Tilting the Balance between Canonical and Non-Canonical Conformations for the H1 Hypervariable Loop of a Llama VHH through Point Mutations

Sai Pooja Mahajan[§], Camilo Velez-Vega[†], and Fernando A. Escobedo[§]*

§Department of Chemical Engineering, Cornell University, Ithaca, New York, 14850

†Department of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093

*fe13@cornell.edu

Additional simulation details: Thermodynamics

In BE metadynamics[1], multiple molecular dynamics (MD) simulations (replicas) are run simultaneously, each biased with the metadynamics potential in one or two collective variables. Every few MD steps, a Monte-Carlo (MC) exchange-move is attempted based on:

$$\min\left\{1, \exp\left[\frac{1}{T}\left(V^{a}{}_{G}(x^{a},t) + V^{b}{}_{G}(x^{b},t) - V^{a}{}_{G}(x^{b},t) - V^{b}{}_{G}(x^{a},t)\right)\right]\right\} \quad \text{(S1)}$$

where x^{a} and x^{b} are the coordinates of replica a and b and $V^{r}{}_{G}(x,t)$ is the metadynamics potential acting on replica $r(a/b)$. This treatment allows for an efficient conformational search in a high-dimensional collective variable space without the requirement of biasing each simulation in more than two collective variables. The MC exchange step enhances conformational search in a fashion similar to that in REM. From each simulation, one obtains low-dimensional projections of the free-energy which can then be used to obtain an estimate of the higher-dimension free-energy in CVs of interest. The low-dimensional free-energy projections after a sufficiently long time have been shown to fluctuate around an

average which estimates the FE of the system in the chosen CV space. However, this is only true when the choice of CVs ensures that all "slow" degrees of freedom that may be important to the conformational transitions in the system of interest have been sampled.

Calculation of percentage of conformations belonging to different clusters

The percentage of conformations belonging to a particular cluster is found by calculating volume under the FE curve in a three-dimensional CV space formed by CVs d_1, ψ_{S30} and ψ_{T31}. The limits for the integration were determined by the largest cube encompassing all conformations with an FE under $3k_BT$ from the FE minima for that cluster.

Additional Simulation details: Kinetics

The analytical solution is based on the equations derived for the milestoning[2, 3] approach. The cited method can be approximately mapped into ours by identifying their milestones with our "bins" in CV space. The authors derive these equations for the case when the jumps between the bins can be described as a Markov jump process, such that the probability of a system to jump from state i at time t, to state j at time Δt, is independent of all states before i. The rate of such a transition can be described in terms of rates q_{ij} for each pair of bins. Then the master equation of the Markov jump process can be written as,

$$\frac{d\rho(t)}{dt} = Q\rho(t) \qquad (S2)$$

where $\rho(t)$ is a row vector in which row i is the probability of finding the system in state i at time t and Q is the rate matrix such that its off-diagonal entries are given by q_{ij} and diagonal entries are given by

$$-\sum_{j \neq i} q_{ij} \; .$$

To obtain an estimate of the mean first passage times, the authors derive an analytical expression which only depends on Q:

$$\hat{Q}T^N = -\mathbb{1} \qquad\qquad (S3)$$

where $\hat{Q}$ is the N-1 $\times$ N-1 matrix obtained by deleting the row and column corresponding to the Nth milestone (bin) rendering it a cemetery state, T^N is a column vector with N-1 entries, i^{th} entry corresponding to the mean first passage time (MFPT) of the ith milestone to the Nth milestone. It is also shown that if one uses optimal milestones (i.e. milestones in which successive transitions are statistically independent), equation S3 gives the exact expression for the mean first passage time to milestone T^N.

We use this approach to obtain MFPTs for transitions among different basins in the CV space, assuming that for a certain optimal grid spacing the transitions become Markovian. Accordingly, we found the maximum grid spacing for which the ratio of transition rates between different basins in a high-dimension CV space remains constant. For each grid-spacing, the FE for each microstate was evaluated as the average over all conformations that lie within the 3D cube formed by the 3 CVs. As the grid-spacing increases, we average over larger regions in the 3D CV space, obtaining increasingly higher coarse-grained estimates of the FE. When the grid spacing is coarse-grained further (potentially erasing important local features of the landscape), we see a reversal of trend in the rates of transition between different basins and the analytical, and KMC results start diverging. We therefore aim to choose a grid spacing which is sufficiently small to accurately describe the changes in the FE in the chosen CV space and sufficiently large such that "fast" motions within each bin can be ignored. We tested this approach with a model potential (Muller potential), which has also been used in a previous work[3]. We were able to obtain identical results from KMC and equation S3 using rate constants between microstates (or bins) of the form 2 (in the main text). By recording the sequence in which the events occur in the KMC simulation, we generate an ensemble of pathways for the transition.

Validation of Force field

To validate the use of the CHARMM27-CMAP[4, 5] force field in explicit solvent for our system, we conducted molecular dynamics and BE metadynamics simulations of the VHH Cab-CA05[6, 7] (PDB code 1F2X). This VHH exhibits the type-1 canonical conformation of the H1 loop. We used the pre-folded framework approach as described in the main article. The FE profiles obtained from BE metadynamics are shown in Figure S1. These simulations were started from the type-1 canonical conformation. The FE profiles confirm the presence of a deep basin corresponding to the stable canonical conformation.
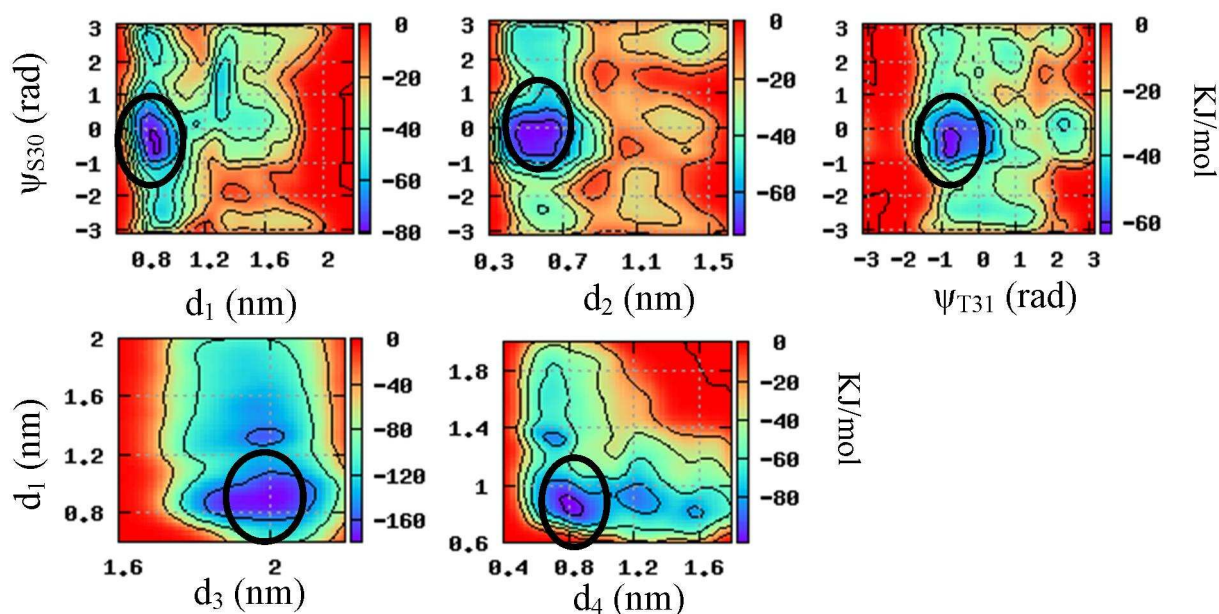

Figure S 1 2-D FE profiles for CAB-CA05 VHH. Black circles mark basins corresponding to the type-1 canonical conformation

FE profiles in all 6 CVs

In the main text, we present only the most informative FE profiles. However, all 6 CVs are important for the purpose of conformational search for the type-1 H1 canonical conformation. We present these additional FE profiles in Figures S1 and S2. Also, all FE profiles for the mutant FFTL, omitted in the main text, are presented here.

Additional details on the type-1 canonical conformation observed for FF, FFSa and FFD

The FE profiles show that a single mutation at position 52a (cases FFSa and FFD) is sufficient to generate additional free-energy minima in the d_1 vs. d_3 and d_1 vs. d_4 profiles in the FFD mutant (Figure S2a, S2b). As explained in the main text, d_3 captures the interaction between the H2 and H4 loops, which can hinder the burial of the F29 side-chain. A small d_3 value implies that the side-chains of residues at positions 52a (on H2) and 76 (on H4) are close to each other and the F29 side-chain cannot bury in the hydrophobic cavity formed by H1, H2 and H4. On the other hand, a small d_1 value and a large d_3 value indicate that the F29 side-chain is buried. The distance d_4 captures the role of the residue at position 71 on the H4 loop. The creation of additional basins in the corresponding CV space is captured in the free-energy profile.

Figure 5a, b further illustrates the difference in the type-1 canonical basins for FFD and FFSa due to the relative positioning of side-chains 52a and 71. The R71 side-chain lies between the H1 and H2 loops in the FFD type-1 H1 canonical basin, whereas it is exposed to water in the FFSa mutant. The location of the R71 side-chain for FFD is similar to the one observed in presence of a type-3 H2 loop. However, for FFD, the H2 loop conformation does not change (remains a type-2) but the R71 side-chain position does. Interestingly, one of the reasons for studying the W52aD mutation was to allow the possibility of a type-3 H2 loop conformation. We also note that in the implicit solvent study, R71 in FFD exhibits the same position as that in FFSa. There could be several reasons for this discrepancy, mainly 1) conformational sampling may not complete in the present study, possibly due to a missing CV that is important for the positioning of the R71 side-chain, 2) insufficient simulation time, and 3) discrepancies between the potential used here and the one used in our implicit solvent study, that may lead to the preferential stabilization of different states. On the other hand, the implicit solvent model used in our prior study may not have correctly captured the role of water molecules in the cavity formed by H1, H2 and H4 loops.
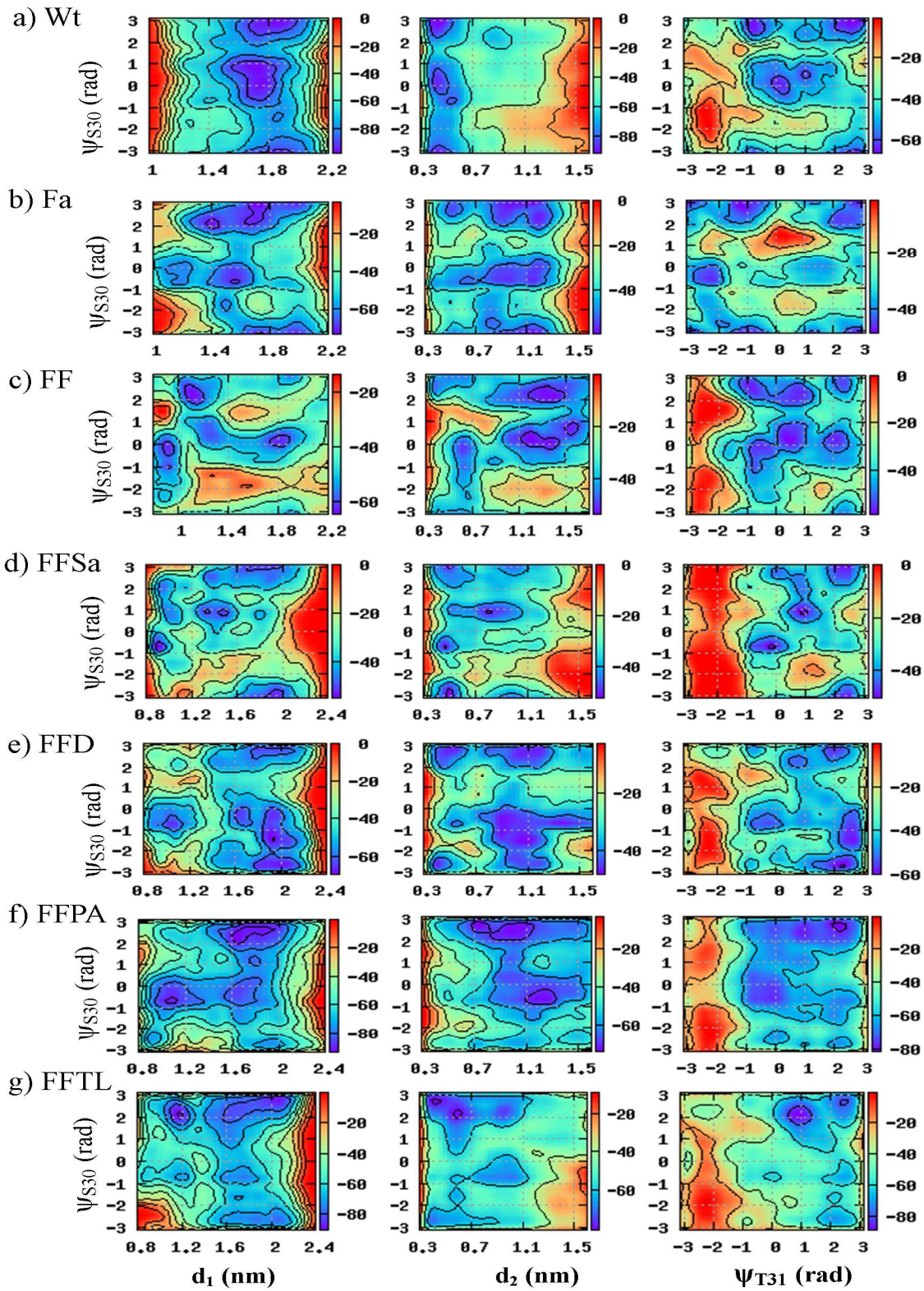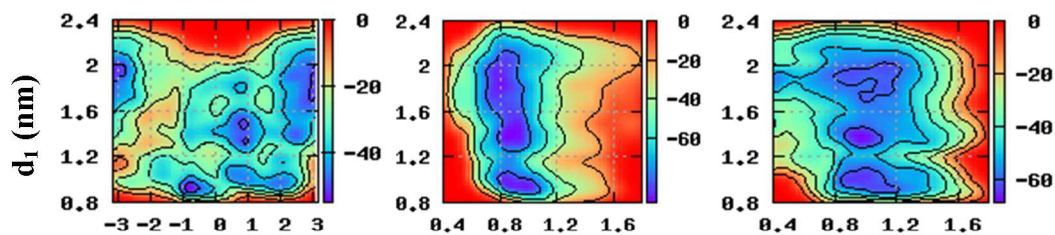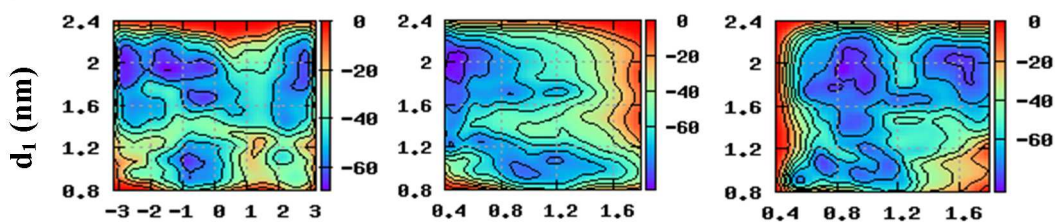
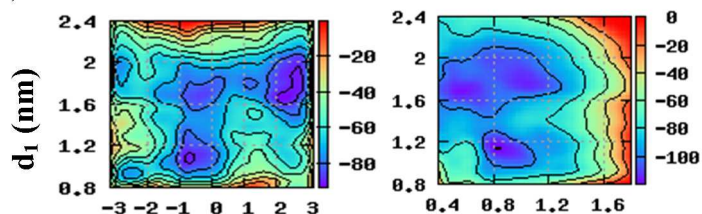**Figure S 2 2-D FE profiles for Wt and mutants in CVs d_1, ψ_{S30}, ψ_{T31}**

a) FFSa

b) FFD

c) FFPA

d) FFTL

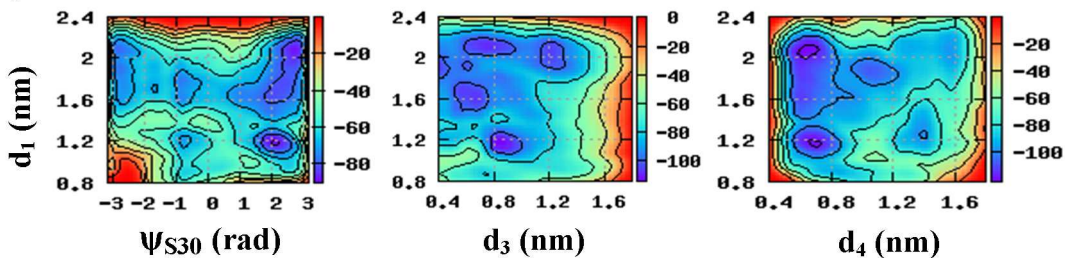ψ_{S30} (rad) d_3 (nm) d_4 (nm)

Figure S 3 2-D FE profiles for mutants FFSa, FFD, FFPA and FFTL in CVs d_1, ψ_{S30}, d_3 and d_4

Isolated H1 loop simulations

For these simulations, the isolated H1 loop is defined as comprising residues 25-34 with backbone restraints for residues 25 and 34. Residue 33, which is restrained in the reduced model,

was unrestrained here (the N and C terminals were capped by acetyl and N-methyl groups, respectively).

The mutants considered in this study have four possible H1 loop sequences – Wt, Fa (R27F mutation), Fb (G29F mutation) and FF (R27F and G29F mutations). We investigated the motion of these loops when they are free in solution (i.e., "isolated" from the rest of the antibody), by conducting a 60-ns unbiased MD simulation for each case.. The cluster analysis tool, g_cluster, available with GROMACS, was used to determine the principal conformations that each of the isolated H1 loops adopts in these simulations; the results are summarized in Fig. S4.
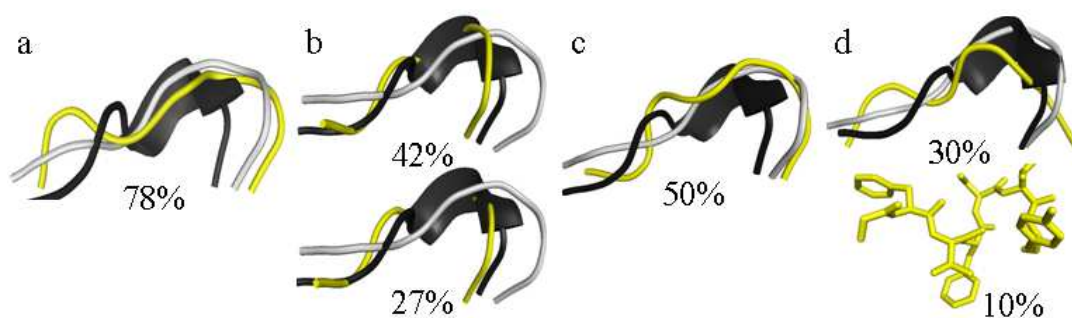

Figure S 4 Representative conformations of the isolated H1 loop, obtained via clustering analysis (yellow). a) Wt b) Fa, two main clusters c) Fb d) FF, Top - most populated cluster, bottom – cluster with F29 side-chain forming a kink. The percentages below each conformation indicate approximate cluster sizes. Each of these representative structures is compared to the 1HCV H1 loop (gray) and the type-1 canonical loop (black)

There is one major cluster obtained from the Wt loop simulations (Figure S4a), whose representative conformation is structurally similar to the Wt non-canonical conformation. Concerning the Fa mutant, its three main clusters are similar to the type-1 canonical conformation (Figure S4b, only the two most populated clusters are shown). Concerning the Fb mutant, there is no evidence of an early formation of a type-1 canonical conformation or a 3-10

helical structure (Figure S4c). Finally regarding the FF mutant, the turn/kink at the F29 position is observed in as early as 6 ns, though this conformation (Figure S4d, bottom) is found only during the initial 11 ns of the simulation and represents 10% of all conformations in the unbiased MD simulation, possibly due to the absence of a hydrophobic core. The FF mutant loop preferentially adopts extended conformations in the remaining part of the simulation. Hence, for some of these mutants, the H1 loop may tend to partially "fold" into a more compact structure. This is further supported by metadynamics simulations of the FF isolated H1 loop which shows that the most stable basin corresponds to a conformation with a well-defined kink at position 29 (see further details in he next paragraph), suggesting that the "burial" of the F29 residue may not be solely driven by the enthalpic gain from the de-solvation of a hydrophobic side-chain. The kink at position 29 is also influenced by the sequence (and associated torsional potentials) of residues at positions 29, 30 and 31. Overall, our results for the isolated H1 loops simply suggest that different mutants have a different intrinsic propensity towards canonical or non-canonical conformations, which of course can be overridden by interactions with proximal residues in the full VHH molecule.

We used BE metadynamics to further explore the behavior of an isolated H1 loop of mutant FF. We use 4 CVs for this case – d_2, ψ_{S30}, ψ_{T31} and an additional new CV not defined in the main text. This is the C_α-C dihedral of the ψ_{F29} residue. In the absence of the rest of the protein, we cannot define distances d_1, d_3 and d_4. In the FE-profiles in Figure S5, the most stable basin corresponds to a conformation in which a kink is observed at residue 29. This suggests that the H1 loop may have a propensity to partially "fold" to a conformation in which residues F29, S30 and T31 exhibit dihedrals close to the type-1 canonical conformation.
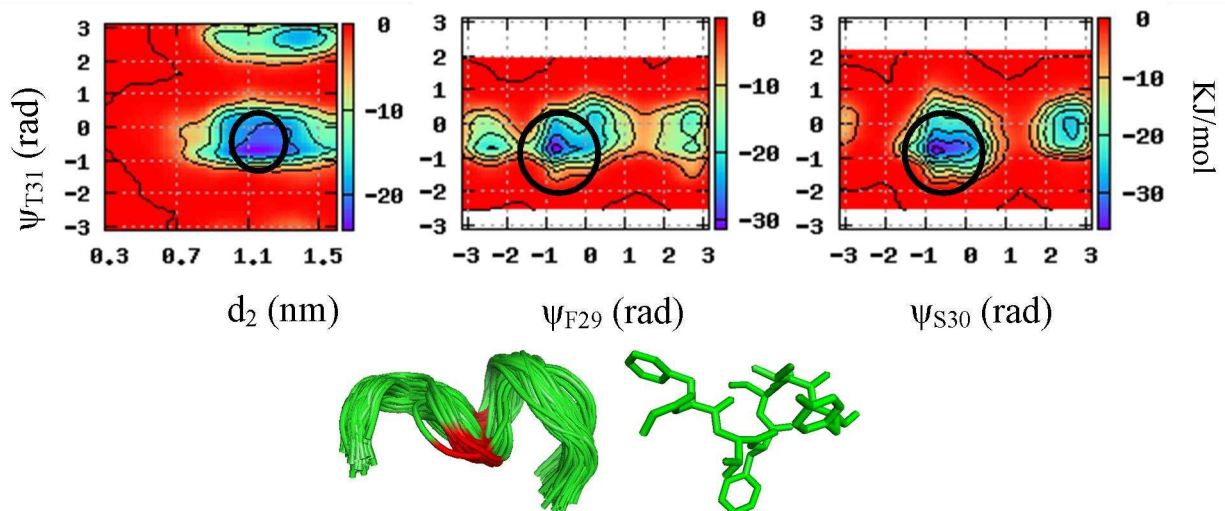
Figure S 5 Top - 2-D FE profiles for FF isolated loop. Bottom – Conformation corresponding to the most stable basin (circled) in cartoon (left) and stick (right) conformations. The kink at position 29 is highlighted in red.

References

(1) Piana, S.; Laio, A. *J Phys Chem B* **2007**, *111*, 4553-4559.

(2) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. *J. Chem. Phys.* **2008**, *129*, 174102.

(3) Vanden-Eijnden, E.; Venturoli, M. *J. Chem. Phys.* **2009**, *130*, 194101-194101-13.

(4) MacKerell et al. *J Phys Chem B* **1998**, *102*, 3586-3616.

(5) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. *J. Chem. Theory Comput.* **2010**, *6*, 459-466.

(6) Lauwereys, M.; Arbabi Ghahroudi, M.; Desmyter, A.; Kinne, J.; Holzer, W.; De Genst, E.; Wyns, L.; Muyldermans, S. *EMBO J.* **1998**, *17*, 3512-3520.

(7) Decanniere, K.; Muyldermans, S.; Wyns, L. *J. Mol. Biol.* **2000**, *300*, 83-91.