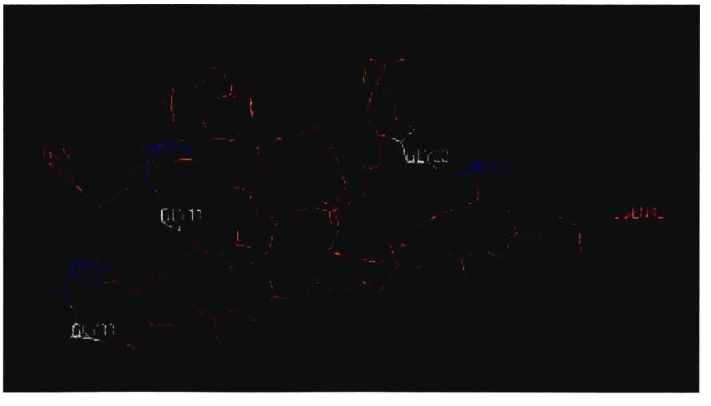
Note: The structures shown here were selected using automated histogram filtering (AHF) cluster analysis (as described in Section 2.3.3 of the thesis). For all structures shown, glutamine residues are highlighted in orange, proline residues are highlighted in blue, glycine residues are highlighted in white, and residues at the ends of the polyglutamine segments are labeled (for clarity) according to their numbers in the amino acid sequence.

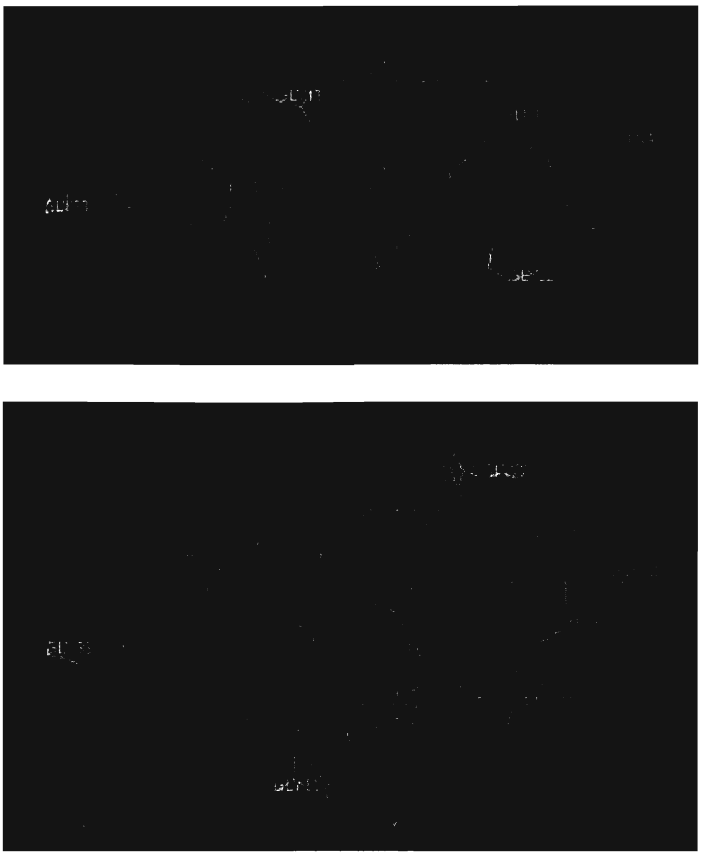
(a) Representative Structures of PGQ_9



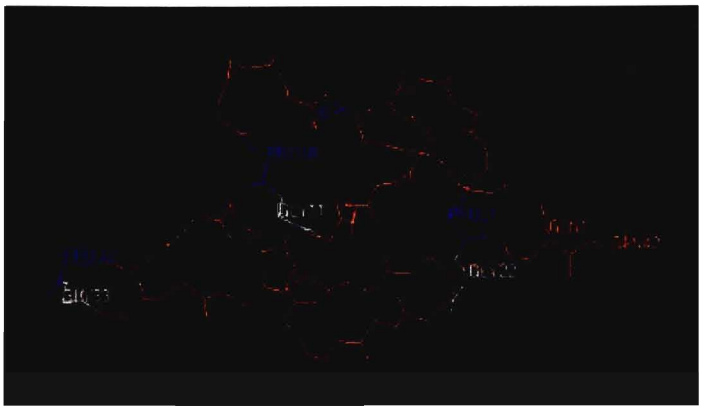


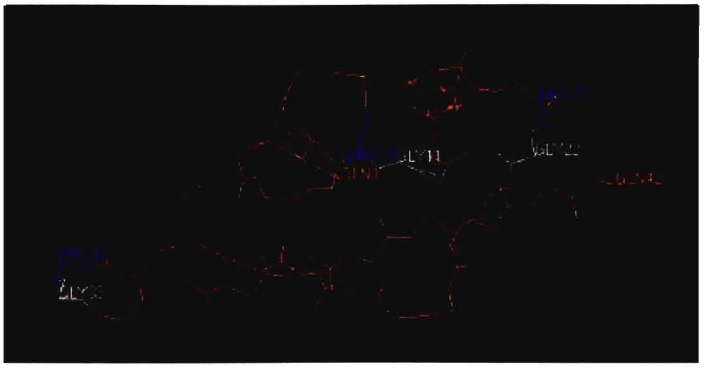


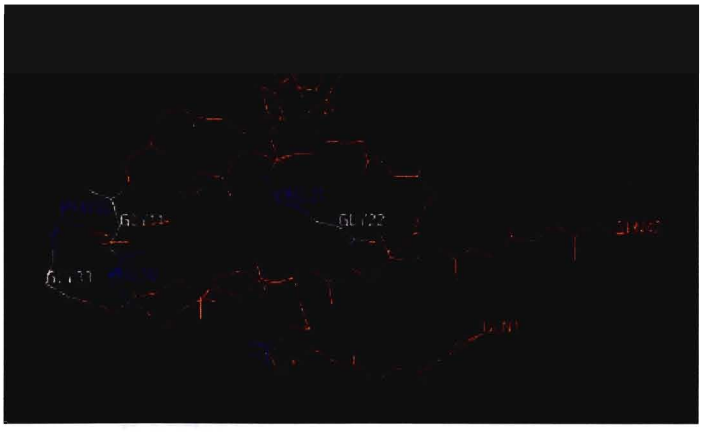


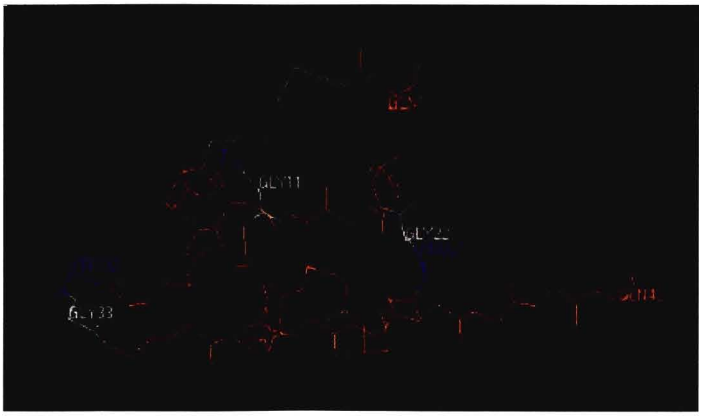


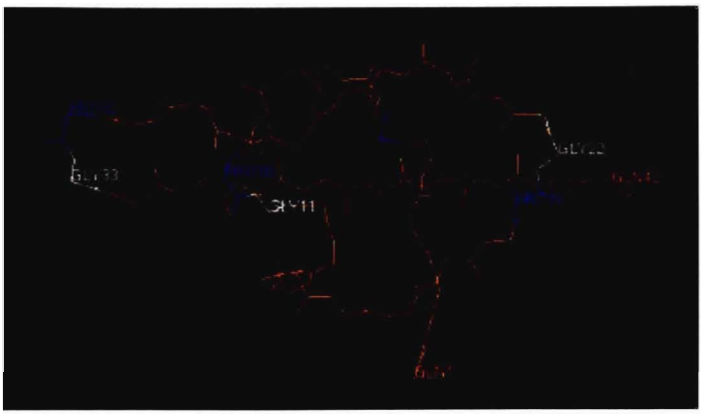
(b) Representative Structures of $PGQ_9(P^1)$

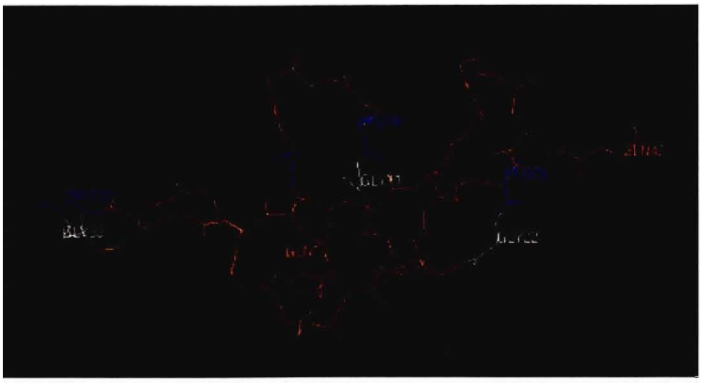


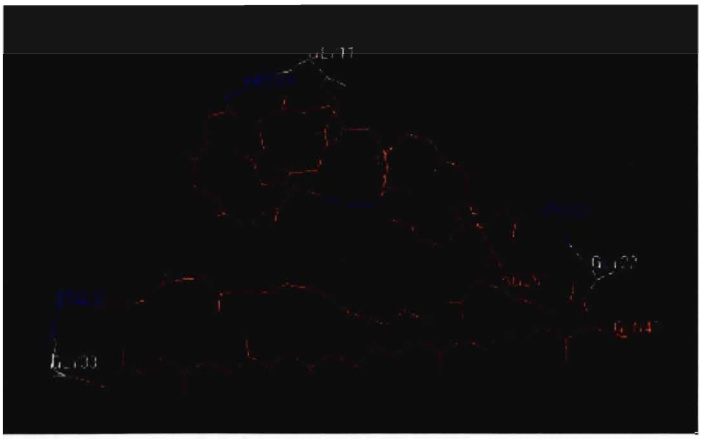


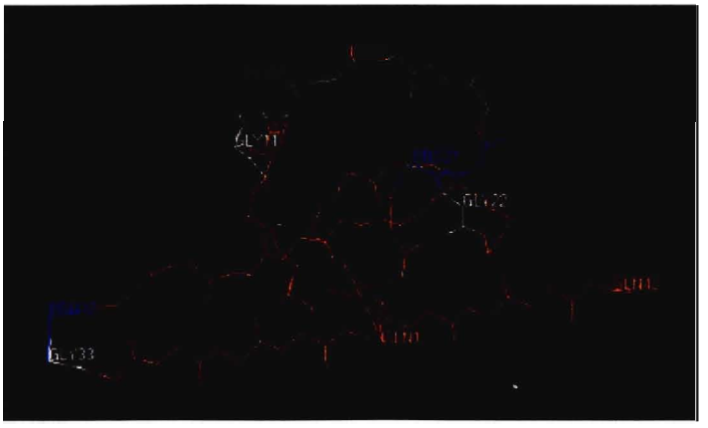


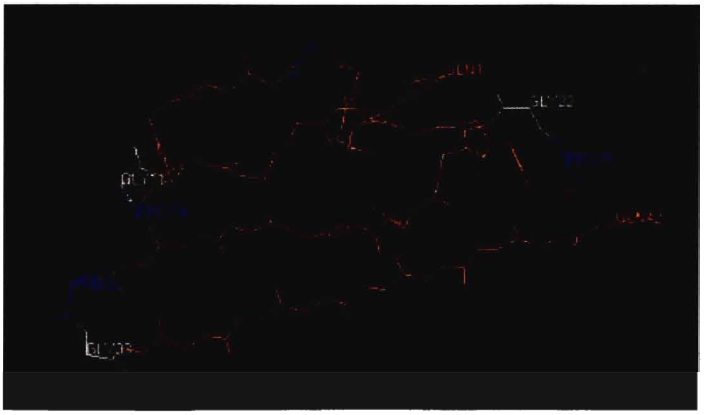


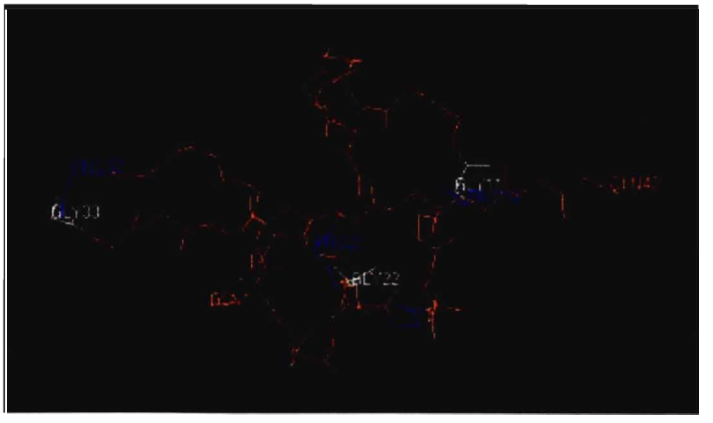




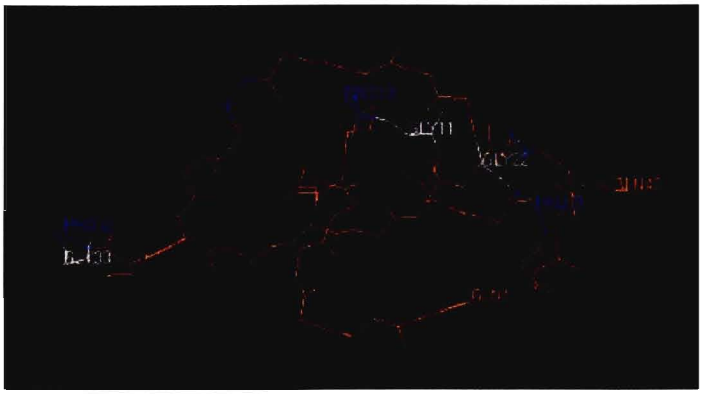


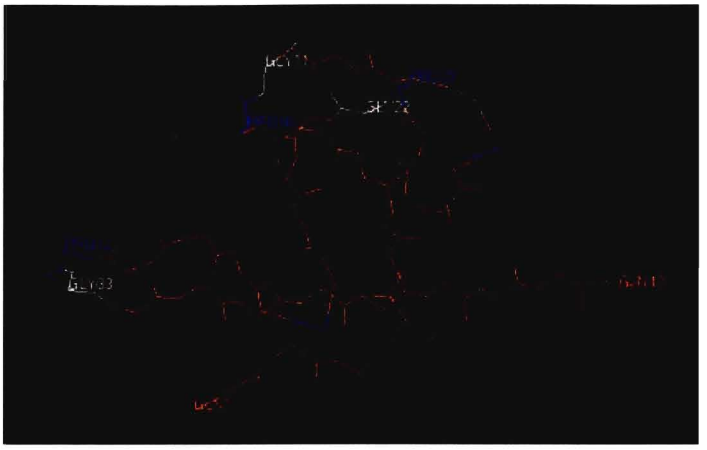


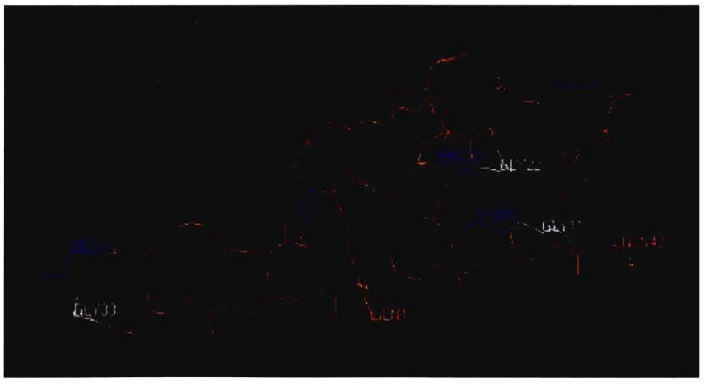


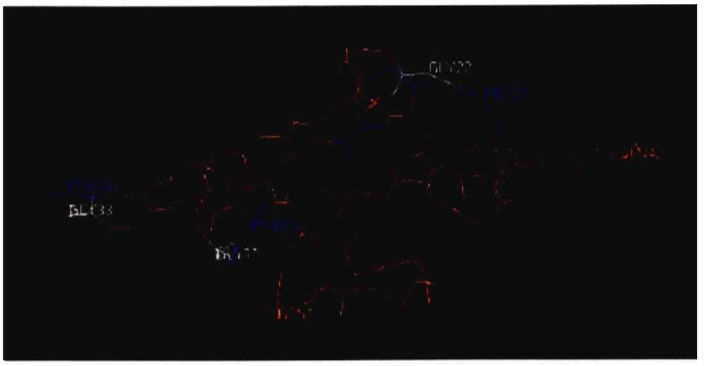


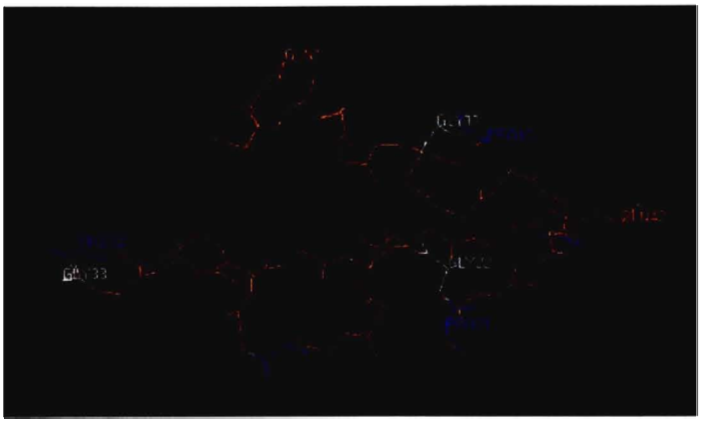
(c) Representative Structures of $PGQ_9(P^{2,3})$

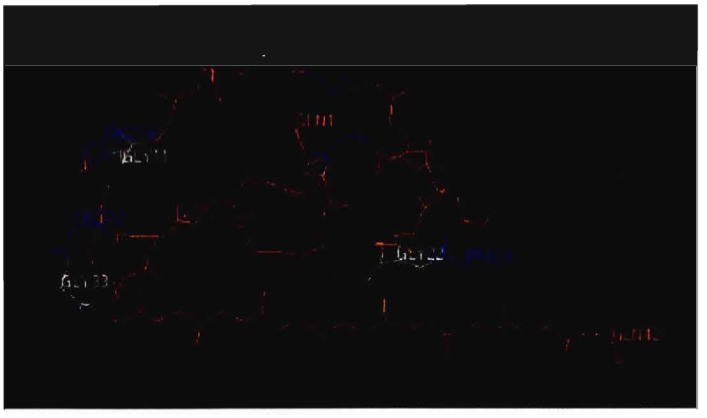


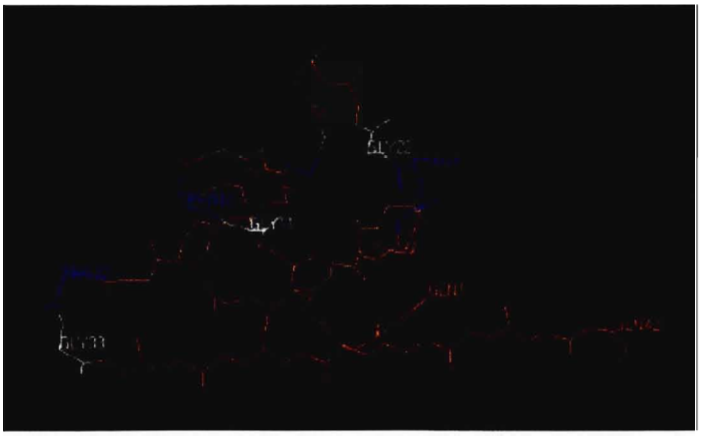


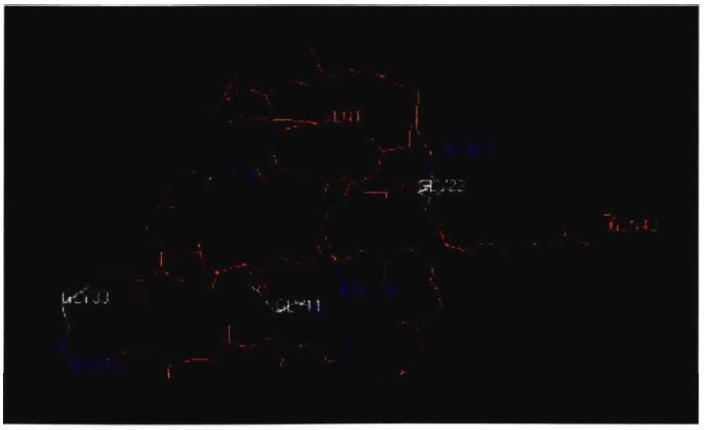


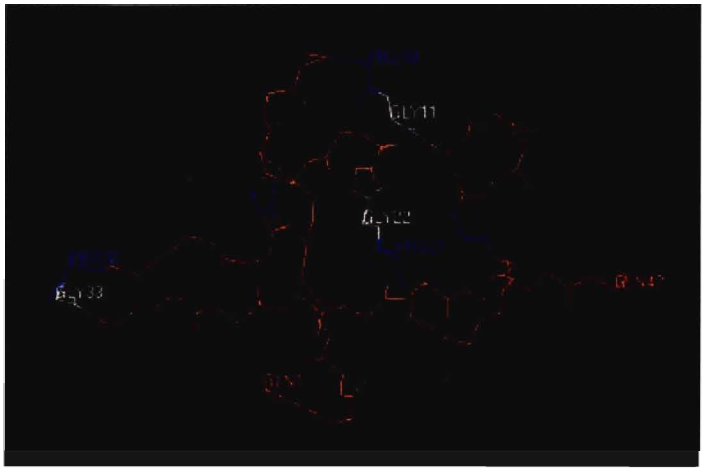


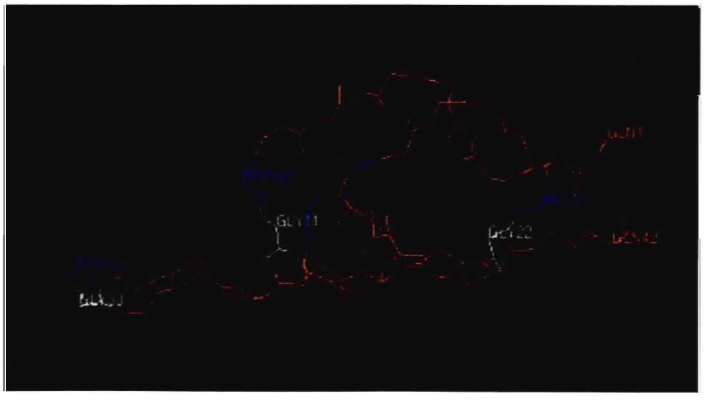












APPENDIX B: Background on CHARMm Force Field

The CHARMm force field is a force field used in the simulation of macromolecules. It represents macromolecules as a series of beads (representing atoms) connected by springs (representing bonds), and is computed based on the following potential energy function:

$$E_{bonded} = E_{bond} + E_{angle} + E_{dihedral} + E_{improper} \quad \text{(B1)}$$

$$E_{non\text{-}bonded} = E_{VDW} + E_{electrostatic} \quad \text{(B2)}$$

$$E_{total} = E_{bonded} + E_{non\text{-}bonded} \quad \text{(B3)}$$

The six energy terms E_{bond}, E_{angle}, $E_{dihedral}$, $E_{improper}$, E_{VDW} and $E_{electrostatic}$ are the total bond-stretching, bond-angle-bending, dihedral (bond rotation), improper dihedral, van der Waals and electrostatic energies, respectively. They are computed as follows:

$$E_{bond} = \sum_{i=1}^{N_{bonds}} k_{b,i}(b_i - b_{0,i})^2 \quad \text{(B4)}$$

describes bond-stretching potentials as positive harmonic functions centered around equilibrium bond lengths. Here, b_i is the current length of a given bond within the molecule; $b_{0,i}$ is the equilibrium length of that type of bond; $k_{b,i}$ is a force constant used to constrain the bond near its equilibrium length; and the sum is computed over all bonds within the molecule.

$$E_{angle} = \sum_{i=1}^{N_{angles}} k_{\theta,i}(\theta_i - \theta_{0,i})^2 \quad \text{(B5)}$$

describes bond-angle-bending potentials as positive harmonic functions centered around equilibrium bond angle values. Here, θ_i is the current value of a given bond angle within the molecule; $\theta_{0,i}$ is the equilibrium value of that type of bond angle; $k_{\theta,i}$ is a force constant used to constrain the bond angle near its equilibrium value; and the sum is computed over all bond angles within the molecule.

$$E_{dihedral} = \sum_{i=1}^{N_{dihedrals}} [k_{\phi 1,i}(\cos(\phi_i - \phi_{1,i})) + k_{\phi 2,i}(\cos(2\phi_i - \phi_{2,i})) + k_{\phi 3,i}(\cos(3\phi_i - \phi_{3,i}))] \quad \text{(B6)}$$

describes dihedral potentials as positive cosine functions, with energy minima centered around sterically-preferred dihedral values. Here, ϕ_i is the current value of a given dihedral angle within the molecule; $\phi_{1,i}$, $\phi_{2,i}$ and $\phi_{3,i}$ are constants corresponding to the preferred values for that type of dihedral angle; $k_{\phi 1,i}$, $k_{\phi 2,i}$ and $k_{\phi 3,i}$ are force constants used to bias the dihedral angle toward its preferred values; and the sum is computed over all dihedral angles within the molecule.

$$E_{improper} = \sum_{i=1}^{N_{imprp-dhd}} k_{\omega,i}(\omega_i - \omega_{0,i})^2 \qquad \text{(B7)}$$

describes improper dihedral potentials as positive harmonic functions centered around equilibrium improper dihedral values. Here, ω_i is the current value of a given improper dihedral within the molecule; $\omega_{0,i}$ is the equilibrium value of that type of improper dihedral; $k_{\omega,i}$ is a force constant used to constrain the improper dihedral near its equilibrium value; and the sum is computed over all improper dihedrals within the molecule. This energy term is imposed in order to maintain stereochemistry at chiral centers, and to maintain the planarity of any planar functional groups (such as aromatic rings).

$$E_{VDW} = \sum_{i}^{N_{atoms}} \sum_{j>i}^{N_{atoms}} 4\varepsilon_{ij}\left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}}\right) \qquad \text{(B8)}$$

describes van der Waals potentials as Lennard-Jones functions, with a short-range repulsive character (accounting for steric repulsion) and longer-range attractive character (accounting for dispersion forces). Here, r_{ij} is the Euclidean distance between atoms i and j; σ_{ij} is the van der Waals contact distance for atoms i and j; ε_{ij} is a parameter defining the strength of van der Waals attraction between atoms i and j at their optimum distance from one another (relative to infinite atom separation); and the sum is computed over all atom pairs within the molecule.

$$E_{electrostatic} = \sum_{i}^{N_{atoms}} \sum_{j>i}^{N_{atoms}} \left(\frac{q_i q_j}{D r_{ij}}\right) \qquad \text{(B9)}$$

describes electrostatic potentials as Coulombic functions, which are purely attractive or repulsive depending on whether the atomic charges in question have opposite or equal signs (respectively). Here, r_{ij} is the Euclidean distance between atoms i and j; q_i and q_j are the partial charges present on atoms i and j; and D is a dielectric constant selected to represent desired solution conditions. As with the van der Waals energy term, the sum is computed over all atom pairs within the molecule.

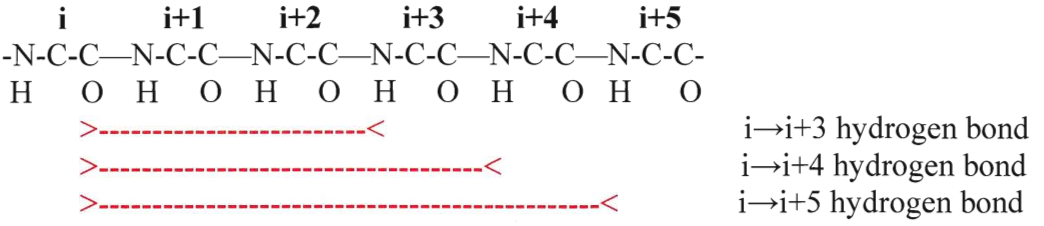
Reference:

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D. and Karplus, M. 1998. "All-atom empirical potential for molecular modeling and dynamics studies of proteins." *Journal of Physical Chemistry B* **102**: 3586-3616.

APPENDIX C: Background on Dictionary of Protein Secondary Structure (DSSP) Software

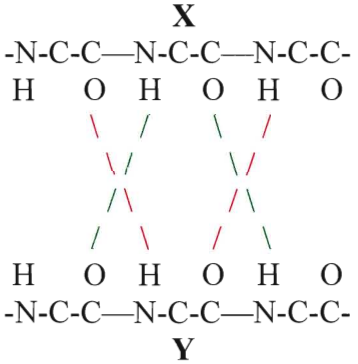
The Dictionary of Protein Secondary Structure (DSSP) software package identifies local secondary structure types adopted by amino acid residues in proteins/peptides, based on a combination of backbone hydrogen bond and geometry criteria. It examines seven possible structure types – 3_{10}-helix, α-helix, π-helix, hydrogen-bonded β-structure, hydrogen-bonded turns, non-hydrogen-bonded bends, and non-hydrogen-bonded loop/irregular structure (which includes PPII helix structure) – and the occurrence of these structure types is identified as follows.

Hydrogen bonds between amide groups of the protein/peptide backbone are first identified. Then, the occurrence of the five hydrogen-bonded structure types is assigned based on configurations of hydrogen bonds that are present within the protein/peptide structure. Part of this process involves identification of hydrogen bonds between the backbone carbonyl oxygen of residue "i" and the backbone nitrogen of residue "i+3", "i+4" or "i+5":

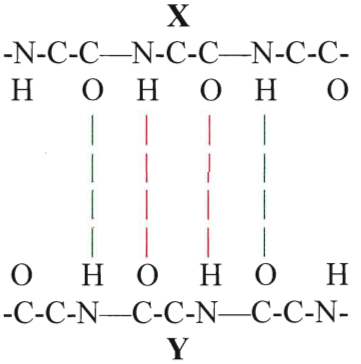


If two or more consecutive i→i+3, i→i+4 or i→i+5 hydrogen bonds are present, then the implicated amino acid residues are identified as forming 3_{10}-helix, α-helix or π-helix structure, respectively. If only one such hydrogen bond is present, however, then the implicated residues are identified as forming a hydrogen-bonded turn.

The other type of hydrogen bonding that is examined involves pairs of hydrogen bonds consistent with hydrogen bond configurations found in parallel or antiparallel β-sheet structure:



parallel (/ and \, or \ and /)

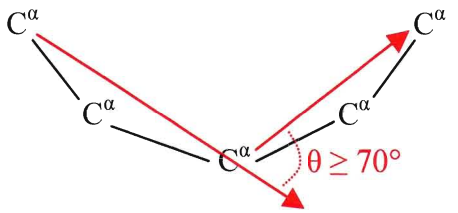


antiparallel (| or |)

If such hydrogen bonds are identified, then residues "X" and "Y" are identified as participating in hydrogen-bonded β-structure. [Notice that two possible hydrogen bond configurations (shown in red and green, respectively) are allowed for each case, due to the natural alternation of spacial orientation exhibited by successive residues in β-strands.]

In the event that a residue is identified as potentially belonging to more than one of the five hydrogen-bonded structure types, then the residue's structure type is assigned based on the following order of structure priorities: α-helix > hydrogen-bonded β-structure > 3_{10}-helix > π-helix > hydrogen-bonded turn.

Finally, residues not exhibiting one of the five hydrogen-bonded structure types are assessed for the formation of non-hydrogen-bonded bends. A non-hydrogen-bonded bend is identified at residue i if the inter-α-carbon vectors from residue i-2 to residue i, and from residue i to residue i+2, are at an angle of at least 70° to one another:



If a residue does not meet this criterion either, then the residue is identified as forming non-hydrogen-bonded loop/irregular structure.

Reference:

Kabsch, W. and Sander, C. 1983. "Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features." *Biopolymers* **22**: 2577-2637.

APPENDIX D: Background on Jonckheere k-Sample Test

The Jonckheere k-sample test is a statistical test that permits simultaneous comparison of two or more samples of data, testing the null hypothesis that the samples are all equal in value to one another. The test compares samples based on rank order of the individual data values from each sample, and a single test statistic is computed for subsequent comparison with an appropriate critical value. The logic involved in the test is as follows.

Let $(X_{1,1}, X_{1,2}, \ldots, X_{1,m1}), \ldots, (X_{i,1}, \ldots, X_{i,\alpha i}, \ldots, X_{i,mi}), \ldots, (X_{k,1}, \ldots, X_{k,mk})$ be k samples of sizes $m_1, m_2, \ldots, m_k$, randomly drawn from populations $F_1(X), \ldots, F_i(X), \ldots, F_k(X)$. Then, for all X, let

$$p_{i,\alpha i,j,\alpha j} = \begin{cases} 1 & if \quad X_{i,\alpha i} < X_{j,\alpha j} \\ 0 & if \quad X_{i,\alpha i} > X_{j,\alpha j} \end{cases} \tag{D1}$$

where $i = 1, \ldots, k\text{-}1$; $j = i+1$; $\alpha_i = 1, \ldots, m_i$; and $\alpha_j = 1, \ldots, m_j$. Then,

$$p_{i,j} = \sum_{\alpha_i=1}^{m_i} \sum_{\alpha_j=1}^{m_j} p_{i,\alpha i,j,\alpha j} \tag{D2}$$

$$S = 2\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} p_{i,j} - \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} m_i m_j \tag{D3}$$

For large samples, once the test statistic S is computed, a corresponding standardized normal statistic Z can be calculated for comparison against standard normal distribution tables (normal approximation):

$$Z = \frac{S}{\sqrt{\mu_2}} = \frac{S}{\sqrt{\frac{1}{18}\left\{n^2(2n+3) - \sum_{r=1}^{k} m_r{}^2(2m_r+3)\right\}}} \tag{D4}$$

where n is the sum of all m_r values. Alternatively, for small samples, a Student's t statistic can be calculated for comparison against standard Student's t tables (Student's t approximation, with ν degrees of freedom):

$$t = S\sqrt{\frac{\nu}{(\nu+1)\mu_2 - S^2}} \tag{D5}$$

where

$$\nu = \frac{-3(2+\gamma_2)}{\gamma_2} \tag{D6}$$

$$\gamma_2 = -\frac{36}{25}\frac{\left\{n^3(6n^2+15n+10)-\sum_{r=1}^{k} m_r{}^3(6m_r{}^2+15m_r+10)\right\}}{\left\{n^2(2n+3)-\sum_{r=1}^{k} m_r{}^2(2m_r+3)\right\}^2}$$

and μ_2 and n are defined in the same manner as with the normal approximation.

Reference:

Jonckheere, A.R. 1954. "A distribution-free k-sample test against ordered alternatives." *Biometrika* **41**: 133-145.