

# Supporting Informations: Effect of urea on $\beta$ -harpin conformational ensemble and protein denaturation mechanism

Anna Berteotti, Alessandro Barducci,<sup>\*</sup> and Michele Parrinello

E-mail: alessandro.barducci@phys.chem.ethz.ch

## Well-tempered metadynamics

Metadynamics<sup>1</sup> is a powerful technique for enhancing sampling in molecular dynamics simulations. It is able to reconstruct the free-energy surface as a function of few selected degrees of freedom, called collective variables (CVs),  $S$ . In metadynamics, an external history-dependent bias potential which is a function of the CVs, is added to the Hamiltonian of the system. The bias potential is built during the simulation as a sum of Gaussian functions centered on the previously visited configurations in the CVs space discouraging the system from revisiting configurations that have already been sampled. The free energy surface, as a function of the CVs, can be reconstructed as the negative of the sum of the added gaussians. At variance with metadynamics, in the well-tempered formalism<sup>2</sup> the initial deposition rate  $\omega_0$  of the bias potential decreases with the bias accumulated over time. This is achieved by rescaling the Gaussian height  $W$  according to:

$$W = \omega_0 \tau_G e^{-\frac{V_G(S,t)}{k_B \Delta T}} \quad (1)$$

---

<sup>\*</sup>To whom correspondence should be addressed

where  $\tau_G$  is the Gaussian deposition stride,  $\Delta T$  a temperature and  $V_G(S, t)$  is the bias potential accumulated in  $S$  over time  $t$ . At variance with standard metadynamics, the bias potential does not fully compensate the FES, but it converges to:

$$V_G(S, t \rightarrow \infty) = -\frac{\Delta T}{\Delta T + T} F(S) \quad (2)$$

where  $T$  is the temperature of the system. In other words, at convergence the CVs are sampled at a (fictitious) higher temperature  $T + \Delta T$  :

$$P(S, t \rightarrow \infty) \propto e^{-\frac{F(S)}{k_B(T+\Delta T)}} \quad (3)$$

Therefore, for  $\Delta T \rightarrow 0$ , ordinary MD is recovered, whereas the  $\Delta T \rightarrow \infty$  corresponds to standard metadynamics. In between one can regulate the extent of FES exploration by tuning  $\Delta T$ . This avoids overfilling and might save computational time when a large number of CVs are used. For a comprehensive description of metadynamics see also.<sup>3</sup>

## Collective variables

We analyzed both WS and US systems as a function of the following variables:

1. radius of gyration calculated on all the  $C\alpha$  ( $R_{ca}$ ) and the distance from the crystallographic state in contact map space ( $Z_{map}$ ). For the  $R_{ca}$  we used the same formula utilized to compute the radius of gyration for the hydrophobic core (Eq.1 in the main text). The distance from the crystallographic state in contact map space was calculated as

$$Z_{map} = \sqrt{\sum_i^{N_{cont}} (C_i(r) - C_i(Xray))^2} \quad (4)$$

where  $C_i(r)$  is the  $i$ th contact for the configuration  $r$  and  $N_{cont}$  is the total number of contacts considered, and  $C_i(Xray)$  is the  $i$ th contact for the crystallographic state. We considered all

the contacts between the C $\alpha$  atoms present in the crystallographic structure and calculated as

$$C(R_{i,j}) = \frac{1 - \left(\frac{r_{i,j}}{r_0}\right)^p}{1 - \left(\frac{r_{i,j}}{r_0}\right)^q} \quad (5)$$

where  $d_{ij}$  is the distance between the  $i$ th and  $j$ th C $\alpha$  atoms of the protein backbone,  $r_0$  is taken to be 8.5 Å,  $n$  and  $m$  are set to 6 and 10 respectively. All the contacts formed by the  $j > i$ th C $\alpha$  atoms have been considered.

2. the total number of hydrogen bonds (H-bonds) between backbone polar groups and solvent ( $HB_{bb-solv}$ ) and  $R_{ca}$ . The  $HB_{bb-solv}$  has been calculated as the sum of contacts between backbone polar groups and water ( $HB_{bb-wat}$ ) plus those between backbone and urea ( $HB_{bb-ure}$ ) :

$$HB_{bb-solv} = HB_{bb-wat} + HB_{bb-ure} \quad (6)$$

In detail

$$HB_{bb-wat} = \sum_{Ob_{Hw}} \frac{1 - (d_{Ob_{Hw}}/2.5)^{80}}{1 - (d_{Ob_{Hw}}/2.5)^{110}} + \sum_{Hb_{Ow}} \frac{1 - (d_{Hb_{Ow}}/2.5)^{80}}{1 - (d_{Hb_{Ow}}/2.5)^{110}} \quad (7)$$

and

$$HB_{bb-ure} = \sum_{Ob_{Hu}} \frac{1 - (d_{Ob_{Hu}}/2.5)^{80}}{1 - (d_{Ob_{Hu}}/2.5)^{110}} + \sum_{Hb_{Ou}} \frac{1 - (d_{Hb_{Ou}}/2.5)^{80}}{1 - (d_{Hb_{Ou}}/2.5)^{110}} \quad (8)$$

where Ob and Hb are oxygens and hydrogens of the backbone, Ow and Hw oxygens and hydrogens of water and Ou and Hu oxygens and hydrogens of urea.

3. the total number of H-bonds between the side chain polar groups and solvent ( $HB_{side-solv}$ ) and  $R_{ca}$ . The  $HB_{side-solv}$  has been calculated as the sum of contacts between the side chain polar groups and water ( $HB_{side-wat}$ ) plus those with urea ( $HB_{side-ure}$ ) :

$$HB_{side-solv} = HB_{side-wat} + HB_{side-ure} \quad (9)$$

The formula and the switching function parameters are the same as for the case of  $HB_{bb-solv}$ .

4. the number of all H-bonds within the backbone ( $HB_{tot}$ ) and  $R_{ca}$ . For  $HB_{tot}$  the same formula as in Eq.3 in the main text has been used, but here all the possible donor-acceptor H-bonds have been taken into account. The parameters were 2.5 Å for  $r_0$ , 20 and 80 for  $n$  and  $m$ .
5. the non native H-bonds ( $HB_{nonnat}$ ) calculated as the difference between  $HB_{tot}$  and the number of the native H-bonds and  $R_{ca}$ . For the native H-bonds we used the same formula as in Eq. 3 in the main text with  $r_0 = 2.5$  Å, 20 and 80 for  $n$  and  $m$ . Only the six H-bonds of the native structure have been considered in the sum.
6. the solvent accessible surface (SAS), calculated by means of g\_sas tool of Gromacs<sup>4</sup> on the hydrophobic side chains of Trp, Tyr, Val and Phe.
7. in order to understand the urea mechanism of action we also reweighted in two separated FESs the contribution of H-bonds between backbone and water ( $HB_{bb-wat}$ ) and those between backbone and urea ( $HB_{bb-ure}$ ) as a function of  $R_{ca}$ .

## H-bonds between solvent and side-chain groups

In Figure S1 the average number of sidechain-solvent H-bonds,  $HB_{side-solv}$ , as a function of  $R_{ca}$  both in WS and US is reported. This analysis has to be compared to Fig.4-C of the manuscript, where an equivalent analysis was performed for backbone-solvent H-bonds,  $HB_{bb-solv}$ . Comparing the two pictures it can be easily noticed that while the average number of both  $HB_{side-solv}$  and  $HB_{bb-solv}$  is greater in US than in WS, the relative differences are much more pronounced for  $HB_{bb-solv}$ . The different behavior is striking in the region from  $R_{ca} = 8$  Å to  $R_{ca} = 10$  Å where the average number of H-bonds in US is  $\sim 30\%$  larger than in WS considering  $HB_{bb-solv}$  whereas the relative difference is less than  $\sim 10\%$  for  $HB_{side-solv}$ .

These results suggest that, at least for GB1 beta-hairpin, urea solution favors the formation of both solvent-backbone and solvent-sidechain H-bonds but the former are likely to play a greater role in urea-driven denaturation in agreement with experimental results.<sup>5</sup>

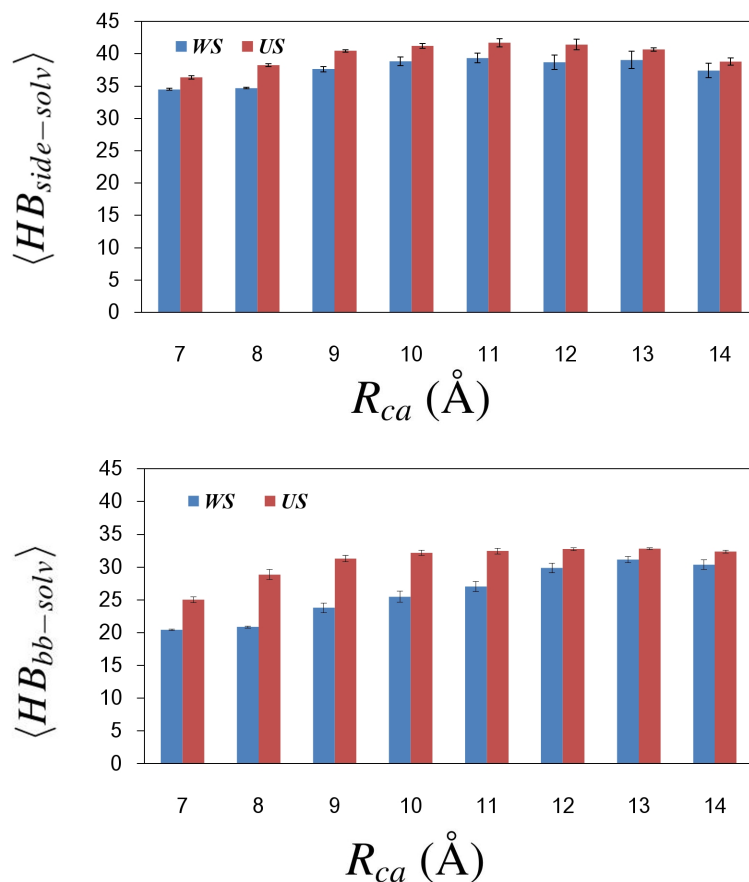


Figure S1: Upper panel: the average number of sidechain-solvent H-bonds in WS and US as a function of  $R_{ca}$ ; lower panel: the average number of backbone-solvent H-bonds in WS and US as a function of  $R_{ca}$ .

## Protein-solvent potential energy in US

The different components of the total protein-solvent potential energy are reported for urea solution system. Particularly, the total protein-solvent potential energy is subdivided into four terms corresponding to the Coulomb and Lennard-Jones contributions to the interaction between the protein and all urea or water molecules within 5 Å from the protein. The dependence of the average of the single contributions on the radius of gyration ( $R_{ca}$ ) is reported as the differences relative to the values at  $R_{ca} = 7$  Å which corresponds to the folded state ( $urea\_LJ(R_{ca}) - urea\_LJ(R_{ca} = 7 \text{ Å})$ ), and so on). The results clearly show that for what matters the interaction between protein and urea molecules, the Coulomb term is greater than the Lennard Jones one. This finding thus agrees

with the proposed mechanism of action of urea based on the formation of hydrogen bonds between the protein and urea as the main driving force of unfolding mechanism.

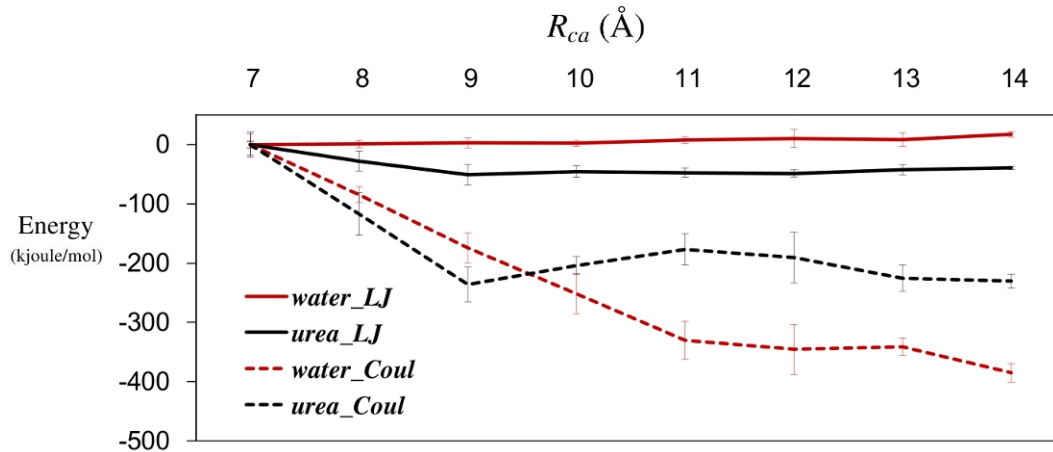


Figure S2: Average values of the interaction energy contributions along the  $R_{ca}$  of the protein with the solvent environment. The differences relative to the values at  $R_{ca} = 7$  Å are reported.

## Convergence of PTMetaD

We monitored the convergence of the PTMetaD simulation by calculating the free energy difference between folded and unfolded states as a function of sampling time:

$$\Delta F_{FU}(t) = -\frac{1}{\beta} \ln \left( \frac{\int_0^3 e^{-\beta F(Z_{map}, t)} dZ_{map}}{\int_3^{35} e^{-\beta F(Z_{map}, t)} dZ_{map}} \right) \quad (10)$$

where  $\beta = (k_B T)^{-1}$  and  $F(Z_{map}, t)$  is the free energy as a function of the collective variable  $Z_{map}$  estimated by means of the reweighting algorithm at time  $t$ . The folded ( $Z_{map} \leq 3$ ) and unfolded ( $Z_{map} > 3$ ) states were defined on the basis of the localization of minimum 1 in Figure 2 and 3 (See main text).

The average values reported in Figure 4, 5, 6 and 7 of the main text were calculated as weighted time averages as follows:

$$\bar{x} = \sum_i^{i=T} \omega_i \langle x \rangle_i \quad (11)$$

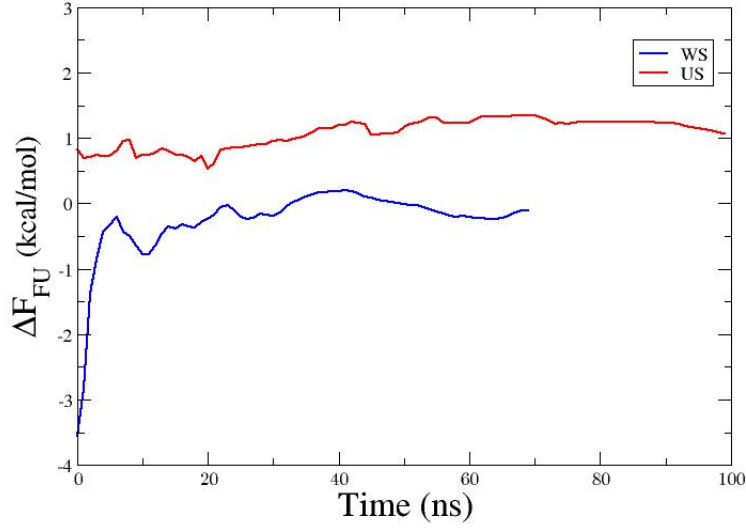


Figure S3: Convergence plot of  $\Delta F_{FU}$  as a function of sampling time. Red line for  $US$  and blue line for  $WS$ .

$$\langle x \rangle_i = \frac{\int x e^{-\beta F(x; t_i)}}{\int e^{-\beta F(x; t_i)}} \quad (12)$$

where  $F(x; t_i)$  is the free energy dependence on  $x$  at simulation time  $t_i$  computed using the reweighting algorithm. We collected the reweighted free energy every 1 ns. On the basis of the convergence properties of the reweighting algorithm,<sup>6</sup> we used  $\omega_i \propto t$ . The error bars were calculated as the standard deviation on the weighted mean  $\bar{x}$ , as a function of time

$$\sigma = \sqrt{\frac{\sum_{i=1}^{i=T} \omega_i (\langle x \rangle_i - \bar{x})^2}{\sum_{i=1}^{i=T} \omega_i}} \quad (13)$$

## References

- (1) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
- (2) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 20603.

- (3) Barducci, A.; Bonomi, M.; Parrinello, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826–843.
- (4) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (5) Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15317–15322.
- (6) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615–1621.