

## Hydrogen Bonding Cooperativity in polyQ $\beta$ -Sheets from First Principle Calculations

Giulia Rossetti,<sup>†,‡,§</sup> Alessandra Magistrato,<sup>\*,†</sup>  
Annalisa Pastore,<sup>||</sup> and Paolo Carloni<sup>†,‡,§</sup>

Statistical and Biological Physics Sector, International School for Advanced Studies (SISSA-ISAS) and CNR-IOM-DEMOCRITOS National Simulation Center, Trieste, Italy, Via Beirut 2-4, Trieste, Italy, German Research School for Simulation Science, FZ-Juelich and RWTH, Germany, Italian Institute of Technology—SISSA Unit, Via Beirut 2-4, Trieste, Italy, and National Institute for Medical Research, The Ridgeway London, NW71AA, United Kingdom

Received September 09, 2009

**Abstract:** Polyglutamine  $\beta$ -sheet aggregates are associated with the derangement of Huntington's disease. The effect of cooperativity of the H-bond network formed by both backbone and side chain groups is expected to be important for the structure and energetics of the aggregates. So far, no direct description and/or quantification of the effect is yet available. By performing DFT and hybrid DFT/MM simulations of polyglutamine  $\beta$ -sheet structures *in vacuo* and in aqueous solution, we observe that the cooperativity of glutamine side chains affects both the directions perpendicular and parallel to the backbone. This behavior is not usually observed in  $\beta$  sheets and may provide significant extra-stabilization together with explaining some of the unique properties of polyglutamine aggregation.

Huntington's and other neurodegenerative diseases depend on the abnormal expansion of polyglutamine (polyQ) tracts in proteins which form aggregates rich in  $\beta$  sheets associated with neurodegeneration.<sup>1–6</sup> The glutamine side chain is similar to the backbone unit. Thus, polyQ tracts can form particular  $\beta$  strands stabilized by a hydrogen bond (HB) net involving both the backbone and the side chains.<sup>5–7</sup> The presence of a

cooperative effect (CE) on this peculiar HB net may play a role in the misfolding and aggregation of polyQ.<sup>5</sup> The CE in hydrogen bonding is very important for both the structure and the energetics of polypeptide systems.<sup>8</sup>

The extra-structural stability of polyQ aggregates due to the CE is related to the number of HBs formed between the backbone and side chains.<sup>9</sup> Nevertheless, the conclusions so far were achieved by classical molecular dynamics calculations that cannot answer the critical issue of how to deal with electronic polarizability. This can be described by first principle methods, which have in fact already been applied in the study of CE on polypeptides, including polyQ chains.<sup>10–18</sup> However, the crucial role of Q side chain HBs on the CE has not been investigated so far by first principles approaches.

Here, we perform first principles DFT-PBE<sup>19–21</sup> calculations on polyQ peptides of increasing complexity, assembled in parallel  $\beta$  sheets (Table 1), a structure well characterized from biochemical and theoretical studies [CE turns out to be stronger in parallel  $\beta$  sheets (like the systems considered here) than in antiparallel ones<sup>22</sup>].<sup>14,22–24</sup> Our models ( $N \times n$  hereafter) differed from each other for the number of strands ( $N = 1, 2, 3, 4$ ) and/or for the number of Qs in each strand ( $n = 1, 2, 3, 4$ ) [the models were built using the *HyperChem 8.0* program<sup>25</sup>]. They are terminated by the addition of  $-NCH_3$  and  $-OCCH_3$  groups. The resulting 16 models range from 29 to 320 atoms (Table 1, see the footnote for more details on notations). Next, because of the obvious role of solvent and temperature effects on polypeptide conformation,<sup>13</sup> we performed 2 ps of hybrid DFT/MM molecular dynamics calculations on a large system, a  $\beta$ -helix nanotube (8 turns of 20 Q, see Figure S1 in the Supporting Information) in aqueous solution.<sup>29–32</sup> [The structure is characterized by Q residues with  $\varphi$  and  $\psi$  angles of  $-162^\circ$  and  $159^\circ$ .<sup>26</sup> Its coordinates were kindly provided by Dr. A. Lesk. Although  $\alpha$  helices have a low probability of forming *in vivo* with respect to other Q structures,<sup>27,28</sup> they have been investigated here because (1) they have been already investigated by classical MD by us<sup>9</sup> and (2) we provide a qualitative description CE, independently from the peculiarity of these conformations. Quantitative predictions, which would require an investigation on a variety of structures proposed, are beyond the scope of the present investigation.]

Taken together, our calculations suggest that the CE is manifested both by the shortening of HB lengths, increasing the number of HBs involved (**structural** aspect) and by the energy stabilization of H-bonded peptides with respect to the isolated ones (**energetic** aspect): We are going to detail in the following some of the crucial features of our results.

\* Corresponding author e-mail: alema@sisa.it.

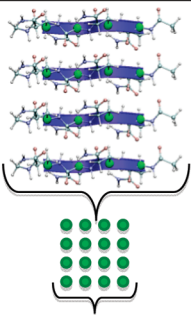
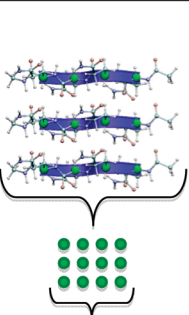
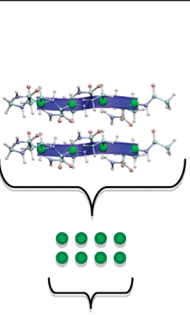
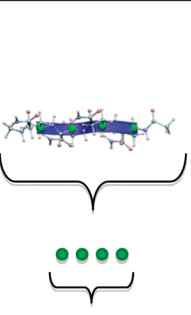












<sup>†</sup> SISSA and CNR-IOM-DEMOCRITOS.

<sup>‡</sup> German Research School for Simulation Science.

<sup>§</sup> Italian Institute of Technology—SISSA Unit.

<sup>||</sup> National Institute for Medical Research.

**Table 1.** polyQ Peptides of Increasing Complexity, Assembled in Parallel  $\beta$  Sheets<sup>a</sup>

Series $N \times 4$	 4x4	 3x4	 2x4	 1x4
Series $N \times 3$	 4x3	 3x3	 2x3	 1x3
Series $N \times 2$	 4x2	 3x2	 2x2	 1x2
Series $N \times 1$	 4x1	 3x1	 2x1	 1x1

<sup>a</sup> Each system studied here is defined in terms of the  $n$  and  $N$  integers, ranging from 1 to 4. The first number counts the Qs in each strand. It defines a group of four systems, each with the same number of Qs per strand, but with a different number of strands (a “series”). The second counts the strands in each system. Thus,  $N \times 4$  indicates systems formed by peptides of 4 Qs ( $1 \times 4$ ,  $2 \times 4$ ,  $3 \times 4$ ,  $4 \times 4$ ),  $N \times 3$  those formed by 3 Qs ( $1 \times 3$ ,  $2 \times 3$ ,  $3 \times 3$ ,  $4 \times 3$ ),  $N \times 2$  those formed by 2 Qs ( $1 \times 2$ ,  $2 \times 2$ ,  $3 \times 2$ ,  $4 \times 2$ ), and  $N \times 1$  those made up of only 1 Q ( $1 \times 1$ ,  $2 \times 1$ ,  $3 \times 1$ ,  $4 \times 1$ ). We built also two other, different  $N \times 3$  series: (A) the  $N \times 3_{SC}$  polyQ series where we varied the side chain conformations and (B)  $N \times 3_{ALA}$ . This is a polyalanine system.

Finally, to prove that such cooperative effects are due only to the peculiarity of polyQ chains, we also considered, as a control study (a) series of models where we varied the initial Q side chain conformations and (b) series of models built with polyalanine.

**Structural Aspects.** CE on a  $\beta$ -sheet system may be present in patterns *perpendicular* to the peptide elongation ( $\perp$ CE) or *parallel* to it ( $\parallel$ CE, Figure 1A).<sup>18</sup>

1. The  $\perp$ CE is manifested (a) by a decrease of HB length with an increasing number of piling strands, and (b) by HBs at the center of the pile shorter than in the rim.<sup>18,22</sup>

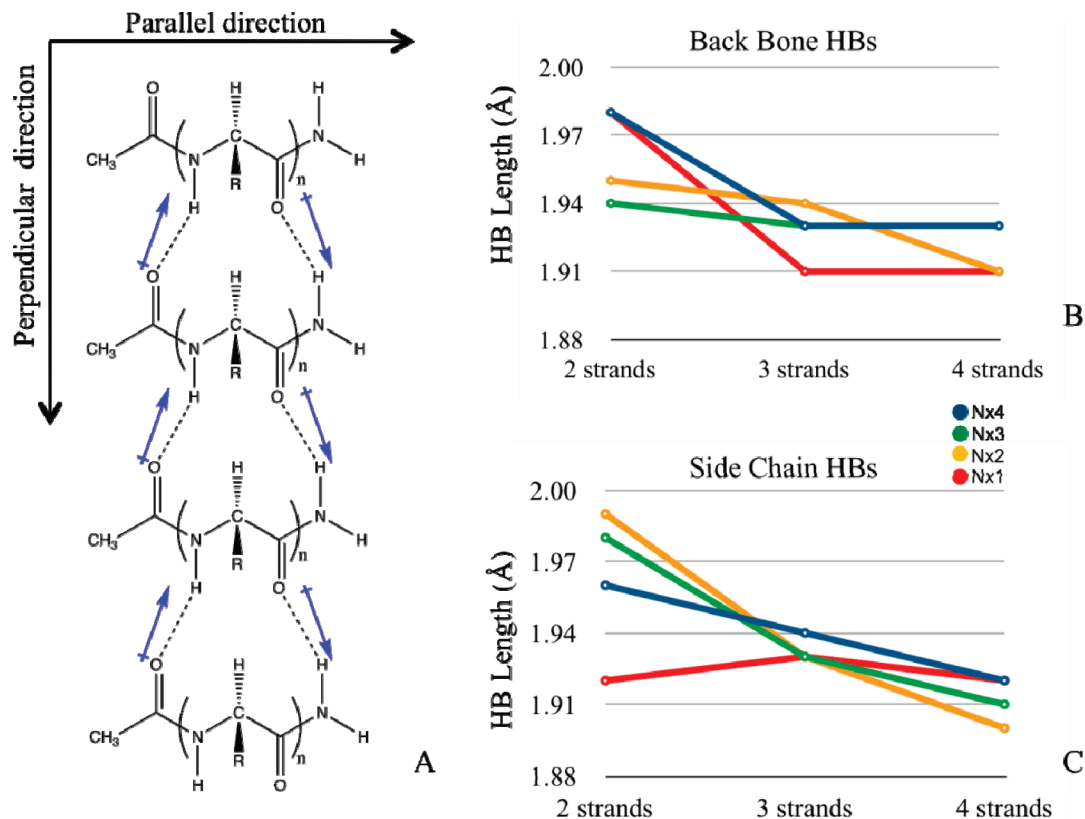
In all the series considered, HB distances decreased with an increasing number of piling strands in both the backbone and the side chains (effect a in Figure 1B,C).

In addition, HB lengths turned out to be shorter at the center of H-bonded chains than at the rim, in the case where at least three HBs are piled up in the perpendicular direction ( $N = 4$ ; effect b in Figures 2–4A and Figure S2, Supporting Information). This feature was observed both for the side chains (Figure 2) and the backbone (Figures 3, 4A, S2).

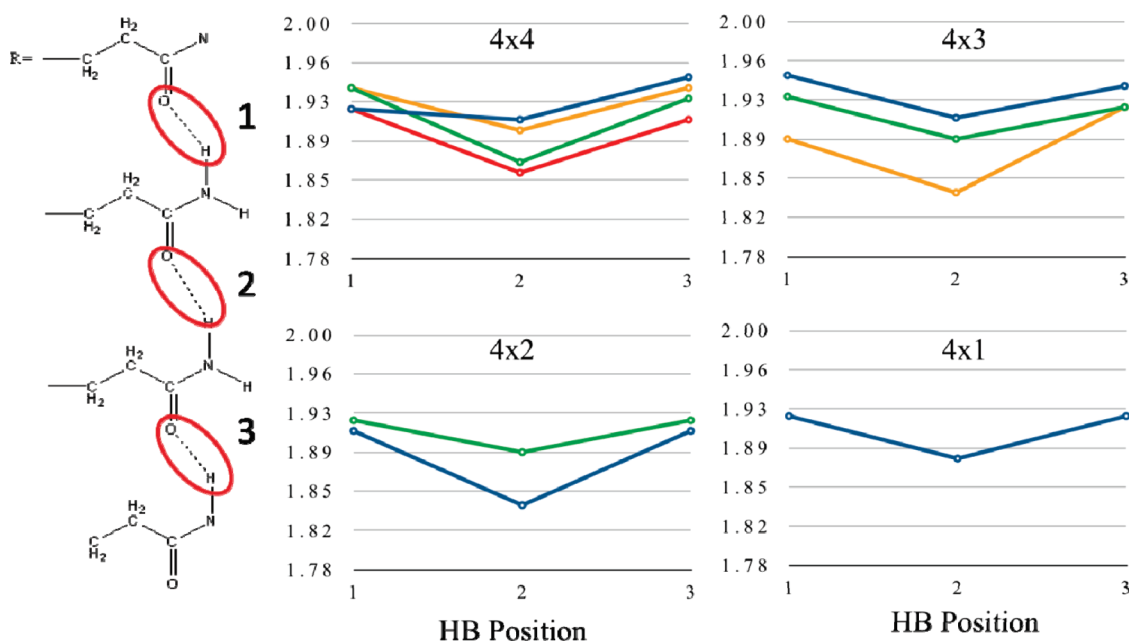
For the backbone, the trend was however observed only when taking averages: the inner HBs turned out not always to be the shortest of the column (Figure 3) [with the term “column”, we indicate the HB chain in the perpendicular direction]. This fact

can be explained, at least in part, by the polarization of the dipoles associated with the HB functionalities ( $C=O \cdots N-H$ ) of both the backbone and side chains. It has been already observed that the backbone dipoles along the same column of  $\beta$  strands have the same orientations (in contrast to those of the adjacent column) and can therefore sum up increasing the polarization of the systems.<sup>22</sup> However, in the case of polyQ  $\beta$ -strands, the glutamine side chains counterbalance this polarization, affecting the inner HBs (Figure 3). Therefore, in the columns where the HB dipole orientations were enhanced by similar side chain HB dipole orientations, a CE is present—the shorter HB was the one in the center of the column; on the other hand, when neighboring side chain columns had HB dipoles oriented in opposite directions (with respect to the column considered), the inner HB was not the shortest of the column (Figure 3).

To prove this conclusion, we perform the same calculations on the  $N \times 3$  polyQ series varying the side chain conformations ( $N \times 3_{SC}$  hereafter). Here, glutamine side chains are twisted in such a way they are not able to establish HBs; thus only backbone HBs are present. As expected, we found both types of  $\perp$ CE (effects a and b). However, remarkably,  $\perp$ CE-type b is not affected by side chain HB dipole orientations due to the absence of side-chain HBs. Thus, backbone HB lengths turned



**Figure 1.** (A) Parallel ( $\parallel$ ) and perpendicular ( $\perp$ ) directions of peptides elongation. (B, C) Structural aspects of CE: (B) Backbone CE ( $\perp$ CE-effect a): mean values of HB lengths of the backbone atoms versus the number of strands for each series of  $n$  Q. (C) Side chain CE ( $\perp$ CE-effect a): mean values of HB lengths for the side chain atoms versus the number of strands for each series of  $n$  Q.



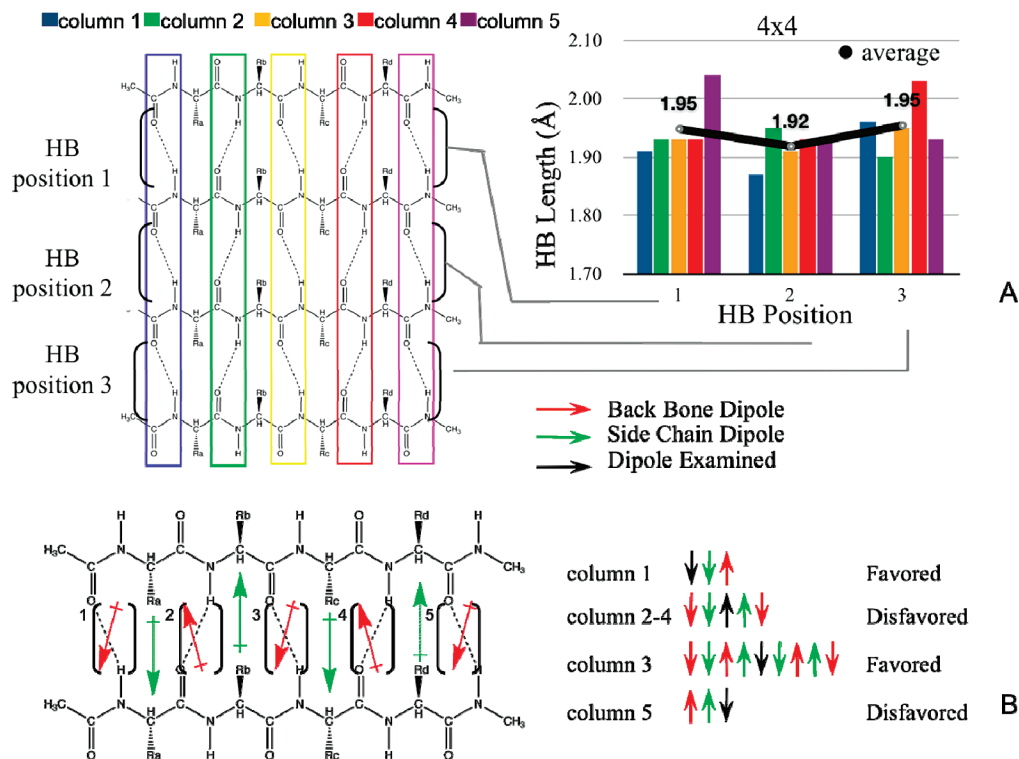
**Figure 2.**  $\perp$ CE-effect b, in the Q side chains: systems  $4 \times 4$ ,  $4 \times 3$ ,  $4 \times 2$ , and  $4 \times 1$ . HB length versus HB position.

out to be shorter at the center of H-bonded chains than at the rim, in each single columns considered, not only taking the average (Figure S3 and Table S2, Supporting Information).

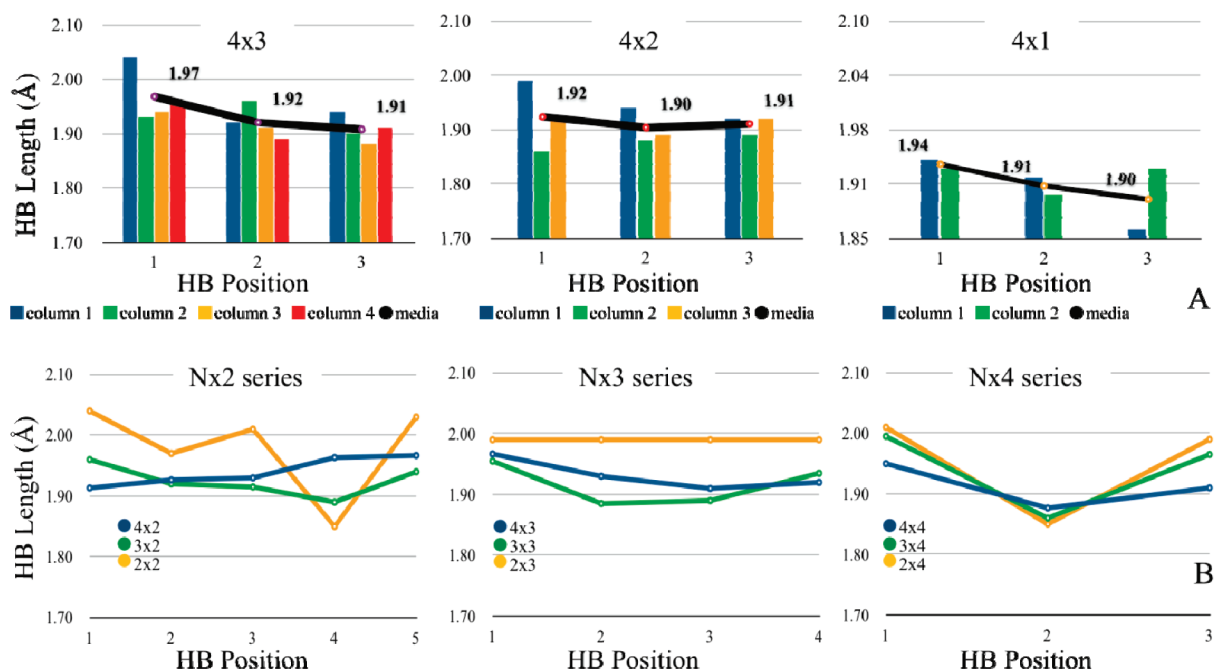
According to this, we have similar results also for  $N \times 3_{\text{ALA}}$ , where there is no possibility for the alanine to form side chain HBs (Figure S3 and Table S2, Supporting Information). Indeed,

we found only a  $\perp$ CE of type b: HB lengths are shorter at the center of H-bonded chains than at the rim, in the case where at least three HBs are piled up in the perpendicular direction ( $N = 4$ ; Figure S3).

2. We observed a  $\parallel$ CE as reflected from shortening of the central HB lengths between two adjacent strands.<sup>18</sup>  $\parallel$ CE is



**Figure 3.**  $\perp$ CE-effect b in system  $4 \times 4$ . (A) In the histograms: HB length of backbone for different positions inside each strand as a function of the position across the different strands. The color of the histogram corresponds to the HBs circled on the top-left picture. The black line represents the mean values over the rows. (B) Orientation of the dipoles associated with the HBs for  $4 \times 4$  ( $4 \times 3$ ,  $4 \times 2$ , and  $4 \times 1$  treated in Supporting Information, Figure S2).



**Figure 4.** (A) Backbone CE in the direction perpendicular to strand elongation ( $\perp$ CE-effect b): systems  $4 \times 3$ ,  $4 \times 2$ , and  $4 \times 1$ . In the histograms: HB length for each column (the position along the strand) versus the HB position (the position perpendicular to the strand direction). The black line represents the mean values over the rows. (B) Backbone  $\parallel$ CE: series  $N \times 2$ ,  $N \times 3$ , and  $N \times 4$ . HB length versus HB position.

usually not present in  $\beta$ -sheets because of the alternative orientation of backbone HB dipoles along the strands (Figure 1A).<sup>18,33,34</sup> However, the dipoles associated with the Q side chains add up in a coherent way for the central HBs between two strands (position 2 in  $N \times 2$  series; positions 2 and 3 in  $N$

$\times 3$  series; positions 2, 3, and 4 in  $N \times 4$  series). As a result, the latter turned out to be shorter than those of the rim (Figure 1B). We performed the same calculation on the  $N \times 3_{SC}$  systems, where there is not a contribution of side chains' HB. As expected, no ICE is found (Figure S4a, Supporting Informa-



tion). These results point to the relevance of glutamine side chains for the structure of polyQ systems.

To confirm that such cooperative effects are specific only for polyQ and not a general feature of polypeptide chains, we performed a control study also on a series ( $N \times 3_{\text{ALA}}$ ) of polyalanine systems (Table S2, Supporting Information). As expected, no  $\parallel\text{CE}$  or  $\perp\text{CE}$  type a was found (Figure S4 b, Supporting Information).

Similar conclusions can be drawn by our hybrid QM/MM calculations of the circular  $\beta$ -helix of the polyQ chain, in which the QM region corresponds to  $4 \times 4$ ,  $4 \times 3$ , and  $3 \times 4$ , and the rest of the polyQ tracts as well as the water molecules were included in the MM region ( $\sim 45\,000$  atoms). These systems were labeled as  $4 \times 4_{\text{MIX}}$ ,  $4 \times 3_{\text{MIX}}$ , and  $3 \times 4_{\text{MIX}}$ . Although the trend of HB lengths in the first two systems qualitatively resembled that of the corresponding *in vacuo* models, we have to remark that the HB lengths were larger (Table S3 and Figure S5, Supporting Information). Moreover, the side chains formed mostly HB with the solvent. These differences are probably due to the presence of the solvent and to temperature effects, which are completely neglected in the *in vacuo* calculations. [We further notice that, because of the very short time-scale of our QM/MM simulation, our structural parameters may also not have reached equilibration.] No CE was observed in the last system ( $3 \times 4_{\text{MIX}}$ ), possibly because of the small number of strands.

**Energetic Aspects.** The stabilization energy associated with the formation of HBs between the different strands of the systems *in vacuo* is calculated as follows. [Unfortunately, the stabilization associated with the addition of a Q unit starts with the fourth amino acid unit,<sup>11</sup> so it cannot be investigated here. In fact, the number of amino acids in our systems is never greater than four. This issue must be addressed in a further study.] First, we define the stabilization energy *per strand* ( $\Delta E_N$ ) as the energy associated with the addition of the  $N$ th Q strand to the  $Q_{N-1}$  ( $E_{N \times n}$ ), minus the formation energy of the  $N$  isolated strand.

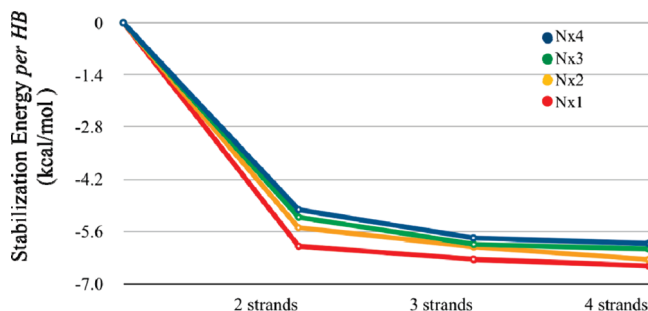
$$\Delta E_N = E_{N \times n} - N \cdot E_{1 \times n}$$

$E_{N \times n}$  is the energy of a system belonging to the  $n$  series and containing  $N$  strands;  $E_{1 \times n}$  is the energy associated with an isolated polyQ strand with  $n$  glutamines. This is the formation energy of a strand constituted by  $n$  glutamines free from long-range effects (i.e., isolated non-interactive strand).

The stabilization energy per hydrogen bond ( $\Delta E_{\text{HB}}$ ) was then defined by dividing  $\Delta E_N$  by the number of hydrogen bonds ( $n_{\text{HB}}$ ) in each system.

$$\Delta E_{\text{HB}} = \Delta E_N / n_{\text{HB}}$$

$\Delta E_{\text{HB}}$  decreased nonlinearly with the number of strands (Figure 5): the variation of  $\Delta E_{\text{HB}}$  in each series is  $\sim 0.8$  kcal/mol, passing from two-strand systems to four-strand systems. This quantity is smaller than the typical DFT-PBE error.<sup>35</sup> However, here, we consider differences of energies in similar systems; thus we can reasonably assume that fortuitous error cancellation errors may increase the accuracy of our calculations.  $\Delta E_{\text{HB}}$  ranged from  $-5.0$  kcal/mol in the smallest system to  $-6.5$  kcal/mol in the larger system ( $4 \times 4$ ), suggesting that a CE



**Figure 5.** Stabilization energy per hydrogen bond ( $\Delta E_{\text{H}}$ ) for the addition of an  $N$ th Q strand to the  $Q_{N-1}$ . The gradual change of  $\Delta E_{\text{H}}$  versus the number of strands showed that the strength of the HBs between layers increases nonlinearly with the number of strands.

effect exists and that for the present systems this is a maximum of 1.5 kcal/mol per HB.

As expected, the stabilization energy depending on CE is smaller for polyA systems with respect to the polyQ, clearly for the absence of side chain HB stabilization. According to this hypothesis, if we compute the CE for the  $N \times 3_{\text{SC}}$  series, where the glutamine side chains are not able to form HBs, we find results comparable with the polyA ones (Table S5, Supporting Information).

In summary, we found that (1) both parallel and perpendicular CEs affect the geometry of polyQ  $\beta$  strands because of the key role of the Q side chains; (2) the formation of cooperative hydrogen bonds stabilized multiple polyQ  $\beta$ -sheet strands with respect to a single isolated strand; (3) within the limitations of the calculations on a single  $\beta$ -stranded structure in a water solution, we suggest that environmental effects on hydrogen bonding CE affects only the magnitude of CE, while the qualitative trend is the same as that found in the *in vacuo* calculation.

**Acknowledgment.** A.P. acknowledges funding from MRC (grant No U117584256).

**Supporting Information Available:** (1) Methods: DFT and DFT/MM calculations. (2) Figure S1, circular  $\beta$ -helix. (3) Figure S2, HB dipole orientations in (a)  $4 \times 3$ , (b)  $4 \times 2$ , and (c)  $4 \times 1$ . (4) Figure S3, (a) backbone CE ( $\perp\text{CE}$ -effect a) in system  $4 \times 3_{\text{SC}}$  and system  $4 \times 3_{\text{ALA}}$ , mean values of HB lengths of the backbone atoms versus the number of strands for each series of  $n$  Q; (b) backbone CE ( $\perp\text{CE}$ -effect b) in the direction perpendicular to strand elongation: system  $4 \times 3_{\text{SC}}$ , system  $4 \times 3_{\text{ALA}}$ . In the histograms: HB length for each column (the position along the strand) versus the HB position (the position perpendicular to the strand direction); the black line represents the mean values over the rows. (5) Figure S4, backbone  $\parallel\text{CE}$ : Series  $N \times 3_{\text{SC}}$ ,  $N \times 3_{\text{ALA}}$ . HB length versus HB position. (6) Figure S5, backbone CE in the direction perpendicular to strand elongation: (a) system  $4 \times 4_{\text{MIX}}$ , (b) system  $4 \times 3_{\text{MIX}}$ . In the histograms: HB length for each column (the position along the strand) versus the HB position (the position perpendicular to the strand direction); the black line represent the mean values over the rows. (7) Table S1, lengths of HBs in backbone and side chains for all the systems studied

with DFT *in vacuo*. (8) Table S2, lengths of HBs in the backbone: series  $N \times 3_{SC}$ ,  $N \times 3_{ALA}$ . (9) Table S3, lengths of HBs in the backbone obtained with DFT/MM MD. (10) Table S4, energies obtained from the DFT calculation *in vacuo*. (11) Table S5, energies obtained from the DFT calculations *in vacuo* for series  $N \times 3_{SC}$ ,  $N \times 3_{ALA}$ . This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Chen, S. M.; Ferrone, F. A.; Wetzel, R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (18), 11884–11889.
- (2) Davies, S. W.; Turmaine, M.; Cozens, B. A.; DiFiglia, M.; Sharp, A. H.; Ross, C. A.; Scherzinger, E.; Wanker, E. E.; Mangiarini, L.; Bates, G. P. *Cell* **1997**, *90* (3), 537–548.
- (3) Masino, L.; Pastore, A. *Brain Res. Rev.* **2001**, *56* (3–4), 183–189.
- (4) Scherzinger, E.; Sittler, A.; Schweiger, K.; Heiser, V.; Lurz, R.; Hasenbank, R.; Bates, G. P.; Lehrach, H.; Wanker, E. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (8), 4604–4609.
- (5) Perutz, M. F.; Johnson, T.; Suzuki, M.; Finch, J. T. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91* (12), 5355–5358.
- (6) Perutz, M. F.; Windle, A. H. *Nature* **2001**, *412* (6843), 143–4.
- (7) Klein, F.; Pastore, A.; Masino, L.; Zederlutz, G.; Nierengarten, H.; Ouladabdelghani, M.; Altschuh, D.; Mandel, J.; Trottier, Y. *J. Mol. Biol.* **2007**, *371* (1), 235–244.
- (8) Ludwig, R. *J. Mol. Liq.* **2000**, *84* (1), 65–75.
- (9) Rossetti, G.; Magistrato, A.; Pastore, A.; Persichetti, F.; Carloni, P. *J. Phys. Chem. B* **2008**, *112* (51), 16843–50.
- (10) Horvath, V.; Varga, Z.; Kovacs, A. *THEOCHEM* **2005**, *755* (1–3), 247–251.
- (11) Horvath, V.; Varga, Z.; Kovacs, A. *J. Phys. Chem. A* **2004**, *108*, 6869–6873.
- (12) Improtà, R.; Barone, V.; Kudin, K. N.; Scuseria, G. E. *J. Chem. Phys.* **2001**, *114* (6), 2541–2549.
- (13) Scheiner, S.; Kar, T. *J. Phys. Chem. B* **2005**, *109* (8), 3681–3689.
- (14) Tsemekhman, K.; Goldschmidt, L.; Eisenberg, D.; Baker, D. *Protein Sci.* **2007**, *16* (4), 761–4.
- (15) Varga, Z.; Kovacs, A. *Int. J. Quantum Chem.* **2005**, *105* (4), 302–312.
- (16) Viswanathan, R.; Asensio, A.; Dannenberg, J. J. *J. Phys. Chem. A* **2004**, *108* (42), 9205–9212.
- (17) Wiczorek, R.; Dannenberg, J. J. *J. Am. Chem. Soc.* **2003**, *125* (27), 8124–9.
- (18) Zhao, Y. L.; Wu, Y. D. *J. Am. Chem. Soc.* **2002**, *124* (8), 1570–1.
- (19) Benedek, N. A.; Snook, I. K.; Latham, K.; Yarovsky, I. *J. Chem. Phys.* **2005**, *122* (14), 144102.
- (20) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101* (18), 6946–51.
- (21) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.
- (22) Koch, O.; Bocola, M.; Klebe, G. *Proteins* **2005**, *61* (2), 310–317.
- (23) Beke, T.; Csizmadia, I.; Perczel, A. *J. Am. Chem. Soc.* **2006**, *128* (15), 5158–5167.
- (24) Perczel, A.; Gaspari, Z.; Csizmadia, I. G. *J. Comput. Chem.* **2005**, *26* (11), 1155–1168.
- (25) *HyperChem 8.0*; Hypercube, Inc.: Gainesville, FL.
- (26) Perutz, M. F.; Finch, J. T.; Berriman, J.; Lesk, A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (8), 5591–5.
- (27) Sikorski, P.; Atkins, E. *Biomacromolecules* **2005**, *6* (1), 425–32.
- (28) Zanuy, D.; Gunasekaran, K.; Lesk, A. M.; Nussinov, R. *J. Mol. Biol.* **2006**, *358* (1), 330–345.
- (29) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.
- (30) *CPMD 3.11.1*; IBM Corp.: Armonk, New York, 1990–2008.
- (31) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116* (16), 6941–6947.
- (32) van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718.
- (33) Hol, W. G.; Halie, L. M.; Sander, C. *Nature* **1981**, *294* (5841), 532–6.
- (34) Kortemme, T.; Ramírez-Alvarado, M.; Serrano, L. *Science* **1998**, *281* (5374), 253–256.
- (35) Piana, S.; Sebastiani, D.; Carloni, P.; Parrinello, M. *J. Am. Chem. Soc.* **2001**, *123* (36), 8730–7.

CT900476E