Supporting Information

Moradi et al. 10.1073/pnas.0906500106

Free Energy Methods and Simulation Details

To calculate accurate free energy maps, we used the recently developed adaptively biased molecular dynamics (ABMD) method (1), with supplementary equilibrium umbrella sampling runs (2). ABMD belongs to the general category of umbrella sampling methods with a time-dependent potential. It is characterized by only two input parameters (a flooding timescale τ_F and the kernel width $4\Delta\xi$) and has been implemented with replica exchange and multiple walker-based enhancements in the latest version of the AMBER software package. In addition, steered molecular dynamics (SMD) runs were used to further investigate specific pathways on these free energy maps (3). Because the details of the theory and implementation of the ABMD and SMD methods are found in the literature, we restrict ourselves here to a very brief discussion of our use of SMD runs and all the associated simulation details.

We have employed the SMD method (3) to examine and compare several pathways between selected peptide configurations. The SMD method is based on the following ideas. Consider a thermodynamic process that changes a system by means of a control parameter λ from λ_0 at time $t = 0$ to λ_t at time t. The second law of thermodynamics states that the work, W, done on the system cannot be smaller than the free energy difference:

$$\Delta F = F(\lambda_t) - F(\lambda_0) \leq W, \qquad [1.1]$$

with the equality holding if the process is quasistatic. Thus, a nonequilibrium process can only provide us with an upper bound for the free energy difference. However, the remarkable Jarzynski (4) identity is an exact equality that holds regardless of the speed of the transformation:

$$\exp\left(-\frac{\Delta F}{k_B T}\right) = \left\langle \exp\left(-\frac{W}{k_B T}\right)\right\rangle; \qquad [1.2]$$

the angular brackets denote the average over all possible realizations of the process connecting the states λ_0 and λ_t.

In practical SMD simulations, a restraining harmonic spring potential is typically added to the Hamiltonian:

$$U_{rest}(t) = \frac{K}{2}\left[\lambda(t) - \sigma(\mathbf{r}_1, \ldots, \mathbf{r}_N)\right]^2, \qquad [1.3]$$

so that the system can be "steered" towards the states with the desired value of the collective variable $\sigma(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ by "pulling" on the other end of the spring $\lambda(t)$. The stiffness of the spring, K, must be large enough as to well approximate the $K \to \infty$ limit. In this work the control parameter $\lambda(t)$ was varied with constant velocity along a prescribed pathway whereas the pathways were approximated by Catmull–Rom splines.

The method is especially easy to implement with 1D collective variables: one simply repeats the same run with different initial configuration and/or random seeds and averages the exponential of the work over those runs. For a multidimensional control parameter, the formalism additionally requires to average over all possible paths that join the points $\lambda(0)$ and $\lambda(t)$. This ••• is rather difficult to achieve, and therefore only a few of the most probable pathways are typically examined in a practical setting.

In this work, the pathways were chosen in several ways. First, we used the ABMD free energy maps to numerically identify the most likely transition paths by using the lowest free energy path (LFEP) method (5). Second, we used a 1D collective variable Ω for steering. Third, we selected pathways along the (Ω, Λ) landscape. We emphasize that here we do not aim to compute accurate free energy differences by using the SMD method; we rather use it as a tool to examine and compare several pathways associated with different transition mechanisms in a qualitative way.

We now turn to all the simulation details. Simulations of the polyproline peptide were carried out both in vacuo and in implicit (6–9) and explicit solvent environments. Initial configurations consisted of the unfolded peptide, which was generated by using the LEAP program of the AMBER-9 simulation package. The simulations used the 1999 version of the Cornell et al. force field (10). The leap-frog algorithm with a 1-fs time step was used along with the Berendsen thermostat. Hexamer, nanomer, and tridecamer molecules were simulated in vacuo and in implicit water, with no cutoff for the nonbonded interactions. Similar simulations (but only for the 9-mer polyproline) were performed in explicit solvents, namely hexane and 1-propanol. The peptide was bathed separately in 426 molecules of the nonpolar solvent hexane and in 912 molecules of 1-propanol. The explicit solvent calculations were carried out in a truncated octahedron cell with periodic boundary conditions with the total bounding box for atom centers being 79.800 Å in all directions. The time step and temperature control was the same as for the in vacuo implicit simulations. The PME method (11, 12) was used with cubic spline approximation and a cutoff of 8.0 Å, a nonbonded "skin" of 1.0 Å, a direct sum tolerance of 10^{-5}, and a grid size of 64 in each dimension.

Before starting free energy calculations, preliminary SMD runs were used to ensure the feasibility of transitions from right-handed polyproline helices (PPI) to left-handed polyproline helices (PPII). All the subsequent ABMD simulations were carried out by using 24 replicas. Initial calculations of the free energy landscapes in vacuo and in implicit water involved 24 replicas at the same temperature $T = 1200$ K without any exchanges between the replicas. These short and inaccurate multiple-walker ABMD runs of 10-ns duration provided us with a rough estimate of the biasing potentials to be encountered during the next phase of the runs. As shown in Fig. S1, the walkers have already covered all of the relevant phase space in the short period of time. Although the resulting free energy surface was found to be rather crude, it was already possible to discern the main minima associated with the different folded structures.

The second step in the simulation process involved long ABMD runs with replica exchange. We used the results from the first step. However, this time the temperature of all the replicas was varied from 300 to 1200 K, distributed as follows: 300, 318, 338, 359, 381, 405, 430, 457, 485, 516, 548, 582, 618, 656, 697, 740, 786, 835, 887, 942, 1001, 1063, 1129, and 1200 K. Each replica had its own biasing potential. Although the free energy surface for the highest temperature did not change very much, more and more details of the lower temperature maps emerged. Fig. 1 shows that though after 10 ns the phase space is only partially covered, it is completely covered after 80 ns. For most of these simulations, a kernel width of $4\Delta\xi = 0.5$ (for both-handedness and the R_g, in which case it is in Å units) and a flooding time scale of $\tau_F = 5.0$ ps were used.

Finally, we repeated the simulations by using the calculated biasing potential for follow-up equilibrium umbrella sampling runs. The kernel estimator method was then used to recover the biased probability density $p^B(H, R_g)$ (13). The typical follow-up correction run time was again 100 ns for most of the simulations but in some cases was extended to 150, 200, or 250 ns.

Turning to the explicit solvent simulation runs, these were performed by using the multiple-walker extension only. The multiple-walker extension was used exclusively because test runs showed that the parallel tempering extension did not prove to be very useful because of the very low measured rate of exchange between the different replicas. All the other parameters and steps were the same. We performed the simulations in hexane and propanol for 9-mer polyproline by using all three sets of collective variables. The ABMD run time as well as follow up correction run time varied from 30 to 100 ns for different sets of collective variables and different solvents.

As a final calculation, we performed PPI to PPII SMD simulations for the 6-mer and 9-mer polyproline peptides in vacuo to compare different transition pathways. We carried out these simulations (*i*) along two LFEP pathways obtained on the (H, R_g) planes; (*ii*) with the control variable chosen as the Ω order parameter; and (*iii*) along selected pathways in the (Ω, Λ) plane. The harmonic constant was set to 100 kcal/mol/X^2 (with X being Å or 1 depending on the nature of the control parameter) and the SMD run time varied from 10 to 100 ns in different simulations.

Puckering of the Proline Pyrrolidine Ring

The pyrrolidine ring is another distinct feature of proline, as compared with the rest of amino acids. This five-membered ring (see the inset in the left frame in Fig. S2) may adopt two different down- and up-puckered conformations (14), which have been known to be almost equally probable from the analysis of X-ray structures of peptides (15–17) and proteins (18–20). Although the correlations of the *cis-trans* population and the prolyl puckering were suggested by these studies, it is known that the *cis-trans* isomerization and the puckering transition is not coupled and that the prolyl ring can flip between the down- and up-puckered conformations with the peptide bond being in either *trans* or *cis* conformation (21). In particular, the conformational preferences of PPII and PPI structures, depending on their puckerings, have been studied (22–25). Recently, the barriers to the puckering transition between the two conformations were estimated to be about 3 kcal/mol in water (26).

In order to elucidate the puckering states and investigate their effect upon the "global" free energy landscapes studied in our paper, we simulated a proline dipeptide Ace-Pro-NMe in order to (*i*) check the time scales involved in the puckering interconversions to make sure that the latter get properly accounted for in our main computations and (*ii*) compute the free energy profiles associated with the puckering transition for both *cis* and *trans* conformations of the peptide bond. The implicit solvation model was employed for these simulations.

A natural collective variable that describes the two puckering states is the χ^1 torsion angle (26), as shown in the inset in the left frame of Fig. S2. The value of χ^1 is approximately 30° and -23° for the down- and up-puckered states, respectively.

To quantify the time scales for the interconversion, we performed regular molecular dynamics at $T = 300$ K. The time dependence of the χ^1 angle for both *cis* and *trans* conformations is shown in the right frame of Fig. S2. It can be seen that both puckering states get visited every $\approx 10 - 20$ ps for both *cis* and *trans* configurations of the proline peptide bond. Clearly, these states are adequately sampled over our simulation time scales, which is of the order of tens to hundreds of nanoseconds.

To investigate possible correlations between the *cis-trans* isomerization and the puckering transition, we also computed the free energy profiles associated with the angle χ^1 for the proline monomer in both *cis* and *trans* conformations at $T = 300$ K. The results that are shown in the left frame in the Fig. S2 agree with the ones reported in ref. 26 and confirm that the puckering landscapes are nearly identical for both peptide bond conformations.

In summary, we note that (*i*) the puckering states get sampled properly in our simulations because of the much shorter interconversion timescales (as compared with the much slower *cis-trans* isomerization and accompanying backbone motions) and (*ii*) the classical force field along with the implicit solvation model used in our study agrees with the more accurate ab initio results reported in ref. 26.

Results and Discussions

All of the results for the (H, R_g), (Ω, R_g), (H, d_{end}), (Ω, d_{end}), and (Ω, Λ) free energy maps were obtained via ABMD runs with follow-up corrective umbrella sampling runs (2). Generally, both here and in the paper, only the final, corrected results are shown. Fig. S3 shows an example of "raw" (H, R_g) free energy data and the "corrections".

The free energy maps for 6,9,13-mer in vacuo and implicit water are given in the main paper (Fig. 2). Table S1 gives details of the different minima (labeling the same as in main text Fig. 2*A*) associated with the in vacuo results.

There are many alternative ways to quantify the spatial extents of the different conformations of polyproline molecule. In practice, either the radius of gyration R_g (as is common in the polymer literature) or the peptide end-to-end distance d_{end} is typically used. Although the main text (Figs. 2 and 3) shows the maps based on R_g, (Figs. S4 and S5) present corresponding figures in terms of d_{end}. As expected, these free energy maps show similar features as the equivalent R_g maps; i.e., both demonstrate the existence of three classes of minima associated with PPI-like, PPII-like, and more compact globular structures.

For completeness, Fig. S6 shows (Ω, d_{end}) and (Ω, Λ) free energy data for 6-mer (*Top*), 9-mer (*Middle*), and 13-mer (*Bottom*) in vacuo, whereas Fig. S7 shows a similar figure based on R_g. The latter data is to be contrasted with the implicit water data shown in the main text (Fig. 3).

Fig. S8 shows 1D free energy profiles for Γ and in the $N_{\Omega,\Lambda}$ space in implicit water for a hexamer and nanomer.

With regard to the SMD runs, other trajectories—aside from the ones described in the paper—have also been analyzed and will be reported on in the future.

1. Babin V, Roland C, Sagui C (2008) Adaptively biased molecular dynamics for free energy calculations. *J Chem Phys* 128:134101.
2. Babin V, Roland C, Darden TA, Sagui C (2006) The free energy landscape of small peptides as obtained from meta dynamics with umbrella sampling corrections. *J Chem Phys* 125:2049096.
3. Izrailev S et al. (1998) *Steered Molecular Dynamics. Computational Molecular Dynamics: Challenges, Methods, Ideas* (Springer, Berlin), pp 39–65.
4. Jarzynski C (1997) Nonequilibrium equality for free energy differences. *Phys Rev Lett* 78:2690–2693.
5. Ensing B, Laio A, Parrinello M, Klein M (2005) A recipe in the computation of the free energy barrier and the lowest free energy path of concerted reactions. *J Phys Chem B* 109:6676–6687.
6. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 112:6127–6129.
7. Hawkins GD, Cramer CJ, Truhlar DG (1995) Pairwise solute descreening of solute charges from a dielectric medium. *Chem Phys Lett* 246:122–129.
8. Hawkins GD, Cramer CJ, Truhlar DG (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J Phys Chem* 100:19824–19839.
9. Tsui V, Case D (2001) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 56:275–291.
10. Cornell WD, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J Am Chem Soc* 117:5179–5197.
11. Darden TA, York DM, Pedersen LG (1993) Particle mesh Ewald: An N log(N) method for Ewald sumsin large systems. *J Chem Phys* 98:10089–10092.
12. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593.
13. Silverman BW (1986) *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability (Chapman and Hall, Boca Raton, FL).
14. Momany FA, McGulre RF, Burgess AW, Scheraga HA (1975) Energy parameters in polypeptides. vii. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. *J Phys Chem* 79:2361–2381.
15. Balasubramanian R, et al. (1971) Studies on the conformation of amino acids. VI. Conformation of the proline ring as observed in crystal structures of amino acids and peptides. *Int J Protein Res* 3:25–33.
16. DeTar DF, Luthra NP (1977) Conformations of proline. *J Am Chem Soc* 99:1232–1244.
17. Madison V (1977) Flexibility of the pyrrolidine ring in proline peptides. *Biopolymers* 16:2671–2692.

18. Milner–White EJ, Bell LH, Maccallum PH (1992) Pyrrolidine ring puckering in cis and trans-proline residues in proteins and polypeptides: Different puckers are favoured in certain situations. *J Mol Biol* 228:725–734.
19. Pal D, Chakrabarti P (1999) Cis peptide bonds in proteins: Residues involved, their conformations, interactions and locations. *J Mol Biol* 294:271–288.
20. Vitagliano L, Berisio R, Mastrangelo A, Mazzarella L, Zagari A (2001) Preferred proline puckerings in cis andtrans peptide groups: Implications for collagen stability. *Protein Sci* 10:2627–2632.
21. Kang YK, Choi HY (2004) Cistrans isomerization and puckering of proline residue. *Biophys Chem* 111:135–142.
22. Tanaka S, Scheraga HA (1974) Calculation of conformational properties of oligomers of l-proline. *Macromolecules* 7:698–705.
23. Zhong H, Carlson HA (2006) Conformational studies of polyprolines. *J Chem Theory Comput* 2:342–353.
24. Tanaka S, Scheraga HA (1975) Calculation of the characteristic ratio of randomly coiled poly(l-proline). *Macromolecules* 8:623–631.
25. Kang YK, Jhon JS, Park HS (2006) Conformational preferences of proline oligopeptides. *J Phys Chem B* 110:17645–17655.
26. Kang YK (2007) Puckering transition of proline residue in water. *J Phys Chem B* 111:10550–10556.

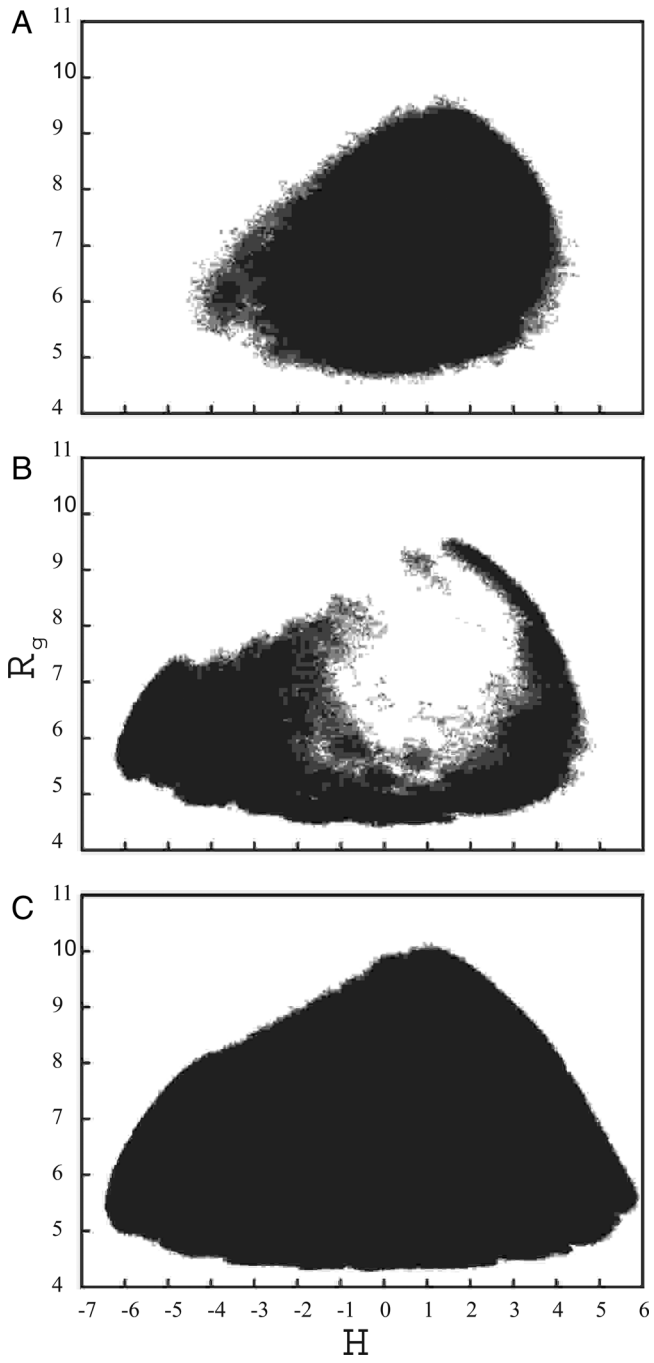

Fig. S1. MD trajectories in (H,R_g) configurational space. (*A*) Most of the phase space was covered quickly (in 10 ns) at the temperature $T = 1200$ K by multiple-walkers ABMD. (*B*) At $T = 300$ K, 10-ns parallel tempering ABMD was not enough to totally explore the phase space. (*C*) It took 80 ns for the 300-K replica to cover most of the phase space. (The ABMD parameters were $\tau_F = 5.0$ and $4\Delta\xi = 0.5$.)

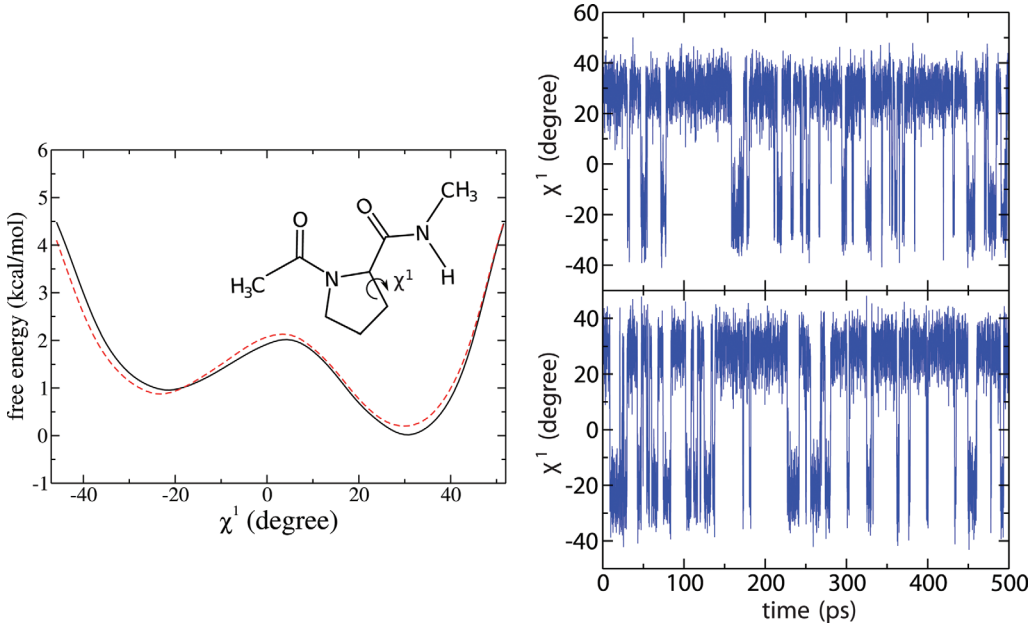

Fig. S2. ••• (*Left*) The Landau free energies associated with χ^1 angle at $T = 300$ K for Ace-Pro-NMe (shown in the *Inset*) in *cis* (solid line) and *trans* (dashed line) conformations. The torsion angle χ^1 is positive for the down-puckered state and negative for the up-puckered one. (*Right*) The χ^1 angle (see the *Inset* in the *Left* frame) in the course of regular molecular dynamics at $T = 300$ K for *cis* (*Upper*) and *trans* (*Lower*) configurations, for the Ace-Pro-NMe.

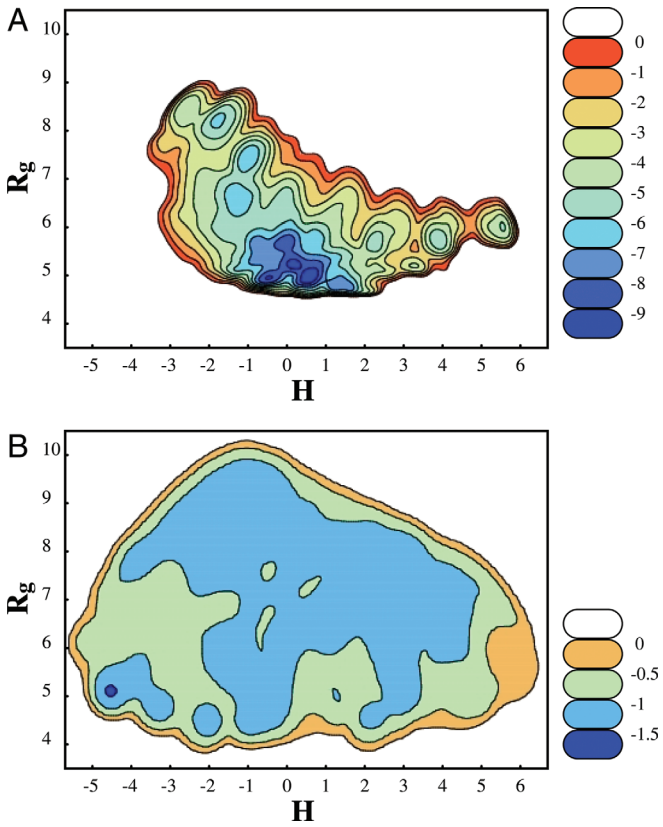

Fig. S3. Maps (kcal/mol) for a 9-mer polyproline in vacuo with handedness and radius of gyration as collective variables. (*A*) Contour plots of the estimated free energies representing the "raw" results from AMBD runs before the corrections. (*B*) Free energy corrections obtained from the biased probability density of equilibrium umbrella sampling runs for a 9-mer polyproline in vacuo. Handedness *H* and radius of gyration R_g were used as collective variables.

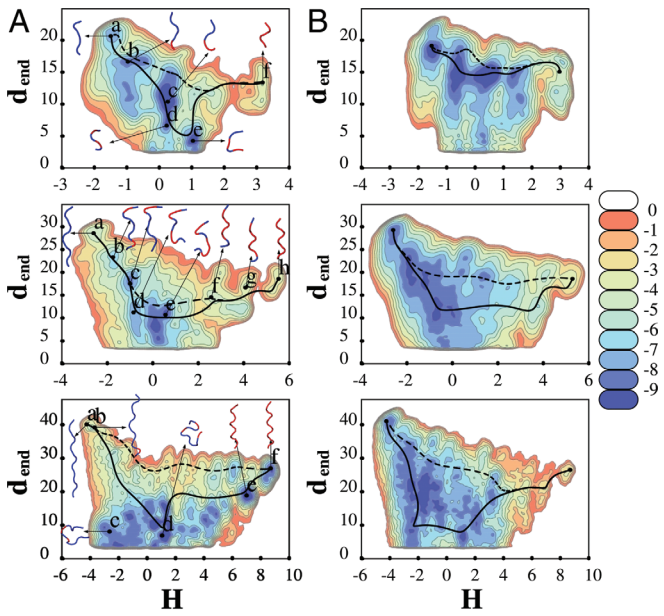

Fig. S4. • • • (*A* and *B*) Free energy landscapes (kcal/mol) as a function of the collective variables (H, d_{end}) for a 6-mer (*Top*), 9-mer (*Middle*) and 13-mer (*Bottom*) polyproline peptide in vacuo (*A*) and in implicit water (*B*). Examples of transition paths between PPI and PPII are shown: The solid line passes through the global minimum whereas the dashed line avoids it. A ribbon representation is used for some of the structures associated with the major minima, with *cis-trans* prolyl bonds highlighted in red (blue), respectively. In the 13-mer case, all the prolyl bonds of structures in both *E* and *F* are *cis*, but the prolyl amide bond in the amidated terminal of the peptide (not shown) is *trans* in *E* and *cis* in *F*.

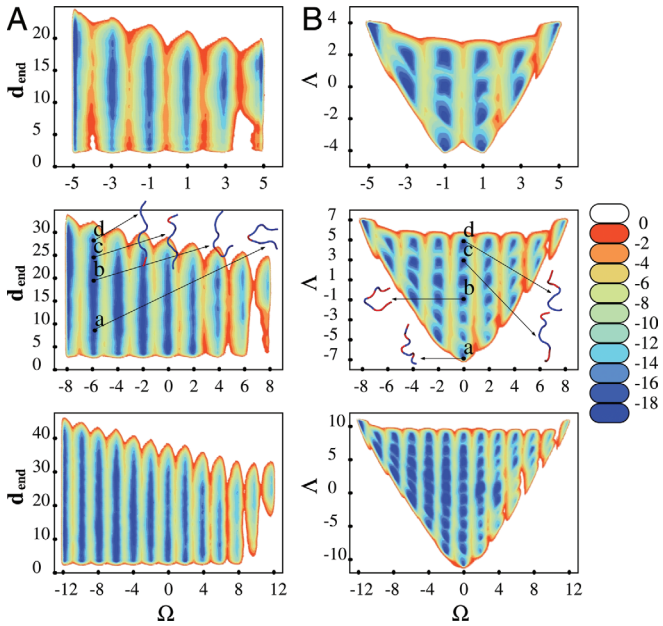

Fig. S5. Here are (Ω, d_{end}) and (Ω, Λ) free energy landscapes (kcal/mol) for a 6-mer (*Top*), 9-mer (*Middle*), and 13-mer (*Bottom*) polyproline in implicit water.

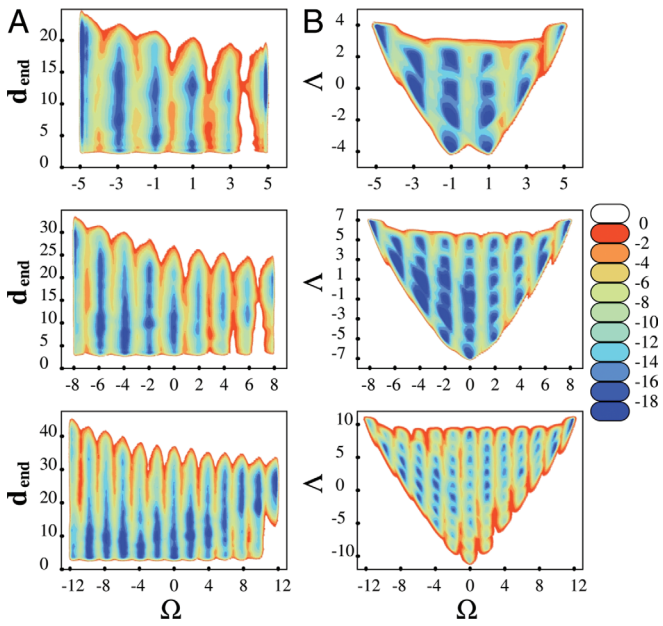

Fig. S6. Here are (Ω, d_{end}) and (Ω, Λ) free energy landscapes (kcal/mol) for a 6-mer (*Top*), 9-mer (*Middle*), and 13-mer (*Bottom*) polyproline in vacuo.

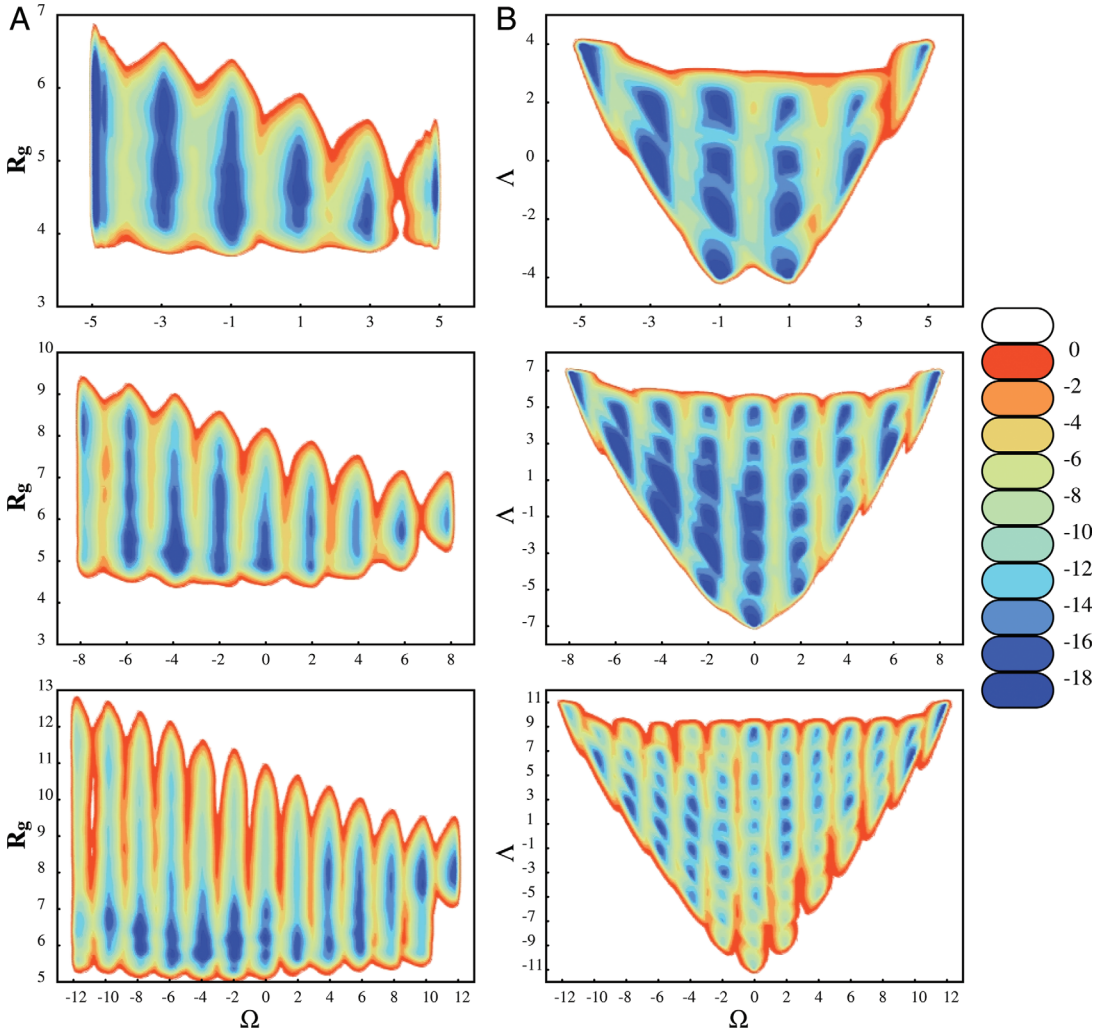

Fig. S7. Here are (Ω, R_g) and (Ω, Λ) free energy landscapes (kcal/mol) for a 6-mer (*Top*), 9-mer (*Middle*), and 13-mer (*Bottom*) polyproline in vacuo.

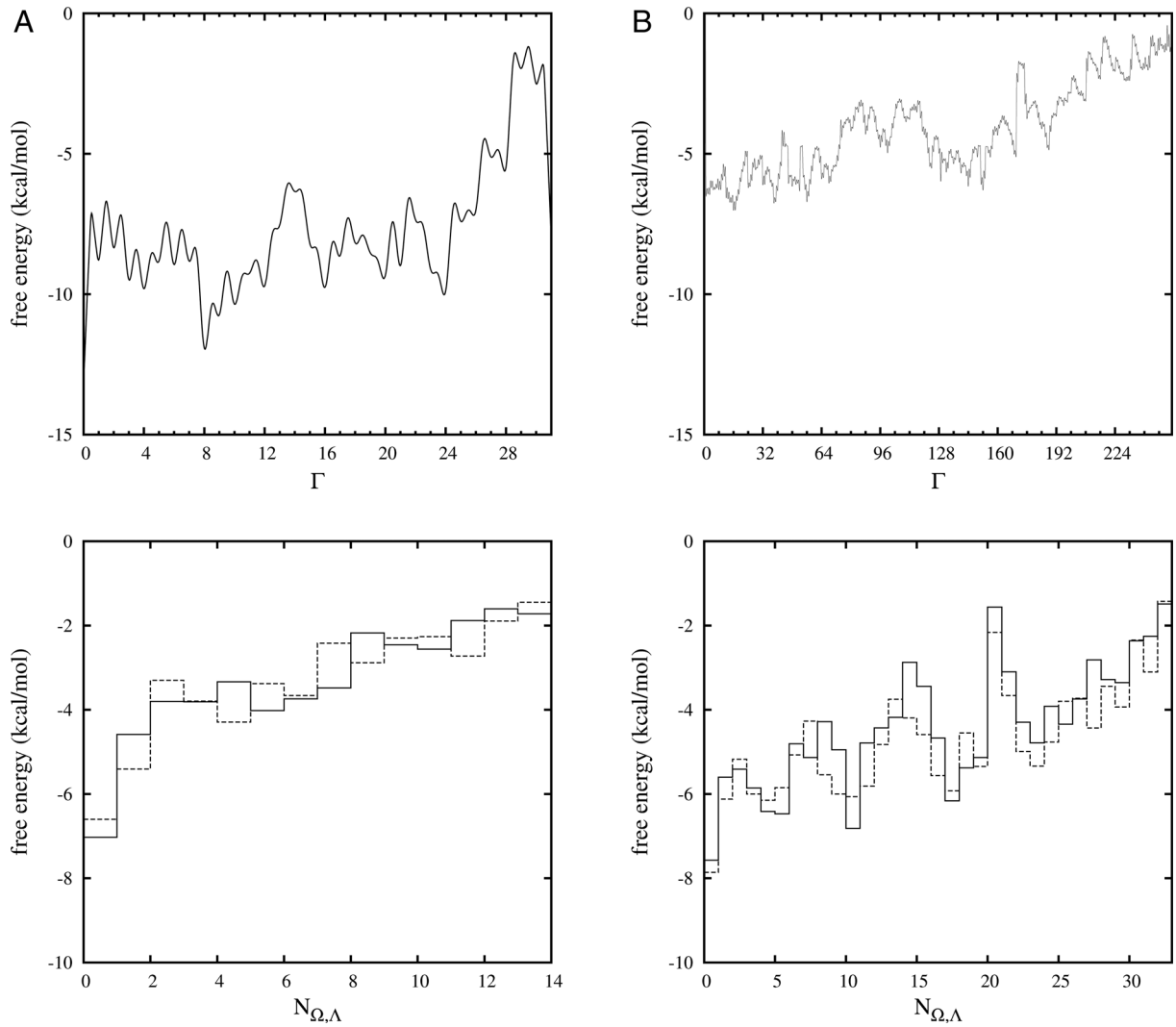

Fig. S8. One-dimensional free energy profiles in terms of Γ for polyproline in implicit water (*Upper*) for hexamer (*A*) and nanomer (*B*). The *Lower* frames compare the results obtained in terms of Γ and Eq. **6** with those obtained from the (Ω, Λ) by mapping onto the $N_{\Omega,\Lambda}$ phase space.

Table S1. Values of handedness H, radius of gyration R_g, free energy FE, and sequence of bond conformations (C = *cis*, T = *trans*), for 6-, 9-, and 13-mer polyproline peptides in vacuo.

Label	H	R_g (Å)	FE (kcal/mol)	Pattern
		6-mer		
a	−1.49	5.97	1.95	TTTTT
b	−0.71	5.65	0.89	TTTTC
c	0.13	4.87	0.00	TTTCC
d	−0.08	4.29	1.17	TCTCT
e	0.97	4.27	1.47	TTCTC
f	3.09	4.53	7.07	CCCCC
		9-mer		
a	−2.54	8.43	4.64	TTTTTTTT
$SP_{a\to b}$	−2.19	8.25	5.32	
b	−1.74	8.19	3.46	TCTTTTCT
$SP_{b\to c}$	−1.39	7.77	6.03	
c	−0.93	7.47	2.89	TCCTTTTT
$SP_{c\to d}$	−1.19	7.01	4.01	
d	−1.19	6.47	3.05	TCCTTCCT
$SP_{d\to e}$	−1.03	6.02	4.09	
e	0.62	4.99	0.00	TCCTCCCT
$SP_{e\to f}$	1.92	5.51	5.58	
f	2.23	5.63	3.96	TCCCCCTT
$SP_{f\to g}$	3.24	5.60	7.83	
g	3.94	5.75	3.89	TCCCCCCT
$SP_{g\to h}$	4.74	5.96	8.49	
h	5.60	5.99	5.69	CCCCCCCC
		13-mer		
a	−3.99	11.82	8.59	TTTTTTTTTTT
b	−3.42	11.60	7.72	CTTTTTTTTCT
c	−2.94	6.26	0.00	CTTCTCTCCTCT
d	1.08	5.90	0.18	CCCCCCCCTCCC
e	7.11	7.80	1.10	CCCCCCCCCCCC
f	8.71	8.03	1.38	CCCCCCCCCCCC

Free energies of saddle points (SP) are included for the 9-mer polyproline. Labeling of minima is as shown on Fig. 2*A*. The FE values have been shifted so that the global minimum in each space corresponds to 0.0 kcal/mol. Because of the degeneracies associated with (H,R_g) configurational space there are other sequences associated with some of the minima, which are not included in the table.