FEBS Letters 585 (2011) 3086–3089

journal homepage: www.FEBSLetters.org

# Conformations of the Huntingtin N-term in aqueous solution from atomistic simulations

Giulia Rossetti [a,b,1], Pilar Cossio [b,1], Alessandro Laio [b], Paolo Carloni [a,b,*]

[a] German Research School for Simulation Science, FZ-Juelich and RWTH Aachen, Wilhelm-Johnen-Straβe, 52428 Jülich, Germany
[b] Statistical and Biological Physics Sector, International School for Advanced Studies (SISSA) and CNR-IOM DEMOCRITOS, Via Bonomea 265, 34136 Trieste, Italy

ARTICLE INFO

*Article history:*
Received 14 April 2011
Revised 20 July 2011
Accepted 23 August 2011
Available online 30 August 2011

Edited by Robert B. Russell

*Keywords:*
Huntingtin
N-terminal
Bias exchange metadynamics
Structural prediction

ABSTRACT

**The first 17 amino acids of Huntingtin protein (N17) play a crucial role in the protein's aggregation. Here we predict its free energy landscape in aqueous solution by using bias exchange metadynamics. All our findings are consistent with experimental data. N17 populates four main kinetic basins, which interconvert on the microsecond time-scale. The most populated basin (about 75%) is a random coil, with an extended flat exposed hydrophobic surface. This might create a hydrophobic seed promoting Huntingtin aggregation. The other main populated basins contain helical conformations, which could facilitate N17 binding on its cellular targets.**



## 1. Introduction

Huntington disease (HD) is an autosomal-dominant neurodegenerative disorder, for which there is no cure [1,2]. It is caused by expanded CAG trinucleotide repeats in the gene that encodes the large (~3500 amino acids) protein Huntingtin (Htt). The resulting mutant, with an extended polyQ tract in the N-terminal region, interacts abnormally with other proteins leading to neuronal dysfunction [2].

Recently, in vivo [3–5], in cell [6–8], in vitro [9–11] and in silico studies [12] showed that the N-terminal 17 amino acids fragment (sequence: MATLEKLMKAFESLKSF – N17 hereafter), modulate Htt fibrillation. This might arise by a variety of mechanisms, including changes in subcellular localization, nucleation of aggregation and/or interaction with cellular partners [5].

Understanding these mechanisms greatly benefits from structural information. NMR [10], CD [10,11] and FRET [10] have established that N17 in aqueous solution does not exhibit one unique structure. It adopts predominantly unfolded, random-coil conformations with transient helical conformations [10]. Indeed, peptides in solution can exist in equilibrium between different conformations [13–17]. Experiments have so far provided information only on **averages** between the populations of different conformers at room temperature in ms time-scale for the NMR and CD technique. Neither the secondary structure content nor the tertiary structure for each conformer is known.

All atom – molecular dynamics (MD) simulations running on tailored machines or on massive collective calculation initiatives (such as folding@home) are a powerful tool to predict the structural determinants of peptides in solution. At times the latter can be predicted also by free energy calculations as a function of a few collective variables (CVs) (see, e.g. [18,19]). Here we use one of these methods, bias exchange metadynamics (BEM), to describe the thermodynamics and the kinetics of N17 in aqueous solution at room temperature. BEM has been already used to address similar problems (see, e.g. [20]).

## 2. Materials and methods

N17's extended coil conformation was built with the Modeller 9v8 program [21]. The D and K residues were considered to be in their ionized state. The peptide was inserted into a cubic box (vector 7.18 nm) of 4100 water molecules and one chloride ion added to achieve electroneutrality. Periodic boundary conditions were applied. The AMBER (parm99) [22], Aqvist [23], and TIP3P force

*Abbreviations:* HD, Huntington disease; Htt, Huntingtin protein; N17, first seventeen amino acids of Htt; MD, molecular dynamics; BEM, bias exchange metadynamics; CV, collective variable

* Corresponding author at: Statistical and Biological Physics Sector, International School for Advanced Studies (SISSA) and CNR-IOM DEMOCRITOS, Via Bonomea 265, 34136 Trieste, Italy. Fax: +49 2461 61 4823.
*E-mail address:* p.carloni@grs-sim.de (P. Carloni).

[1] These authors contributed equally to this work.

0014-5793/$36.00 © 2011 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2011.08.036

**Table 1**
Selected properties of the four basins emerging from our calculations.

| Basin | Population (%) | Gyration radius (nm) | SASA ($nm^2$) | Phobic area ($nm^2$) | Phylic area ($nm^2$) | End-to-end distance (nm) |
|---|---|---|---|---|---|---|
| B1 | 11 | 0.83 ± 0.01 | 27.30 ± 0.08 | 12.84 ± 0.03 | 14.45 ± 0.06 | 1.77 ± 0.06 |
| B2 | 75 | 1.08 ± 0.01 | 29.45 ± 0.04 | 13.27 ± 0.02 | 16.18 ± 0.03 | 2.45 ± 0.04 |
| B3 | 10 | 0.96 ± 0.01 | 28.35 ± 0.09 | 12.92 ± 0.04 | 15.43 ± 0.06 | 1.11 ± 0.04 |
| B4 | 4 | 0.84 ± 0.01 | 27.88 ± 0.10 | 12.60 ± 0.05 | 15.27 ± 0.06 | 2.24 ± 0.06 |

fields [24], were used for the protein, the counter ions, and water, respectively. Long-range electrostatic interactions were calculated with the particle mesh Ewald [25] method. A grid spacing of 0.12 nm was used. A fourth-order cubic spline interpolation [26] was used to compute the potential and forces between grid points. The cutoff radius for the real part of electrostatics, as well as that for the Lennard-Jones interactions, was set to 1.2 nm. The NPT ($T$ = 300 K, $P$ = 1 bar) ensemble was simulated using the Nosè-Hoover [27] and Andersen-Parrinello-Rahman [28] coupling schemes for temperature and pressure, respectively. The LINCS algorithm [29] was used to constrain all bond lengths involving hydrogen atoms. The time-step was set to 2 fs. The system underwent 100,000 steps of energy minimization whilst imposing harmonic position restraints of 1000 KJ/(mol $nm^2$) on water. It was then heated from 0 to 300 K by increasing the temperature by 25 K every 100 ps of MD. It then underwent 100 000 steps of energy minimization and finally 20 ns of MD. The simulations were performed with the GROMACSv4.0.7 program [35]. The free energy was calculated using BEM [30] as a function of six dimensionless collective variables (CVs, see Supplementary data for details). The first three ($CV_1$, $CV_2$, $CV_3$) count the number of hydrophobic contacts, of $C_\alpha$ contacts, and of backbone hydrogen bonds. $CV_4$ and $CV_5$ monitor the helical content in the whole and central part of the peptide. $CV_6$ is the dihedral correlation between successive $\psi$ dihedrals. The Gaussian widths (3.0, 6.0, 3.0, 0.6, 0.6 and 0.6 for $CV_1$–$CV_6$, respectively) were optimized as in Ref. [31]. The total bias simulation time was 240 ns (40 ns for replica). Convergence was reached after 12 ns in each replica. The CV space was divided in a grid of clusters [20]. The free energy of each cluster is estimated by a weighted-histogram approach [32]. The average value of observables <O>, including all properties of Table 1) was calculated as: $\langle O\rangle = \sum_i (O_i * \exp(-F_i/T))/\sum_i(\exp(-F_i/T))$, where the sums run over all the clusters, $T$ is the temperature and $O_i$ is the average value of $O$ in the cluster $i$. If the cluster size is small enough, the bias potentials are approximately constant for the configurations belonging to the same cluster [33].

The transition rates between each cluster are estimated by the kinetic model described in [20], where a Markovian diffusive behavior is assumed [34] (see Supplementary data). Estimations of thermodynamic and kinetic properties require finding the minimal number of independent CVs able to describe with a good statistic the behaviors of the clusters. These turn out to be here $CV_2$, $CV_4$, $CV_5$ and $CV_6$. The set of clusters was thus defined by partitioning this 4 dimensional CV space in small hyper-rectangles of sizes 9.2, 1.19, 0.6 and 1.27, respectively, for each CV.

The helix propensity, hydrophobicity and number of buried residues of N17 as well as those of the mutants in Table 1 were estimated using the AGADIR [36] at http://agadir.crg.es/), PEPINFO [37] at http://emboss.sourceforge.net/ and JPRED3 [38] at http://www.compbio.dundee.ac.uk/, respectively.

## 3. Results and discussion

Our calculations suggest that there are four kinetic basins of N17 in aqueous solution at 300 K (B1–B4 in Fig. 1). Each of them is characterized by a population and by an attractor, which is de-

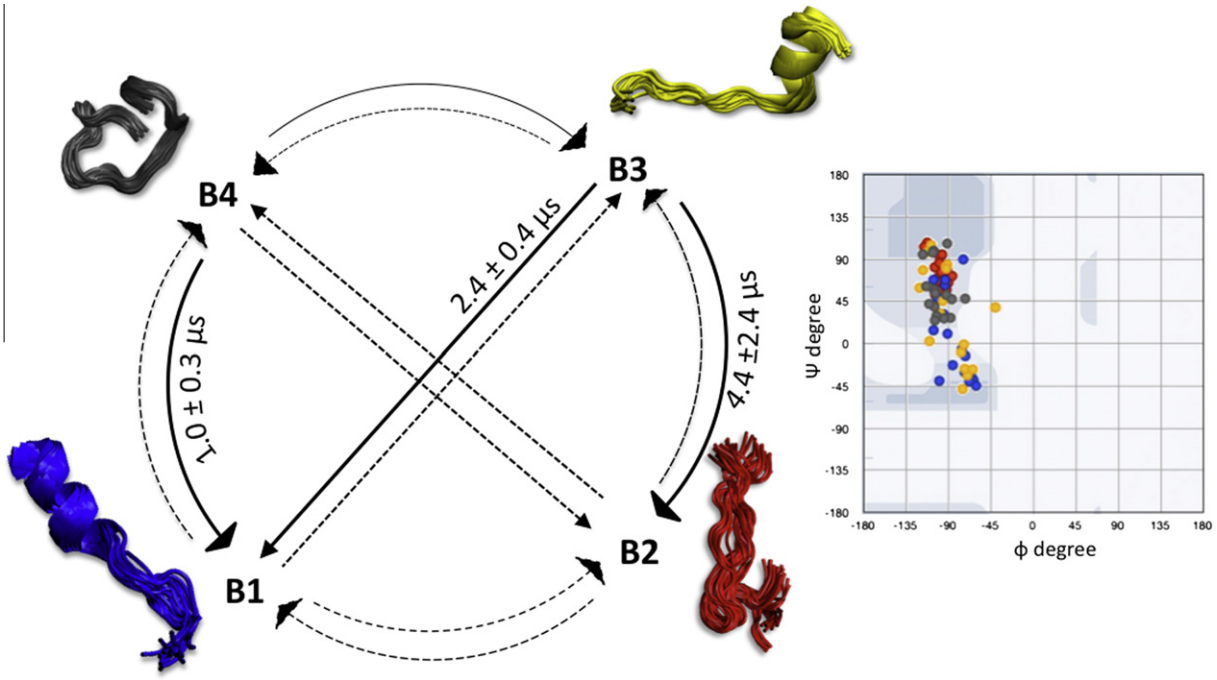


**Fig. 1.** Four basin (B1–B4) of N17 in aqueous solution emerging from this computational study. B1–B4 are characterized by their *population* and by their *attractor*. This is defined as the cluster with lowest free energy in the basin. Only the attractors' structures and their correspondent Ramachandran plots are shown for clarity. The attractors structures of B1 are colored in red, those of B2 in blue, those of B3 in yellow and those of B4 in grey. The calculated interconversion rates with their corresponding statistical errors are reported. Dotted arrows are used for rates >2 μs.