Supporting information

The role of the N-terminal tail for the oligomerization, folding and stability of human frataxin

Santiago Faraj[a], *Leandro Venturutti*[a], *Ernesto A. Roman*[a], *Cristina B. Marino-Buslje*[b], *Astor Mignone*[a], *Silvio C.E. Tosatto*[c], *José M. Delfino*[a*], *and Javier Santos*[a*]

[a]Department of Biological Chemistry and Institute of Biochemistry and Biophysics (IQUIFIB), School of Pharmacy and Biochemistry, University of Buenos Aires, Junín 956, C1113AAD, Buenos Aires. Argentina.

[b]Bioinformatics Unit. Fundacion Instituto Leloir, Buenos Aires, Argentina.

[c] Department of Biology, University of Padova, Viale G. Colombo 3, 35131 Padova.

Keywords: Frataxin, helical propensity, protein stability, oligomerization, folding.

* Correspondence to: Javier Santos and José María Delfino; Department of Biological Chemistry and Institute of Biochemistry and Biophysics (IQUIFIB), School of Pharmacy and Biochemistry, University of Buenos Aires, Junín 956, C1113AAD, Buenos Aires, Argentina. E–mail: jsantos@qb.ffyb.uba.ar and delfino@qb.ffyb.uba.ar.

Phone: 54 11 4964 8291, extension 108. Fax: 54 11 4962 5457.

Running title: Folding and oligomerization of human frataxin

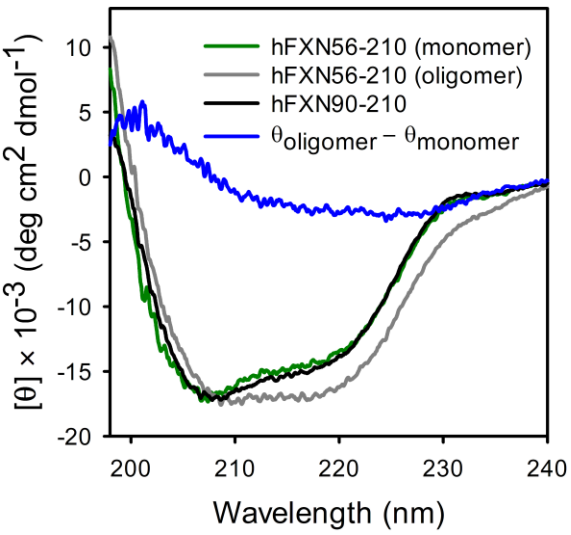

Figure S1. Far-UV CD spectra of hFXN90-210, the monomeric (green line) and oligomeric (grey line) forms of variant hFXN56-210. The difference spectrum (oligomer minus monomer) is also shown (blue). As a control sample, hFXN90-210 was included (black line). All proteins were dissolved in 20 mM Tris-HCl buffer, pH 7.0, containing 100 mM NaCl and 1 mM EDTA. All spectra were acquired at 20 °C.

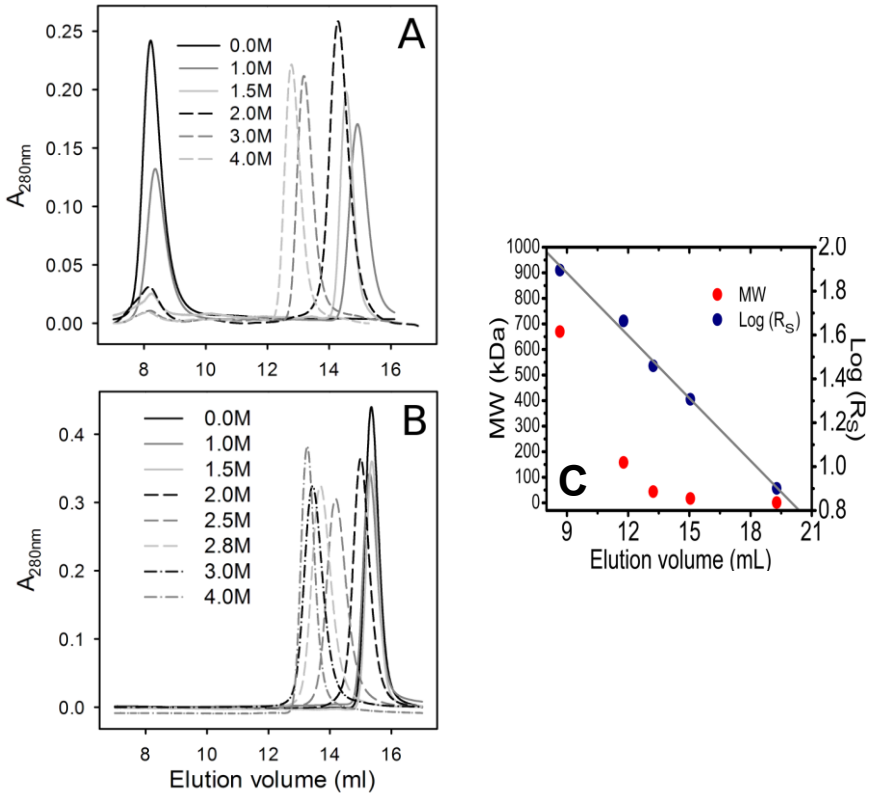

Figure S2. GdmCl-induced unfolding transition of hFXN56-210 (A) and hFXN90-210 (B) variants followed by the change in the elution volume after SEC-FPLC. As in Figure S1, proteins were incubated for 16 h at the indicated GdmCl concentration. Afterwards, samples (100 µL) were injected into the SEC-FPLC system. The column was previously equilibrated at the same concentration of denaturant prepared in buffer 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.0. Flow rate was 0.2-0.4 mL/min, depending on the viscosity of the GdmCl solution. All measurements were performed at room temperature. (C) Size standards (thyroglobulin, 670 kDa; bovine γ globulin, 150 kDa; chicken ovoalbumin, 44 kDa; equine myoglobin,17 kDa; vitB$_{12}$, 1.35 kDa) were run in buffer 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.0. The elution volume is plotted as a function of the log(R_S) or as a function molecular weight (MW). The relation between RS and MW [1] is log(R_S)=0.369×log(MW) -0.254.

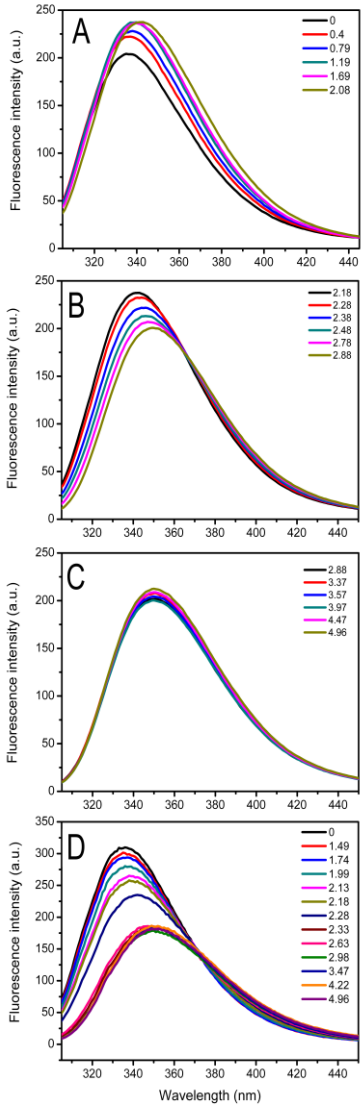

Figure S3. GdmCl-induced unfolding transition of hFXN56-210 (A-C) and hFXN90-210 (D) variants followed by changes in Trp fluorescence emission. Samples were incubated for 16 h a room temperature. Fluorescence spectra (in the range 310–450 nm) were acquired using an excitation wavelength of 295 nm and a spectral slit–width set at 3 nm for both monochromators. Buffer was 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.0. All measurements were performed at 20 °C.

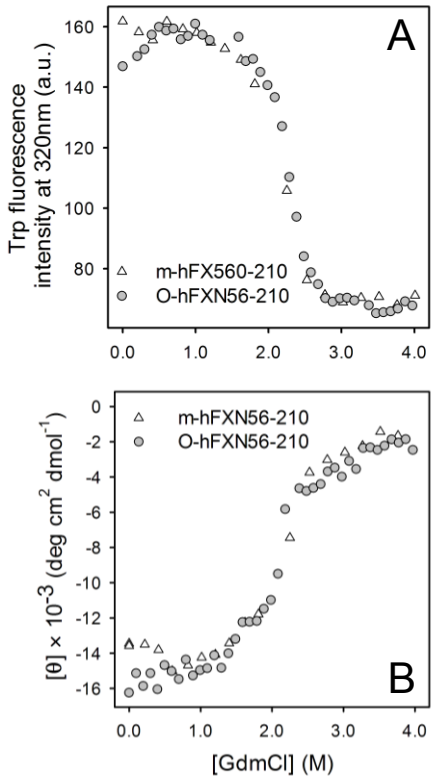

Figure S4. Comparison between unfolding curves corresponding to hFXN56-210 variant purified as the monomeric (m-hFXN56-210, Δ) or oligomeric forms (O-hFXN56-210, ○). Unfolding was followed by (**A**) far-UV CD at 220 nm and by (**B**) Trp fluorescence intensity. Buffer was 20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.0. All measurements were performed at 20 °C.

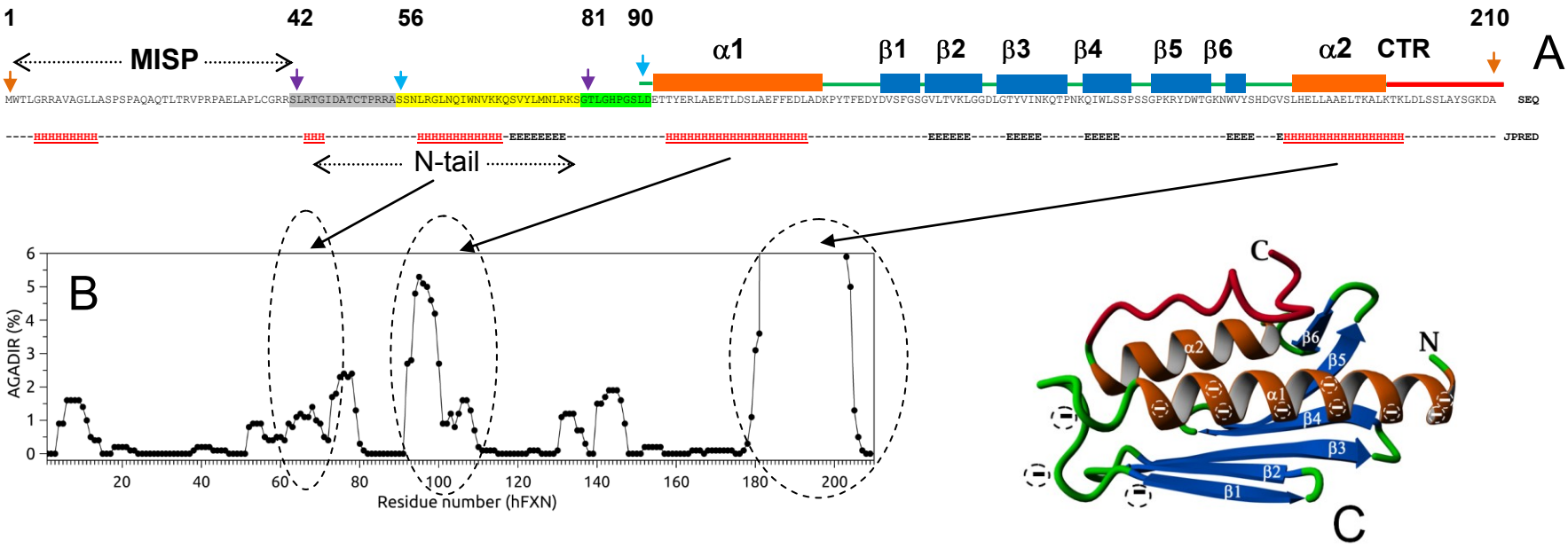

Figure S5. The sequence of hFXN (residues 1-210) and the tendency of N–terminal tail of hFXN 56–210 to form α-helical and β structure according to Jpred are shown (A). MISP is the mitochondria import signal peptide. The sequence corresponding to peptide hFXN56-81 is highlighted in yellow. In addition, boxes corresponding to α-helices, loops and β-strands are represented accordingly to PDBID: 1EKG in orange, green and blue, respectively. The C-terminal region (CTR) of hFXN is highlighted in red. Finally, the amino acid sequence 42-56 is highlighted in gray. Arrows in magenta represent points of processing during the import to the mitochondria (residues 42, and 81). Arrows in cyan indicate the N-terminal of hFXN56-210 and hFXN90-210, both studied in this work. In addition, the first and the last residues of the protein (1 and 210) are decorated with brown arrows. (B) The prediction of secondary structure

stability by AGADIR, pH 7.0, 100mM NaCl at 5 ºC. (C) The 3D structure of hFXN90–210 (PDB ID: 1EKG) represented in ribbons. Acidic residues involved in iron binding are denoted with the sign (-).

List of sequences used for secondary structure prediction on the MSA of the N-terminal tail of FXN.

gi|397508736|ref|XP_003824802.1| [Pan paniscus]
gi|426361957|ref|XP_004048150.1| [Gorilla gorilla gorilla]
gi|332249458|ref|XP_003273877.1| [Nomascus leucogenys]
gi|402897579|ref|XP_003911830.1| [Papio anubis]
gi|355753397|gb|EHH57443.1| [Macaca fascicularis]
gi|386781775|ref|NP_001247670.1| [Macaca mulatta]
gi|296189789|ref|XP_002742919.1| [Callithrix jacchus]
gi|332832118|ref|XP_001137864.2| [Pan troglodytes]
gi|403301852|ref|XP_003941591.1| [Saimiri boliviensis boliviensis]
gi|395819437|ref|XP_003783094.1| [Otolemur garnettii]
gi|472355878|ref|XP_004397585.1| [Odobenus rosmarus divergens]
gi|431898667|gb|ELK07047.1| [Pteropus alecto]
gi|350579295|ref|XP_003480578.1| [Sus scrofa]
gi|338719548|ref|XP_001490501.3| [Equus caballus]
gi|512829179|ref|XP_004881608.1| [Heterocephalus glaber]
gi|344271243|ref|XP_003407450.1| [Loxodonta africana]
gi|158262735|ref|NP_001103428.1| [Canis lupus familiaris]
gi|507933632|ref|XP_004678052.1| [Condylura cristata]
gi|440902536|gb|ELR53319.1| [Bos grunniens mutus]
gi|507538043|ref|XP_004653043.1| [Jaculus jaculus]
gi|426220354|ref|XP_004004381.1| [Ovis aries]
gi|410978095|ref|XP_003995432.1| [Felis catus]
gi|466041703|ref|XP_004276467.1| [Orcinus orca]
gi|444722414|gb|ELW63111.1| [Tupaia chinensis]
gi|511899504|ref|XP_004769299.1| [Mustela putorius furo]
gi|470641967|ref|XP_004325704.1| [Tursiops truncatus]
gi|488525972|ref|XP_004454825.1| [Dasypus novemcinctus]
gi|505768208|ref|XP_004600382.1| [Sorex araneus]
gi|300794591|ref|NP_001178881.1| [Rattus norvegicus]
gi|31077081|ref|NP_000135.2| [Homo sapiens]

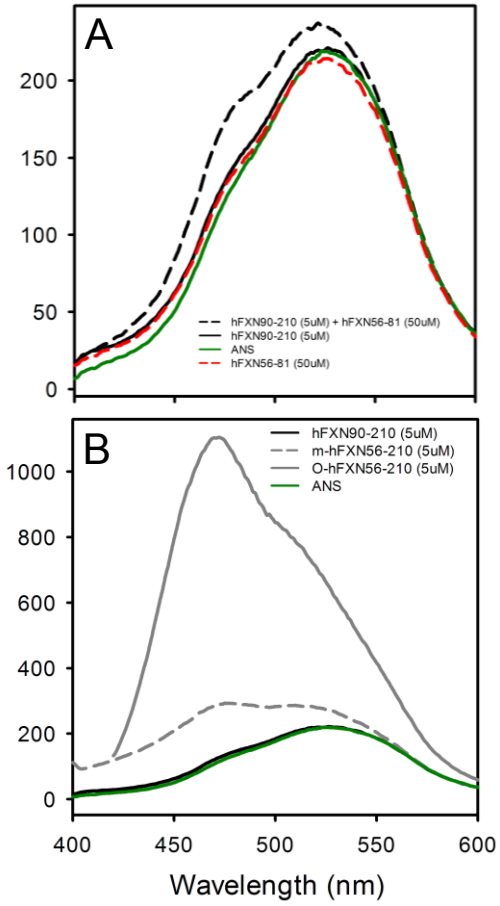

Figure S6. ANS fluorescence spectra of hFXN preparations. **(A)** Variant hFXN90-210 (5 μM), isolated peptide hFXN56-81(50 μM), or a mixture of hFXN90-210 (5 μM) and peptide pFXN56-81A (50 μM) were mixed with ANS (50 μM). **(B)** hFXN56-210 purified as a monomer (grey dashed line) or as an oligomer (grey solid line) was incubated with ANS (50 μM). The spectrum of a mixture of variant hFXN90-210 and ANS is also shown. Buffer was 20 mM Tris-HCl, 100 mM NaCl, pH 7.0 and samples were incubated for 5 min at room temperature before measurements. The excitation wavelength was 350 nm.

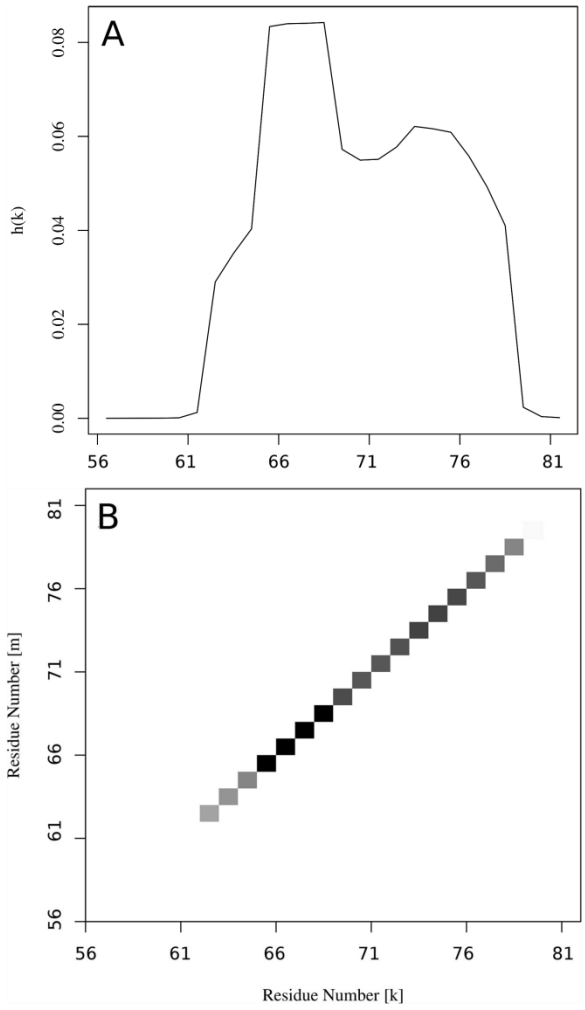

Figure S7. Prediction of amyloid structure aggregation using PASTA on the hFXN56–81 sequence. The hFXN sequence was submitted to the server PASTA (Prediction of amyloid structure aggregation, [2]). The aggregation profile (A) and pairing matrix plot (B) are shown. The former illustrates the aggregation probability for each position along the sequence. The latter represents a 2D self-alignment matrix where shades of grey indicate the probability for two residues to be part of a β -pairing. β-pairings known to appear in cross β-fibrillar aggregates generally yield PASTA energies ≤ -4.0.

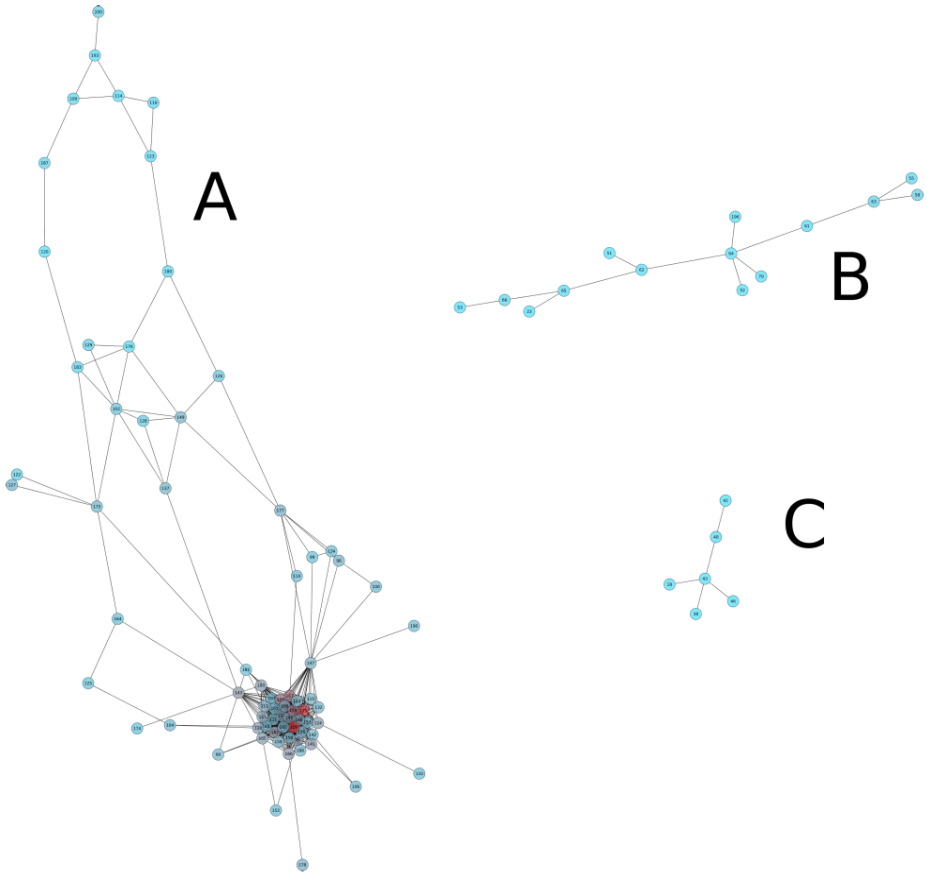

Figure S8. Co-evolution network analysis of mutual information on FXN sequences containing N-terminal extensions. Three networks (A, B and C) result from this analysis**.** Nodes are residues colored according to conservation across the MSA (from cyan to red, to indicate lower to higher conservation). Edges represent a significant MI relationship. Sequences containing N-terminal extensions (n=165) were selected from the global alignment of the CyaY family (n=948 sequences in PFAM). Mutual information and conservation scores were calculated as in references [3].

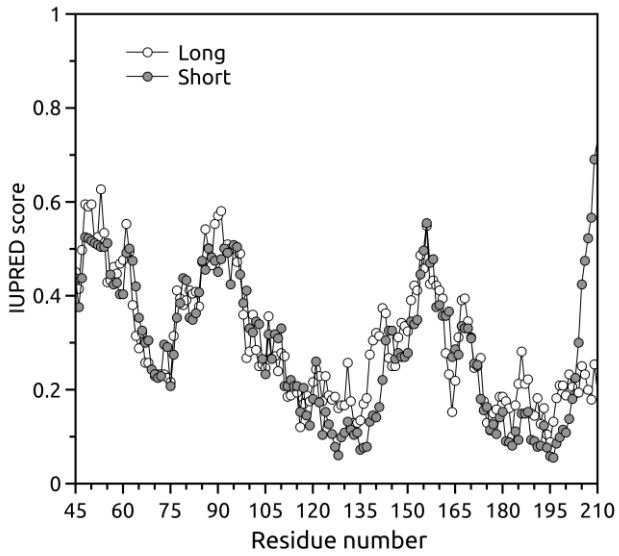

Figure S9. Prediction of intrinsically disordered segments along the sequence of hFXN using IUPRED [4]. Residues 66-75 share a lower tendency to form disordered regions, by comparison to residues 45-61 and 76-90. Predictions for both long or short disordered stretches were in general coincident.

Table S1. Prediction of amyloid structure aggregation using PASTA on the sequence of peptide hFXN56–81.

Pairing	PASTA Energy[1]	Length (in residues)	Interaction between Segments	Relative orientation (parallel or antiparallel)
1	-4.888	14	65-78 and 65-78	parallel
2	-4.775	4	65-68 and 65-68	parallel
3	-4.677	17	62-78 and 62-78	parallel
4	-4.564	7	62-68 and 62-68	parallel
5	-4.113	13	65-77 and 65-77	parallel
6	-3.989	6	73-78 and 73-78	parallel
7	-3.967	12	65-76 and 65-76	parallel
8	-3.902	16	62-77 and 62-77	parallel
9	-3.902	16	63-78 and 63-78	parallel
10	-3.884	11	65-75 and 65-75	parallel

[1]β-pairings known to appear in cross-β fibrillar aggregates generally yield PASTA energies ≤ -4.0 [2].

References

1. Uversky VN (1993) Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry* **32**, 13288-13298.
2. Trovato A, Seno F & Tosatto SC (2007) The PASTA server for protein aggregation prediction. *Protein Eng Des Sel* **20**, 521-523.
3. Aguilar D, Oliva B & Marino Buslje C (2012) Mapping the mutual information network of enzymatic families in the protein structure to unveil functional features. *PLoS One* **7**, e41430.
4. Dosztanyi Z, Csizmok V, Tompa P & Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434.