Modeling Protein Homopolymeric Repeats: Possible Polyglutamine Structural Motifs For Huntington's Disease

Richard H. Lathrop(1), Malcolm Casale(1), Douglas J. Tobias(2), J. Lawrence Marsh(3), Leslie M. Thompson(4)

(1) Information & Computer Science, (2) Chemistry, (3) Developmental & Cell Biology, (4) Biological Chemistry
University of California, Irvine, CA 92717 USA
{ rickl | mcasale | dtobias | jlmarsh | lmthomps }@uci.edu

Abstract

We describe a prototype system (Poly-X) for assisting an expert user in modeling protein repeats. Poly-X reduces the large number of degrees of freedom required to specify a protein motif in complete atomic detail. The result is a small number of parameters that are easily understood by, and under the direct control of, a domain expert. The system was applied to the polyglutamine (poly-Q) repeat in the first exon of *huntingtin*, the gene implicated in Huntington's disease. We present four poly-Q structural motifs: two poly-Q β-sheet motifs (parallel and antiparallel) that constitute plausible alternatives to a similar previously published poly-Q β-sheet motif, and two novel poly-Q helix motifs (α-helix and π-helix). To our knowledge, helical forms of polyglutamine have not been proposed before. The motifs suggest that there may be several plausible aggregation structures for the intranuclear inclusion bodies which have been found in diseased neurons, and may help in the effort to understand the structural basis for Huntington's disease.

Introduction

Several normal and abnormal proteins contain a short sequence motif of one or more amino acid residues repeated several times in succession, called a protein sequence repeat. For example, expanded polyglutamine (poly-Q) repeats are known to cause at least eight progressive autosomal dominant neurodegenerative diseases, including Huntington's disease and several forms of spinocerebellar ataxia (Ross 1997; Warren 1996; Warren & Ashley 1995). Perutz (Perutz 1996) reviews the molecular aspects of glutamine repeats and inherited neurodegenerative diseases. Repeats also are involved in certain important structural proteins such as silk and collagen, which involve short repeating approximate motifs of a few amino acid residues. We have developed a prototype system for assisting an expert user engaged in molecular modeling of repeat structures at atomic detail. The computational task is to enable an expert to explore alternate conformations rapidly, by quickly producing a reasonable trial conformation that falls into the desired energy minima under conventional force fields and molecular modeling software. This facilitates building symmetrical, repetitive structures in the repertoire of current modeling packages. The goal in this paper is to assist in exploring poly-Q structures that may have relevance to Huntington's disease and related syndromes.

Figure 1 shows the N-terminal end of the *huntingtin* protein sequence (HDCR Group 1993). The region implicated in Huntington's disease is the long poly-Q repeat near the N-terminal end, beginning at residue number 18. The length of the poly-Q repeat determines the presence and progression of the disease. Poly-Q repeats of 10 to 34 residues occur in normal individuals, while repeats of 37 to 100 residues occur in Huntington's disease patients (HDCR Group 1993). Above 37 residues, increasing poly-Q repeat length correlates with an increasing rate of disease progression and a decreasing age of onset (Penny *et al.* 1997). The gene has other repeat regions as well; note the two proline repeats (poly-P) shortly after the poly-Q repeat. The *huntingtin* protein is expressed ubiquitously throughout the body, but only in afflicted nerve cells does it cause problems leading to severe neurodegeneration. There, *huntingtin* aggregates into intranuclear inclusion bodies (Davies *et al.* 1997; DiFiglia *et al.* 1997). Similar aggregation is seen in other neurodegenerative diseases, for example β-amyloid plaque formation in Alzheimer's disease. It is thought that β-sheet formation may play a role in this process and possibly others.

Figure 2 shows consensus secondary structure predictions (Geourjon & Deleage 1994; 1995) for the disease regions in both normal (Figure 2.A) and disease-bearing (Figure 2.B) sequences. Glutamine normally favors helix formation, and helix is favored in Figure 2. However, β-sheet is mentioned as a possible secondary structure (Levin method (Levin & others 1986)), and the Levin β-sheet prediction increases with increasing length of the poly-Q repeat. Perutz *et al.* (Perutz *et al.* 1993) proposed the poly-Q β-sheet structure shown in Figure 3 as a polar zipper that could form a stable lattice

through intermolecular hydrogen bonds. In a landmark study, Perutz *et al.* (Perutz *et al.* 1994) obtained circular dichroism spectra from poly-Q fibrils showing that poly-Q indeed can form β-sheets. Stott *et al.* (Stott *et al.* 1995) showed that inclusion of glutamine repeats makes proteins oligomerize, and indicated that the glutamine repeats in dimers and trimers formed β-sheets.

Neither the function nor the structure of the *huntingtin* protein are known, and it has no appreciable sequence similarity to any other known sequence. The effects of the poly-Q repeat on protein structure and function are unclear. Whether the Huntington's disease pathology is due to specific effects mediated by the *huntingtin* protein containing the polyglutamine tracts, or whether the pathology is a consequence of the glutamines per se, is unclear. Clues to the structure are obviously important because they may lead to better understanding of the disease process and ultimately to a treatment or a cure. Here we describe a method to develop structures which are valuable tools for generating testable hypotheses about the molecular basis for the disease and avenues of approach for a treatment.

Methods

Protein repeats represent an opportunity for molecular modeling because of the possibility of reducing the degrees of freedom that must be considered explicitly. Poly-X exploits hierarchy and symmetry to reduce the many parameters that specify a protein motif in atomic detail down to an understandable and manageable few, using interaction with the user for choosing and enforcing symmetry. The motif is decomposed into a number of discrete ungapped chains, each chain is decomposed into residues, each residue is represented by a residue proxy, and each residue proxy is represented by backbone and sidechain atoms. Figure 4 illustrates this.

The motif is placed by establishing a global coordinate system and then iterating over each chain. Each chain is placed by iterating over each chain residue. At each residue, a coordinate transform is constructed from the proxy residue local coordinate system into a coordinate system attached to the chain residue. The transform is based on orthonormal coordinate systems constructed from the backbone Cα, N, and C$'$ (carbonyl) atoms. It is used to map the proxy residue sidechain atoms onto the chain. For efficiency this is implemented by pre-computing and caching the proxy residue as full-atom rotamers. Pseudo-code is:

```
FOR I = 1 TO number-of-chains
  FOR J = 1 TO number-of-residues-in-chain(I)
    P = get-cached-proxy-residue(I,J)
    R = get-chain-residue(I,J)
    T = get-coordinate-transform(P,R)
    FOR K = 1 to number-of-atoms-in-residue(P)
      write-atom(T,P,K)
    END K
  END J
END I
```

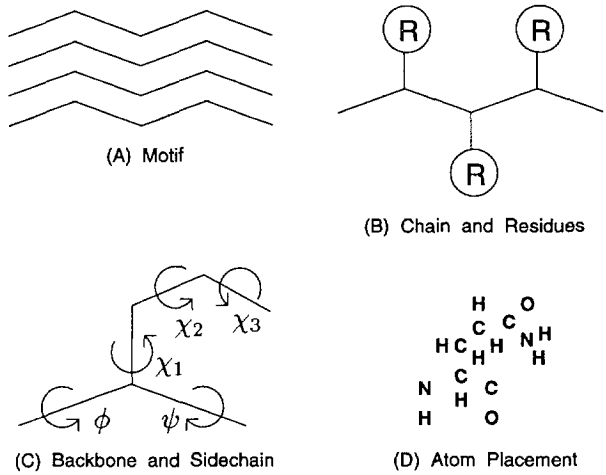

Figure 4: The problem decomposition by hierarchy and symmetry. (A) A motif is decomposed into discrete ungapped chains. (B) A chain is decomposed into residues. (C) Each residue is decomposed into backbone and sidechain. (D) Atoms are placed where specified by their chain, residue, and proxy.

Poly-X mediates between an expert user and a modeling package as shown in Figure 5. First the expert specifies initial parameters for the backbone and sidechains. Alternatively, the expert indicates parameters which should be copied from existing structures, e.g., from a database, or from previous Poly-X or molecular modeling runs. Then the system produces a candidate structure that embodies this. It is dumped as a formatted structure file, input to a standard molecular modeling package, and energy-minimized or otherwise processed. Then, a second formatted structure format file is written from the modeling package and read by the system. The expert interacts with the system to select or modify desired degrees of freedom. The process repeats until the expert is satisfied with the structure.

In this initial study, the structures shown below first were generated based on standard secondary structural motifs. These were minimized using the Amber (Weiner *et al.* 1984) force field. Then candidate residues were chosen as interesting structures based on their hydrogen bonding potential, and designated by the expert as proxy residues. Proxy residue conformations were propagated to the rest of the motif as described above. The resulting motif was reminimized, and the process repeated until the expert was satisfied with the structure. The final step was always a final energy minimization. Consequently, the resulting structures represent a local energy minimum to which they were guided by the expert.

```
>gi|1170192|sp|P42858|HD_HUMAN HUNTINGTIN (HUNTINGTON'S DISEASE PROTEIN)
MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPQLPQPPPQAQPLLPQPQPP
PPPPPPPPGPAVAEEPLHRPKKELSATKKDRVNHCLTICENIVAQSVRNSPEFQKLLGIAMELFLLCSDD...
```

Figure 1: The first 160 amino acid residues of normal human *huntingtin,* or Huntington's disease protein (there are 3,594 total residues).

```
>From : hd normal
                1        10        20        30        40        50        60
                |        |         |         |         |         |         |
hd normal       MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQPPPPPPPPPPPQLPQPPPQA
Gibrat method   HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCC
Levin method    HHHHHHHHHHHHHHHHHCCHCCSCEEEHHHHHHHHHHSSCCSSSCCCCCCCCCCCCCCSC
DPM method      CCHHHHHHHHHHCECCEHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCC
SOPMA predict   HHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Consensus       HHHHHHHHHHHHHHHHHHHHHHCCHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCC

hd normal       QPLLPQPQPPPPPPPPPPGPAVAEEPLHRPKKELSATKKDRVNHCLTICENIVAQSVRNS
Gibrat method   CCCCCCCCCCCCCCCCCCCCCHHHHHHCHHHHHHHHHHHCECCCCCEECCEECEECCCC
Levin method    CCECCCCCCCCSCCCCCCCCCCCHCCCCCCTHCCCHHCHSHHHHCCEEHHHHHHHCCCCC
DPM method      CCCCCCCCCCCCCCCCCCCCCHHHHCCCCCCCCHCHCTCCCECCEEEEEEEEEEEEECCT
SOPMA predict   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHCCCCCCCEEEECCCHHHHHHHCCC
Consensus       CCCCCCCCCCCCCCCCCCCCCCCHCCCCCCCHCHCHCCCCCECCCCEEECHHHHHECCCC
```

(A) Normal *huntingtin* protein sequence.

```
>From : hd mutant
                1        10        20        30        40        50        60
                |        |         |         |         |         |         |
hd mutant       MATLEKLMKAFESLKSFQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
Gibrat method   HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
Levin method    HHHHHHHHHHHHHHHHHCCHCCSCEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
DPM method      CCHHHHHHHHHHCECCEHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
SOPMA predict   HHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Consensus       HHHHHHHHHHHHHHHHHHHHHHCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

hd mutant       QQQQQQQQQQQQPPPPPPPPPPPQLPQPPPQAQPLLPQPQPPPPPPPPPPGPAVAEEPLH
Gibrat method   HHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHH
Levin method    HHHHHHHHHSSCCSSSCCCCCCCCCCCCCCSCCCECCCCCCCCSCCCCCCCCCCHCCCC
DPM method      HHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHCCC
SOPMA predict   CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Consensus       HHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHCCCC
```

(B) Disease-bearing *huntingtin* protein sequence.

Figure 2: The first 160 amino acid residues of consensus secondary sequence predictions for *huntingtin.*
(A) normal sequence. (B) disease-bearing sequence.
Key: H = helix, E = extended (β-strand/β-sheet), C = coil, T = turn, S = bend.

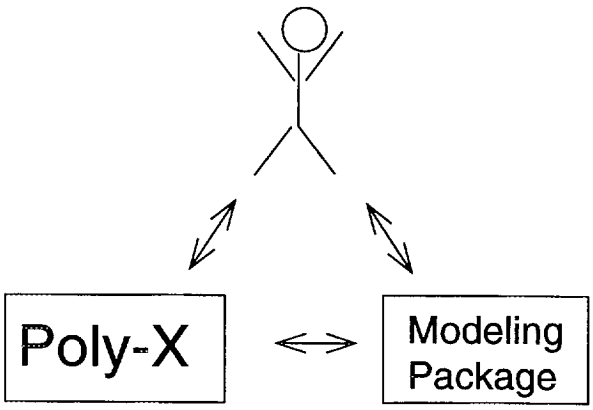

Figure 5: The interaction between the expert, Poly-X, and an atomic modeling package.

Results: Novel Proposed Protein Structural Motifs

We used the process described above to find novel proposed protein structural motifs for the poly-Q repeat region of the *huntingtin* protein. Figure 6 gives atomic coordinates.

Figure 7 shows a parallel β-sheet and Figure 8 shows an anti-parallel β-sheet. These constitute plausible alternatives to a similar β-sheet structural motif previously proposed by Perutz *et al.* (Perutz *et al.* 1994) and shown in Figure 3.

Two of the novel structural motifs are helical. Figure 9 shows an α-helix and Figure 10 shows a π-helix. To our knowledge, helical forms of polyglutamine have not been proposed before.

Discussion

One potentially important result is that there seem to be several plausible aggregate structures. In addition to major differences from alternate backbone secondary structure conformations, minor differences arise from alternate sidechain conformations. It could be that all of them play a role in the intranuclear inclusion bodies (Davies *et al.* 1997; DiFiglia *et al.* 1997). X-ray data from Perutz *et al.* (Perutz *et al.* 1994) suggests that the sheet-type arrangements appear to be favored *in vitro* by short solubilized model peptides in aqueous solution, while in the longer diseased proteins *in vivo,* a variety of patterns may co-exist. The different motifs all may have roughly the same stability per residue (to within small variations on the order of thermal energy), in which case the possibility of multiple aggregation motifs could be favored entropically (i.e., compared to a single motif).

One important question concerns the stability of the proposed structures in solvated and hydrophobic environments. Water would supply many of the exposed hydrogen bonds, disrupting their ordered surface arrangement. Indeed, preliminary modeling with a continuum aqueous solvation model indicates that this occurs. The sidechain hydrogen bonds appear to be disrupted by water, while the backbone hydrogen bonds, and consequently the secondary structure, appear to be stable. On the other hand, inclusion bodies formed by aggregation appear in afflicted neurons. The interior of aggregated inclusion bodies should be protected from solvent, and so the surface poly-Q hydrogen bonds would be protected from disruption. Preliminary modeling with water excluded indicates that the ordered bonds are remarkably stable. This might help explain the persistence of inclusion bodies once formed. It might offer a point of attack for a drug to penetrate the aggregation, disrupt hydrogen bonds spaced at precise poly-Q intervals, and so dissolve the inclusion bodies or open them to proteolytic attack.

Another important question concerns how differences in the length of the poly-Q repeat can trigger Huntington's disease and the formation of inclusion bodies. Perutz (Perutz 1996) suggests that thermodynamic considerations of loss of glutamine translational and rotational entropy, balanced against gain of entropy from liberated waters, might provide such a critical length effect. On the other hand, Perutz (Perutz 1997) discusses "chameleon" sequences that can adopt either α-helix or β-sheet folds depending on context, and suggests that this may be a mechanism for enzyme polymerization. The secondary structure predictions in figure 2 also suggest a helix to sheet transition. Figure 11 shows a hypothetical helix to sheet pathway that incorporates a poly-Q length trigger based on steric considerations. Helical poly-Q structures could account for variable-length poly-Q repeats up to a limit fixed by the spatial separation of the two fixed points (left-hand side of figure). Longer poly-Q repeats could be absorbed by forming more turns of the helix. Beyond that limit the helical form could not absorb the entire poly-Q repeat, and so would be inaccessible due to steric clashes. Intermolecular β-sheet formation would be enabled (right-hand side of figure) because the β-sheet is energetically favorable relative to coil and the alternative helix structure is no longer available. Other plausible possibilities are easily imaginable; e.g., stochastic lateral shear forces or the helix dipole moment might reach a critical stability threshold. We do not propose that any of these are the mechanism that occurs in nature. We do suggest that the availability of appropriate modeling interaction tools is important in studying and analyzing the different hypothetical possibilities.

Future work will include extending Poly-X to model aggregates, such as might arise through β-sheet lattice formation, or the transitions sketched in Figure 11. Extending Poly-X to assemble more than one chain of residues, especially where sidechain conformations may differ or the backbone may deviate from standard secondary structure geometry, will require the ability for the expert user to restrict specified parameters to specific subsets of the structure. The proxy residue must

be generalized to range over several alternate rotamers and local environments, under control of the expert. The intrinsic problem of optimizing the relative orientation of the modeled chains in order to analyze and optimize inter-chain interaction probably will require a local search of relative orientations and sidechain conformations. For this it might be worth considering simpler criteria such as hydrogen-bond geometry between chains, accessibility, excluded volume, and the like, rather than only the modeled force field.

Future detailed studies must be done of the energy and stability of the proposed structures in various environments, including estimated energies of the different structures and comparisons involving: (1) energies of random coil in several conformations (e.g., quench dynamics randomly from high temperatures and then minimize), (2) energies of sheet and helix with disordered sidechains, and (3) energies in solvated vs. hydrophobic environments. Wet-lab experiments are planned to measure certain observable parameters implied by the proposed structures in an attempt to determine whether they exist *in vivo*. Although much remains to be done, the results presented above clearly demonstrate the utility of the approach taken by Poly-X.

Summary

We have described a system for assisting an expert engaged in the task of modeling protein sequences with repeated motifs. A large number of degrees of freedom are required to specify a protein motif in complete atomic detail. Poly-X reduces these to a small number of parameters that are easily understood by, and under the direct control of, a domain expert. The system was applied to the poly-Q repeat in the first exon of *huntingtin*, the gene implicated in Huntington's disease. Poly-X was used to describe four poly-Q structural motifs: two poly-Q β-sheet motifs (parallel and anti-parallel) that constitute plausible alternatives to a previously published poly-Q β-sheet motif (Perutz *et al.* 1994), and two novel poly-Q helices (α-helix and π-helix). To our knowledge, helical forms of poly-Q have not been proposed before. The structures may prove to be relevant to Huntington's disease because they may help to understand the formation of inclusion bodies and how to disrupt or dissolve them.

Acknowledgments

Terry LePage rendered valuable assistance in modeling. We thank Nancy Wexler, Ethan Signer, Allan Tobain, James Nowick, Keith Dunker, and Jim Fallon for useful discussion, and two anonymous referees whose helpful comments improved the paper.

Funding has been provided by the National Science Foundation under grant IRI-9624739.

Poly-X is available from University/Industry Research and Technology, 380 University Tower, University of California, Irvine, 92717 USA. The atomic coordinate files are available electronically from the authors.

References

Davies, S.; Turmaine, M.; Cozens, B.; DiFiglia, M.; Sharp, A.; Ross, C.; Scherzinger, E.; Wanker, E.; Mangiarini, L.; and Bates, G. 1997. Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the hd mutation. *Cell* 90:537–548.

DiFiglia, M.; Sapp, E.; Chase, K.; Davies, S.; Bates, G.; Vonsattel, J.; and Aronin, N. 1997. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* 277:1990–1993.

Geourjon, C., and Deleage, G. 1994. Sopm: A self optimised prediction method for protein secondary structure prediction. *Protein Engineering* 7:157–164.

Geourjon, C., and Deleage, G. 1995. Sopma: Significant improvements in protein secondary structure prediction by prediction from multiple alignments. *Comput. Applic. Biosci.* 11:681–684.

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.

Levin, et al. 1986. *FEBS Letters* 205:303–308.

Penny, J. J.; Vonsattel, J.-P.; MacDonald, M.; Gusella, J.; and Myers, R. 1997. Cag repeat number governs the development rate of pathology in huntington's disease. *Ann. Neurol.* 41:689–692.

Perutz, M.; Staden, R.; Moens, L.; and DeBaere, I. 1993. Polar zippers. *Curr. Biol.* 3:249–253.

Perutz, M.; Johnson, T.; Suzuki, M.; and Finch, J. 1994. Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. USA* 91:5355.

Perutz, M. 1996. Glutamine repeats and inherited neurodegenerative diseases: molecular aspects. *Curr. Opinion in Struct. Biol.* 6:848–858.

Perutz, M. 1997. Mutations make enzyme polymerize. *Nature* 385:773–775.

Ross, C. 1997. Intranuclear neuronal inclusions: A common pathogenic mechanism for glutamine-repeat neurodegenerative diseases? *Neuron* 19:1147–1150.

Stott, K.; Blackburn, J.; Butler, P.; and Perutz, M. 1995. Incorporation of glutamine repeats makes protein oligomerize: Implications for neurodegenerative diseases. *Proc. Natl. Acad. Sci. USA* 92:6509–6513.

Warren, S., and Ashley, C. 1995. Triplet repeat expansion mutations: the example of fragile x syndrome. *Ann. Rev. Neurosci.* 18:77–79.

Warren, S. 1996. The expanding world of trinucleotide repeats. *Science* 271:1374–1375.

Weiner, S.; Kollman, P.; Case, D.; Singh, U.; Ghio, C.; Alagona, G.; Profeta, S.; and Weiner, P. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

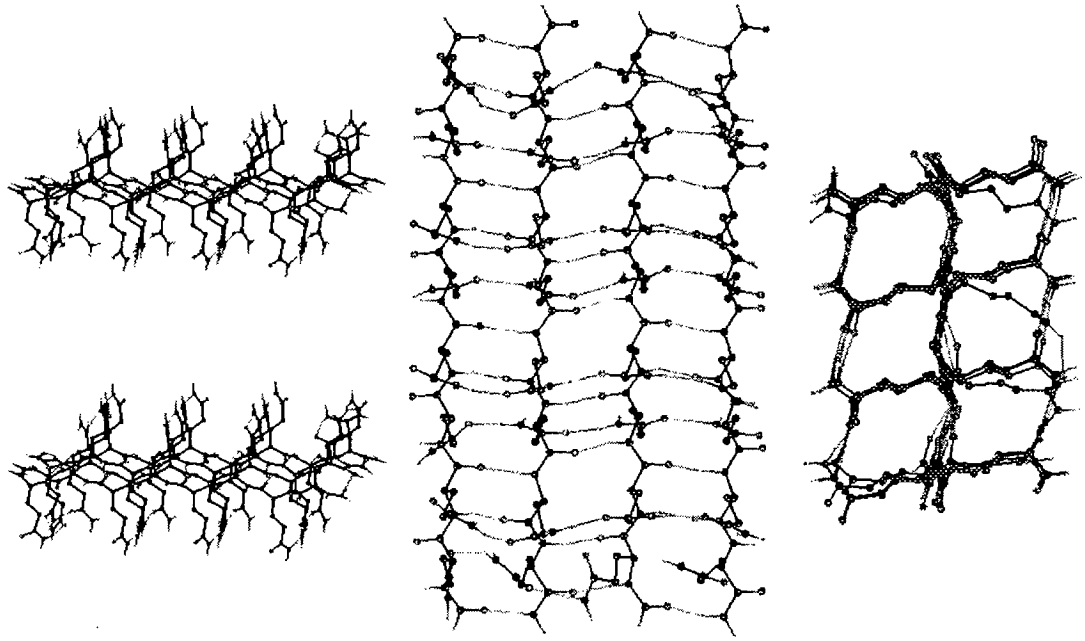

Figure 3: The poly-Q anti-parallel β-sheet structure ("polar zipper") proposed by Perutz *et al.* (Perutz *et al.* 1994) (generated using Poly-X following Perutz *et al.* (Perutz *et al.* 1994)). Stereogram, top view, end view. Hydrogen bonds are dashed. Non-polar hydrogens are omitted for viewing clarity.

	Perutz (Fig. 3)			Parallel (Fig. 7)			Antiparallel (Fig. 8)			α-helix (Fig. 9)			π-helix (Fig. 10)		
Atom	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z
N	-4.61	-4.01	2.07	1.87	-2.63	-0.14	-9.37	-5.84	7.34	2.28	-0.87	10.26	5.20	4.99	8.07
Cα	-3.48	-3.12	2.23	2.98	-1.71	-0.15	-8.63	-4.60	7.22	1.42	-1.99	10.67	4.92	5.13	9.49
C	-2.33	-3.72	1.43	4.08	-2.33	-1.00	-7.28	-4.98	6.63	1.96	-3.35	10.23	3.55	4.52	9.84
O	-2.14	-4.93	1.47	4.25	-3.55	-0.97	-6.70	-5.99	7.04	1.88	-4.31	10.98	3.05	4.69	10.95
Cβ	-3.09	-2.98	3.71	3.50	-1.46	1.27	-8.52	-3.89	8.57	-0.03	-1.82	10.20	5.04	6.60	9.92
Cγ	-4.27	-2.62	4.61	2.40	-1.07	2.26	-9.91	-3.59	9.15	-0.91	-3.00	10.67	5.53	6.64	11.37
Cδ	-3.81	-2.18	5.99	2.99	-0.57	3.57	-9.90	-2.48	10.19	-2.39	-2.86	10.36	5.64	8.04	11.94
Oϵ	-3.26	-1.10	6.15	3.17	0.62	3.76	-8.96	-1.68	10.25	-2.87	-1.88	9.82	5.37	9.03	11.28
Nϵ	-4.01	-3.01	7.01	3.32	-1.47	4.48	-10.94	-2.37	10.99	-3.21	-3.92	10.71	6.06	8.13	13.20

Figure 6: Motif proxy residue atomic coordinates, heavy atoms only. The subsequent minimization allows relaxation, adjustment, and distortion.

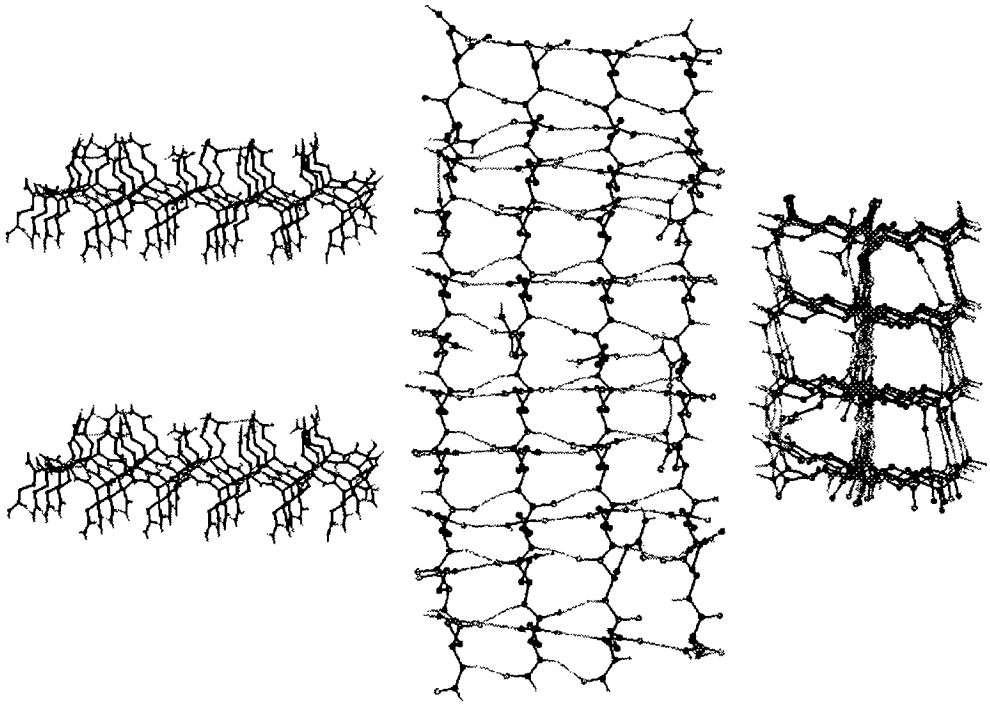

Figure 7: Proposed parallel poly-Q β-sheet structure. Stereogram, top view, end view. Hydrogen bonds are dashed. Non-polar hydrogens are omitted for viewing clarity. Side-chain conformations are similar to Figure 3, but backbone hydrogen bonds differ.

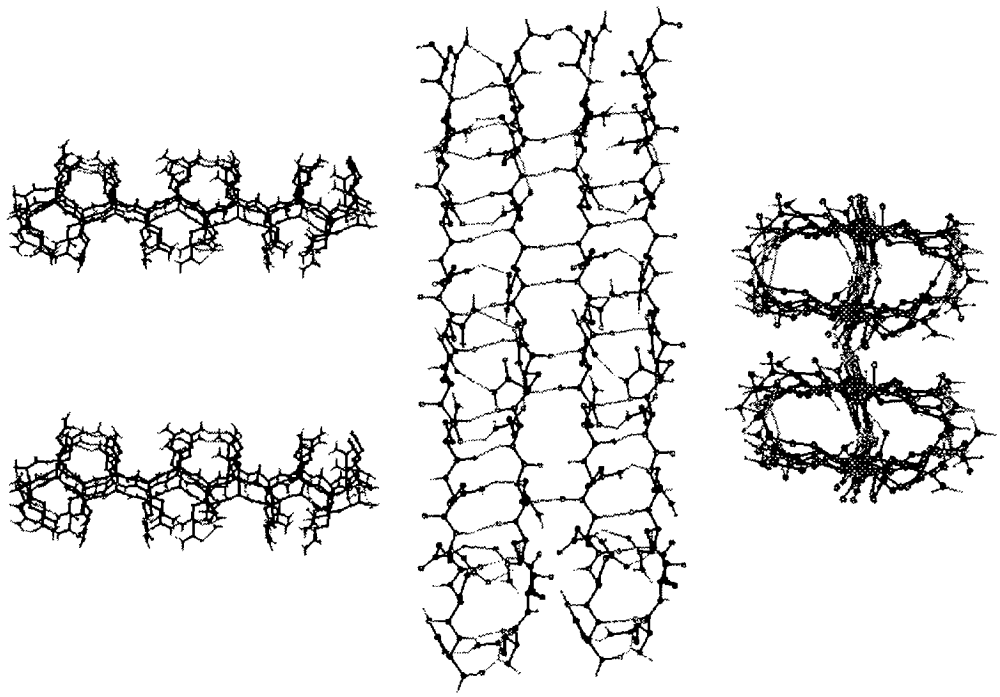

Figure 8: Proposed alternate anti-parallel poly-Q β-sheet structure. Stereogram, top view, end view. Hydrogen bonds are dashed. Non-polar hydrogens are omitted for viewing clarity. Backbone hydrogen bonds are similar to Figure 3, but side-chain conformations differ.

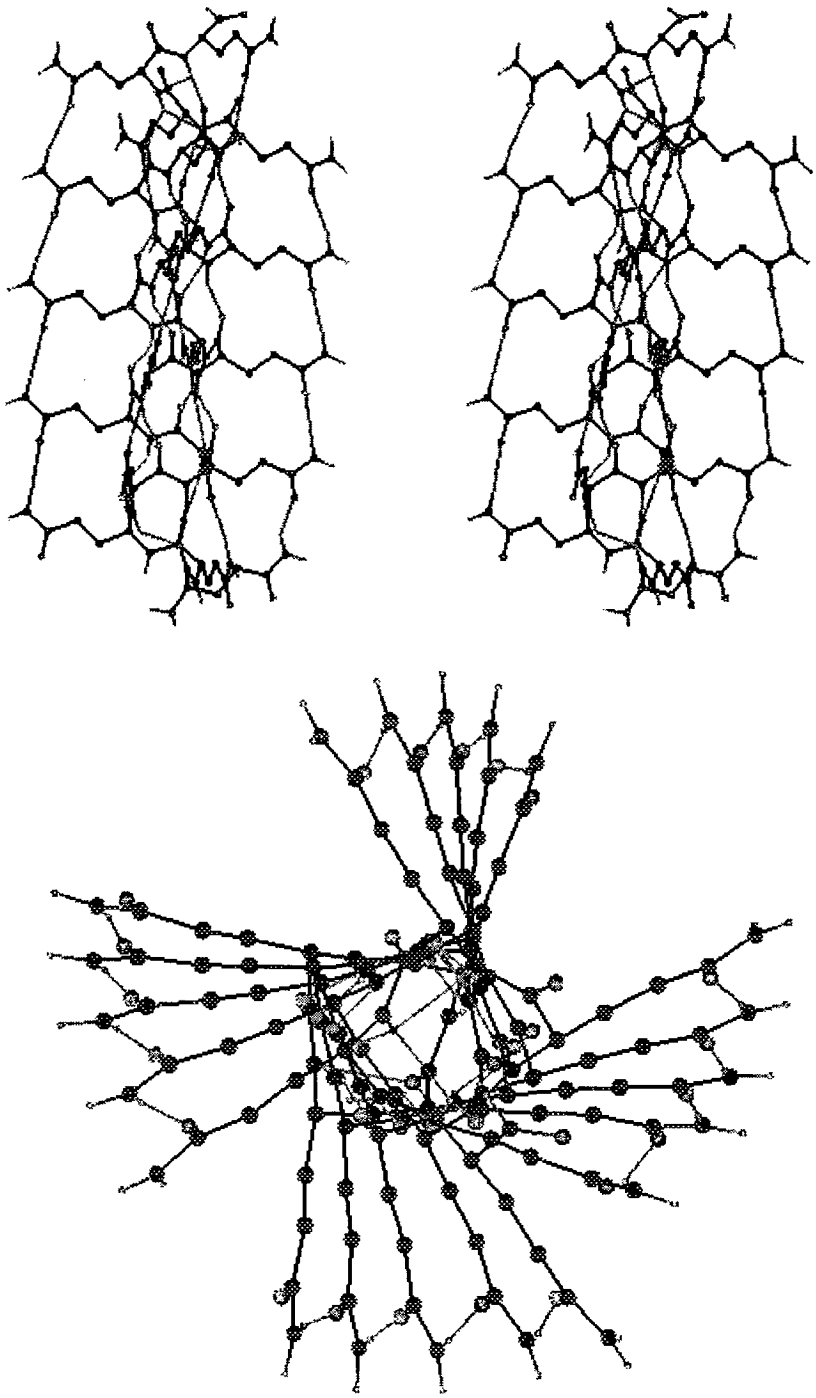

Figure 9: Proposed poly-Q α-helix structure. Stereogram, end view. Hydrogen bonds are dashed. Non-polar hydrogens are omitted for viewing clarity. The sidechain hydrogen bond network rotates in the opposite direction to Figure 10, and the backbone hydrogen bonds differ.

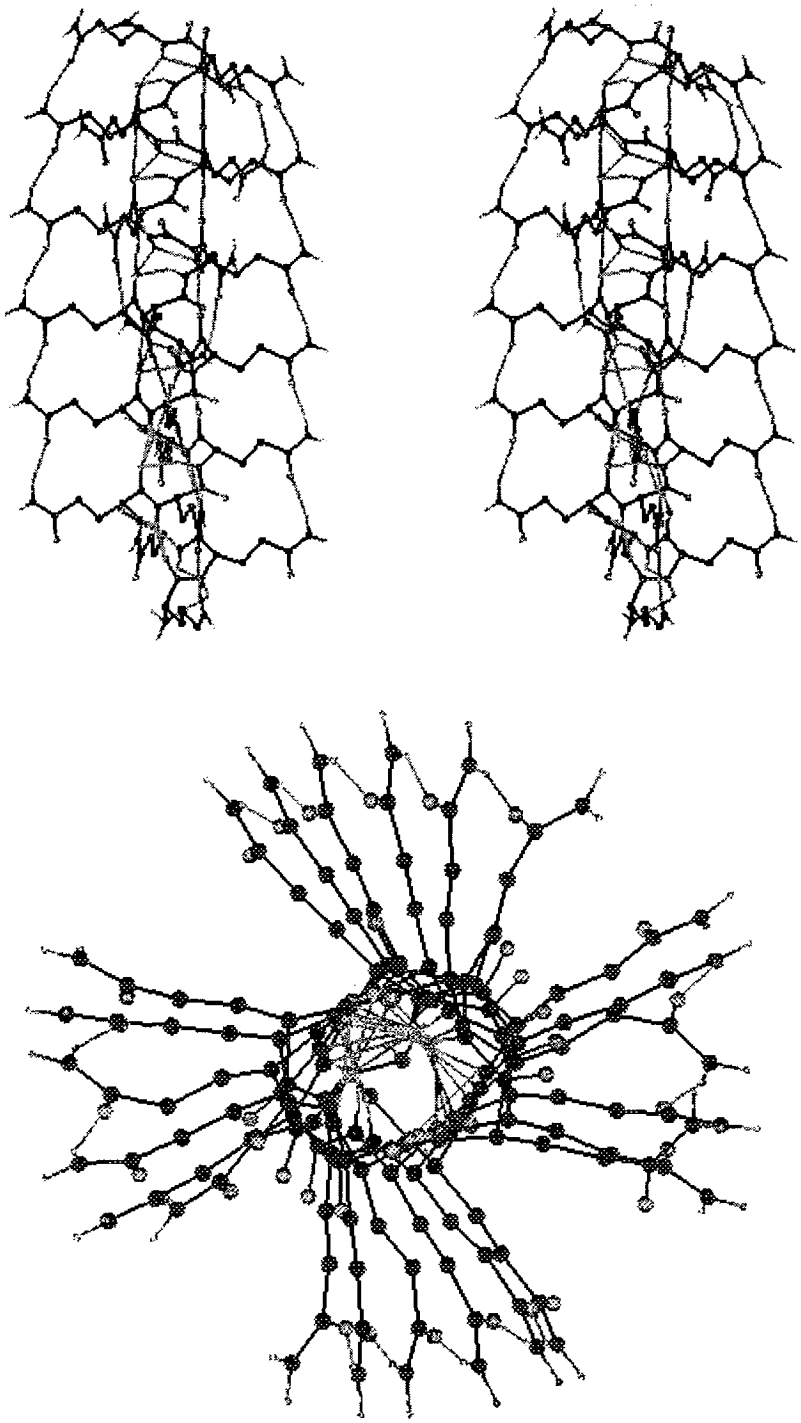

Figure 10: Proposed poly-Q π-helix structure. Stereogram, end view. Hydrogen bonds are dashed. Non-polar hydrogens are omitted for viewing clarity. The sidechain hydrogen bond network rotates in the opposite direction to Figure 9, and the backbone hydrogen bonds differ.

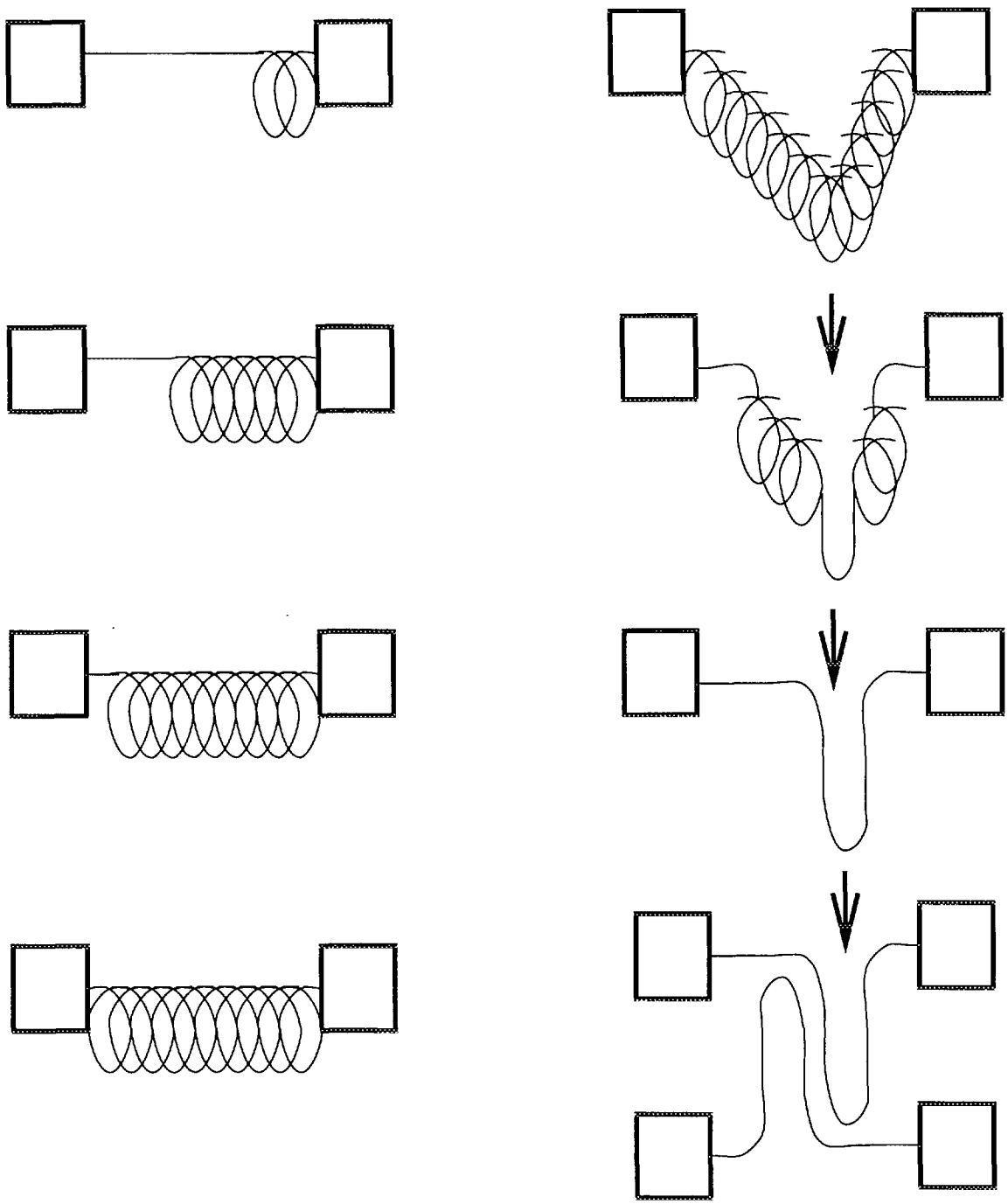

Figure 11: A hypothetical mechanism by which excessive lengths in a poly-Q repeat might trigger β-sheet aggregation. Boxes represent parts of the protein sequence that are assumed to be fixed in the tertiary structure, e.g., anchored in the protein core or pinned by a dimer contact. Thin lines represent the poly-Q polypeptide. Ovals represent helix, parallel thin lines represent β-strands of a β-sheet. The left column represents increasing poly-Q lengths being absorbed by increasing helix turns, up to a limit. The right column represents transitions above that limit. Compare Figure 2.