A solvation induced ring puckering effect in fluorinated prolines and its inclusion in classical force-fields

Ajay Muralidharan,* J.R. Schmidt, and Arun Yethiraj

Department of Chemistry, University of Wisconsin-Madison

E-mail: *amuralidhar4@wisc.edu

Abstract

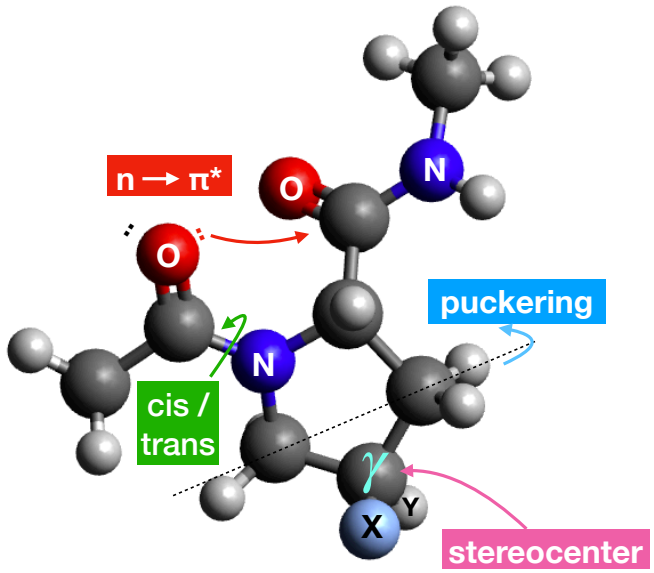
Strategic incorporation of fluorinated prolines can accelerate folding and increase thermal stability of proteins. It has been suggested that this behavior emerges from puckering effects induced by fluorination of the proline ring. We use electronic structure calculations to characterize the potential energy surface (PES) along puckering coordinates for a simple dipeptide model of proline and its fluorinated derivatives. Comparison of gas phase and implicit solvent calculations shed light on the effect of solvation on electronic structure and conformational preferences of the ring. This effect is unknown in the context of prolines, however, recently reported for furanoses in carbohydrates. The PES based on implicit solvent is then utilized to construct a correction for a classical force-field. The corrected force-field accurately captures the experimental conformational equilibrium including the coupling between ring puckering and cis-trans isomerism in fluorinated prolines. This method can be extended to other rings and substituents besides fluorine.

1 Introduction

The incorporation of non-cannonical amino acids has been an effective strategy for introducing novel properties into peptides and proteins.[1–3] A popular choice for this purpose is the use of fluorinated amino acid residues, especially, fluorinated prolines. This is due to the ability of highly electronegative fluorine to modify chemical and structural properties of polypeptide systems such as polarity, hydrophobicity, and secondary structure propensities.[4,5] As such, the development of various synthesis methods for introducing fluorine into proteins have expanded the utility of fluorinated amino acids in protein design. A recent study[6] highlighted the use of fluorinated prolines in an experiment where a single atom (fluorine) substitution accelerated the folding of a polypeptide chain (RNAase) chain and increased its thermostability. Such effects emerge due to the unique structure of proline, lending it an important role in structural biology.[7–10]

Proline is the only naturally occurring amino acid with a side chain α-amino group that connects back to the main chain forming a 5 membered pyrrolidine ring (Fig. 1). This structural feature leads to interesting conformational properties for the peptide backbone (ϕ, ψ and ω, Fig. 2). Firstly, most peptide bonds overwhelmingly adopt the trans isomer due to its higher stability. In contrast, the X-Proline peptide bond (where X represents any amino acid) is known to populate both cis and trans isomers. Secondly, the trans state is typically more stable than cis by $\approx$ 1kcal/mol. This is because of an interplay of weak interactions including electrostatics, dispersion and $n \rightarrow \pi^*$ interactions[11,12] between adjacent carbonyl groups. Since the $n \rightarrow \pi^*$ interaction interaction arises primarily from the proximity of the two carbonyl groups only the trans conformation of the peptide bond allows for such an interaction (Fig. 1). Finally, the 5 membered ring preferentially adopts a mixture of Cγ endo/exo puckers (Fig. 2).

The (endo/exo) ring puckering behavior in peptides containing proline is coupled with the cis-trans isomerism. This is because the exo pucker results in a favorable distance between the carbonyl groups in the trans state (compared to the endo pucker) leading to a stronger



Proline dipeptide & derivatives

atom X	atom Y	shorthand
H	H	HH
F	H	RF
H	F	SF

Figure 1: Dipeptide model of proline (and its fluorinated derivatives) depicting the cis-trans isomerism, $n \rightarrow \pi^*$ interaction and ring puckering effects. Due to coupling between these effects, the cis-trans conformational equilibrium of the peptide bond can be tuned by selecting the puckering preferences of the ring.

interaction.[11] Further, substitution of fluorine at the C_γ position is known to shift the relative stabilities of endo and exo states.[13] Hence, the cis-trans conformational equilibrium of the X-Proline peptide bond can be tuned (indirectly) by selecting the stereochemistry of F at the C_γ carbon.

A majority of simulation efforts[14–16] directed at addressing the conformational behavior of prolines and its derivatives utilize electronic structure methods restricted to gas-phase and implicit solvent. These methods do not include solvent degrees of freedom (DOF) explicitly due to the high computational cost. However, the influence of H-bonding[17] and charge fluxes induced by solvent have been shown to be crucial for treating conformational equilibria in furanose rings[18] (carbohydrates) and peptides containing cyclobutane.[19]

On the other hand, classical force-fields include explicit solvent but do not capture solute conformational properties accurately, especially, puckering of substituted rings. Incorporation of a pucker correction based on accurate quantum mechanical calculations could potentially fix classical force-fields and make the best of both worlds. In this work, we identify the

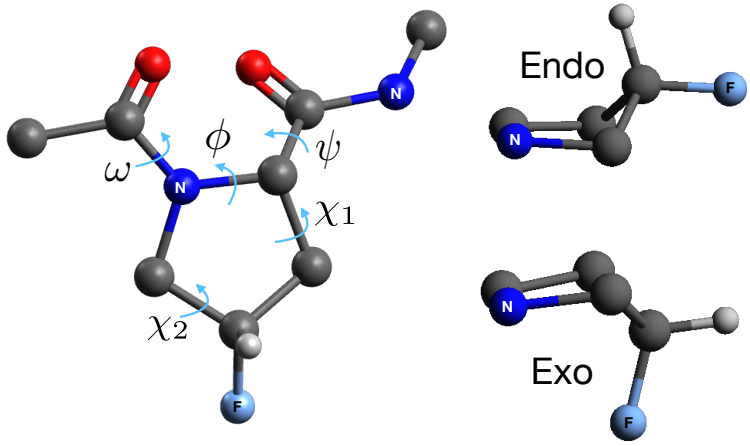


Figure 2: Left: Torsions specifying backbone (ϕ, ψ and ω) and pucker state (χ_1,χ_2) of the ring in (4S)-4-fluoroproline (SF in short). Some of the hydrogen atoms are suppressed here for clarity. Right: The ring puckering preference between endo (right top) and exo (right bottom) puckers can be controlled by selecting the stereochemistry of F at the C_γ carbon.

extent of solvation effects on ring puckering and employ a correction to include those effects in classical force-fields.

2 Methods

2.1 2D potential energy scan (PES) of the pyrrolidine ring:

High-level quantum mechanical methods are relatively successful compared to force-fields and semiempirical methods at describing the energetics of puckered ring conformations.[20] Hence, we perform relaxed energy scans at the MP2 level of theory for dipeptides of proline (HH in short), 4R-fluoroproline (or RF) and 4S-fluoroproline (or SF) using Gaussian 16.[21] A 6-31G(d,p) basis set is used for geometry optimization followed by a single point energy calculation with the 6-31++G(d,p) basis set.

For ring puckering coordinates, we adopt a definition used by Huang et al.[20] in the context of 5 membered furanose rings that requires only two proper endocyclic torsions within the ring:

$$Z_{\mathrm{X}} = \frac{\chi_1 + \chi_2}{2\cos(4\pi/5)} \tag{1}$$

$$Z_{\mathrm{Y}} = \frac{\chi_1 - \chi_2}{2\sin(4\pi/5)} \tag{2}$$

χ_1 and χ_2 are proper torsion angles shown in figure 2. Z_{X} and Z_{Y} can be interpreted as bending and twisting modes for the ring.[22] Here, (Z_{X}, Z_{Y}) is fixed at values ranging from $-60°$ to $60°$ in steps of $6°$ (441 optimizations), and all other degrees of freedom were relaxed to produce a 2D PES. We then obtain the difference in electronic energies, ΔE_{MP2}:

$$\Delta E_{\mathrm{MP2}} = E(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) - E_{min}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) \tag{3}$$

The energy differences are evaluated at MP2 level in gas phase ($\Delta E_{\mathrm{MP2/GP}}$) and implicit solvent phase ($\Delta E_{\mathrm{MP2/implicit}}$). The polarizable continuum model[23] (PCM) is used for implicit solvent calculations with a relative permittivity of $\epsilon \approx 80$. The resulting 2D PES is discussed for HH, RF and SF ($441 * 12 = 5292$ optimizations total) in section 3.2 and figure 4.

2.2 Force-field parameterization

For the canonical proline dipeptide (HH), the Amber-ff14sb force-field[24] is used to perform molecular dynamics (MD). For non-canonical derivatives of proline dipeptide, we adopt the strategy decribed in section 2.2.1 below. The TIP3P model is used for water in all MD calculations.

2.2.1 Non-cannonical aminoacid residues (RF and SF)

We begin with Amber-ff14sb for intramolecular bonds, angles and dihedral parameters that exclude fluorine atoms. Parameters for bonds, angles and dihedrals containing the fluorine atom are obtained from the general amber force-field[25] (GAFF).

Point charge representation:

Atomic partial charges are obtained following the Amber amino acids protocol.[26] In short, multiple conformations (4 per peptide) are used with restrained electrostatic potential (RESP)

fitting to derive the partial charges. The starting configurations (2 each of cis and trans) are obtained from Table 2 of Cieplak et al.[27] and optimized with HF/6-31G(d) theory and basis set in Gaussian16 software package.[21] The respgen program within Ambertools18[28] is then used to obtain partial charges from electrostatic potentials generated by Gaussian.

Dispersion:

Lennard-Jones (LJ) dispersion parameters are obtained from Amber-ff14sb for all atoms except fluorine (F) and the hydrogen (H_F) bonded to a fluorinated carbon. For these exceptions, the dispersion parameters are obtained from Robalo et al.[29] In short, the LJ parameters of F were optimized to reproduce the hydration free energies of CF_4 and the molar volume of a 50% mix of CF_4 and CH_4. Those of H_F were subsequently[30] optimized to reproduce the hydration free energy and the molar volume of CHF_3, CH_2F_2 and CH_3F. The Amber-ff14sb force-field combined with these dispersion parameters and point charges evaluated for non-canonical residues is dubbed as FF in further discussions.

Ring puckering correction:

It is known that classical force-fields and semi-empirical quantum theories do not capture the populations of ring puckered conformations accurately.[20] For correcting classical force-fields, we begin with an expansion of the required QM free energy difference between the puckered endo state (A) and exo state (B) in explicit water (QM/explicit),

$$\begin{aligned} & \Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{QM/explicit}) \\ = \; & \Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{QM/implicit}) + \{\Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{QM/explicit}) - \Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{QM/implicit})\} \\ \approx \; & \Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{QM/implicit}) + \{\Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{FF/explicit}) \; - \Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{FF/implicit})\} \end{aligned} \tag{4}$$

The approximation here is that the free energy difference specified within the curly braces is insensitive to QM or force-field description of the solute. Further, if the force-field in implicit solvent phase (FF/implicit) is corrected by fitting to the potential energy surface

(PES) from QM calculations (QM/implicit), then we expect a cancellation (Eq. 5). The PES fitting procedure also reinforces the approximation made in Eq. 4.

$$\begin{aligned} &\approx \cancel{\Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{QM/implicit})} + \left\{\Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{FF_C/explicit}) - \cancel{\Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{FF_C/implicit})}\right\} \\ &\approx \Delta G_{\mathrm{A}\rightarrow\mathrm{B}}(\mathrm{FF_C/explicit}) \end{aligned} \tag{5}$$

($\mathrm{FF_C}$/explicit) is the ring pucker corrected force-field that approximates the hypothetical (QM/explicit) reference. Equations 4 and 5 outline a theoretical construct for using implicit water calculations to obtain a solute intramolecular correction for use with explicit waters. The calculation and implementation of the correction term is described below.

Implementation of an intramolecular energy correction term:

In order to correct puckering trends in force-fields we introduce a correction term $C(Z_{\mathrm{X}}, Z_{\mathrm{Y}})$,

$$C(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) = \Delta E_{\mathrm{MP2/Implicit}}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) - \Delta E_{\mathrm{FF/Implicit}}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) \tag{6}$$

$C(Z_{\mathrm{X}}, Z_{\mathrm{Y}})$ accounts for difference in the implicit solvent PES between MP2 theory and force-field (FF). $(Z_{\mathrm{X}}, Z_{\mathrm{Y}})$ are coordinates that define the pucker state of the ring. $\Delta E_{\mathrm{MP2/Implicit}}$ is the PES evaluated with MP2/6-31++G(d,p) and the PCM[23] implicit solvation method. $\Delta E_{\mathrm{FF/Implicit}}$ is the PES evaluated with classical force-field (FF) and a modified GBSA[31] implicit solvent model. FF is Amber-ff14sb force-field, but with recalculated point charges for the non-canonical residues and optimized dispersion parameters for fluorine and hydrogen bonded to a fluorinated carbon.

The correction term, $C(Z_{\mathrm{X}}, Z_{\mathrm{Y}})$, is nearly identical for configurations obtained from cis and trans dipeptide cases and are averaged to produce $\overline{C}(Z_{\mathrm{X}}, Z_{\mathrm{Y}})$.

$$\overline{C}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) = \frac{C_{\mathrm{Cis}}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) + C_{\mathrm{Trans}}(Z_{\mathrm{X}}, Z_{\mathrm{Y}})}{2} \tag{7}$$

This correction is introduced as an external bias to molecular dynamics calculations by

using the PLUMED plugin.[32–34]

$$E_{\mathrm{FF_C}}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) = E_{\mathrm{FF}}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) + \overline{C}(Z_{\mathrm{X}}, Z_{\mathrm{Y}}) \tag{8}$$

Here, $\mathrm{FF_C}$ is the ring pucker corrected force-field. The potential energy scan evaluated with $\mathrm{FF_C}$/implicit agrees better with MP2 calculations when compared with FF/implicit (Figure 3). We also demonstrate (section 3.3) that $\mathrm{FF_C}$ quantitatively captures the experimental puckering behavior for fluorinated prolines.

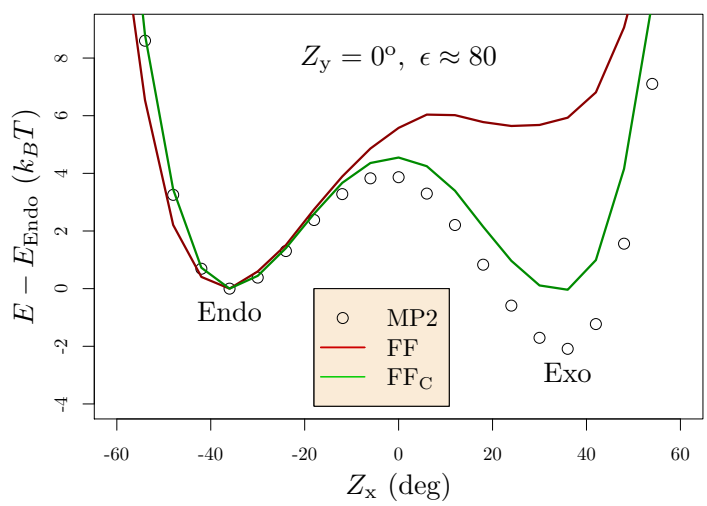


Figure 3: 1D Potential energy scan in implicit solvent ($\epsilon \approx 80$) for trans 4R-fluoroproline (RF) along Z_{x} coordinate. $Z_{\mathrm{y}} = 0°$. The modified GBSA model[31] was used with force-field and the PCM model[23] was used with MP2 calculations. The corrected force-field ($\mathrm{FF_C}$) agrees better with MP2 calculations. The full 2D PES is available in SI, Figure S2.

2.3 Molecular dynamics (MD) simulations

All molecular dynamics calculations utilized the gromacs simulation package. The calculations use a time step of 2 fs and adopt the Nosé thermostat procedure[35] at 298 K and a Parrinello-Rahman[36] barostat at 1atm.

Ring conformational equilibrium:

For evaluating the conformational equilibrium of the ring, we perform a 200 ns trajectory for each dipeptide (HH /RF /SF) in 500 waters, saving the Z_{X} puckering coordinate every 1 ps. From the probability distribution for Z_{X}, the free energy difference between endo and exo states is obtained and converted to relative populations (Table 1). The simulations are carried out using Amber-ff14sb for HH. For RF and SF we perform calculations using a force-field with ($\mathrm{FF_C}$) and without (FF) the ring pucker correction term for comparison.

Umbrella sampling for cis-trans equilibrium:

We perform umbrella sampling calculations to calculate the potential of mean force along the cis-trans torsion angle (ω, Figure 1). A system consisting of a single dipeptide molecule (HH/ RF/ SF) and 500 waters is simulated using Gromacs simulation package.[37] For the production runs, a harmonic force constant is applied to 41 windows centered at $\omega \in (-3.14, 3.14)$ in steps of 0.157 radian. The force constant is set to 1000 kJ $\mathrm{mol^{-1}rad^{-2}}$ for windows with a center between $(-2.2, -0.78)$ and $(0.78, 2.2)$ (near huge energy barriers). Otherwise, a value of 250 kJ $\mathrm{mol^{-1}rad^{-2}}$ is used. The PLUMED[32–34] plugin is used to apply the harmonic restraints. The calculation adopts the Nosé thermostat procedure[35] at 298 K and a Parrinello-Rahman[36] barostat at 1atm. A time step of 2 fs is chosen to realize a 200 ns trajectory per window, saving ω every 0.2 ps. Finally, the weighted histogram analysis method[38] (wham) is employed to obtain the potential of mean force curves from the histograms of ω. The last 75% of the trajectory (150 ns) is used for the wham analysis.

3 Results

3.1 Electronic structure calculations reveal subtle effects of solvation on ring puckering

Substitution of fluorine at the C_γ position is known to shift the relative stabilities of endo and exo ring states of proline rings. To quantify this effect, we performed relaxed energy scans in gas phase and implicit solvent phase for cis and trans dipeptides of HH, SF and RF along ring puckering coordinates. The puckering conformations of 5 membered rings are generally prescribed by 2 modes:[22] bending (Z_{X}) and twisting (Z_{Y}) as defined in section 2. We observed two energy minima (Figure 2), one at negative Z_{X} (endo) and the other at positive Z_{X} (exo). The energy difference (ΔE_{MP2}) was used to identify the energetically preffered puckering state.

$$\Delta E_{\mathrm{MP2}} = E(Z_{\mathrm{X}}, Z_{\mathrm{y}}) - E_{min}(Z_{\mathrm{X}}, Z_{\mathrm{y}})$$

where $E(Z_{\mathrm{X}}, Z_{\mathrm{y}})$ is the energy of any conformer and $E_{min}(Z_{\mathrm{X}}, Z_{\mathrm{y}})$ is the lowest energy conformer.

From experiments,[39] we know that the exo pucker is stabilized (relative to HH) for the case of RF and the endo pucker is stabilized for SF. A stereo-chemistry dependent gauche effect[40] between vicinal N and F atoms was considered to be the primary reason for this reversal in conformational behavior. Surprisingly in our gas phase MP2 calculations, the exo puckers were stabilized for both RF and SF (Figure 4, top panel). However, when the relaxed scans were performed with implicit water solvent, the trends qualitatively agreed with experiment (Figure 4, bottom panel). This indicates that besides the gauche effect, the influence of solvation on electronic structure is crucial to capture the relative conformational energies of the ring. This solvation induced effect is previously unknown in the context of prolines, however, recently reported for furanoses[18] in carbohydrates.

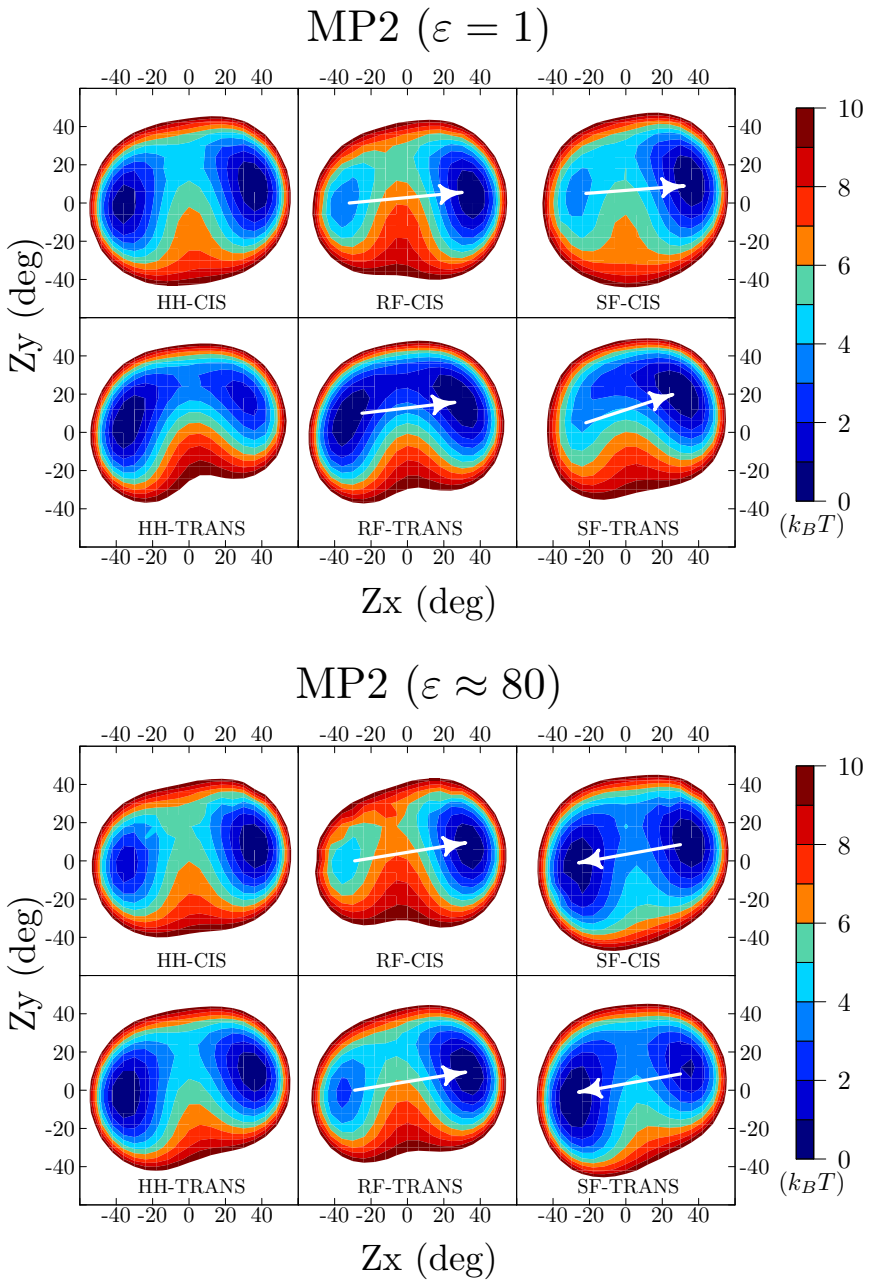


Figure 4: Relaxed energy scans for cis and trans dipeptides of proline (HH), 4R-fluoroproline (RF) and 4S-fluoroproline (SF) along collective variables Zx and Zy in gas phase (top, $\varepsilon = 1$) and implicit solvent (bottom, $\varepsilon \approx 80$). The arrows indicate the ring pucker preference in RF and SF relative to HH. The scans where done at MP2/6-31G(d,p)//MP2/6-31++G(d,p). In gas phase, introducing a fluorine stabilizes the exo ring pucker in RF and SF. However, in implicit solvent, the exo pucker is stabilized in RF and the endo pucker is stabilized in SF in agreement with experimental observations. This reversal in preferred conformations due to solvent is unknown in the context of prolines, however, recently reported for furanoses.[18]

3.2 A harmonic approximation based on implicit solvation fails to produce accurate free energy difference between cis(aq.) and trans(aq.)

We evaluated the free energy difference between cis(aq.) and trans(aq.) for proline dipeptide using a harmonic approximation (peptide experiences small harmonic vibrations about the minima) within an implicit solvent polarizable continuum model[23] (PCM). The free energy obtained this way ($-0.3\ k_BT$) is significantly less than the experimental value[41] for this conformational equilibrium ($\approx -1.6\ k_BT$). Hence, we constructed a thermodynamic cycle to evaluate individual legs of the cycle and investigate the possible sources of error.

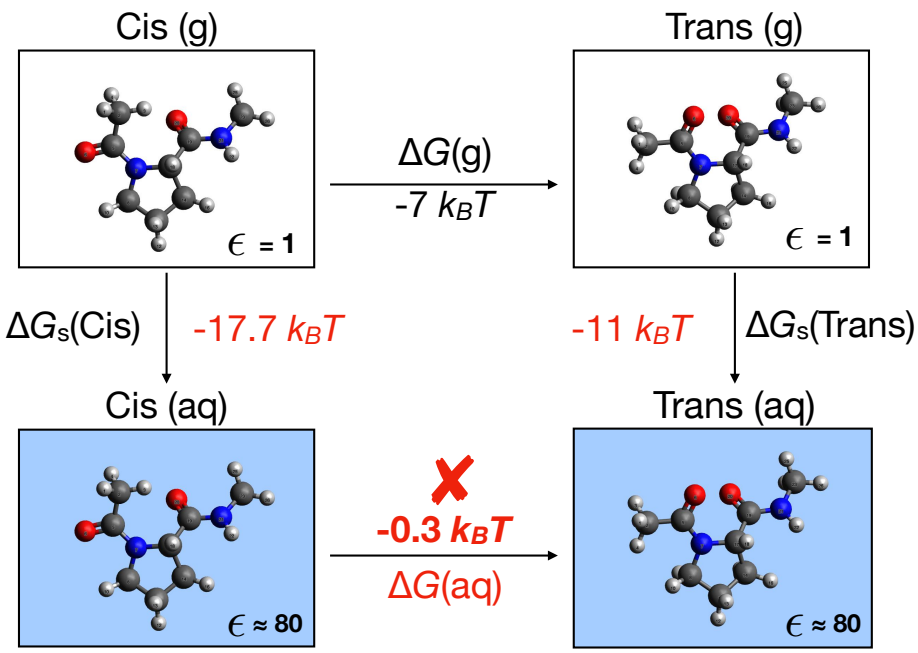


Figure 5: Thermodynamic cycle for evaluation of ΔG(aq.), the free energy change from cis(aq.) to trans(aq.) for proline dipeptide (HH). $T = 298$ K. An implicit water model combined with a harmonic approximation underestimated the free energy change ($-0.3\ k_BT$) when applied to geometry relaxed structures. The experimental value is $\approx -1.6\ k_BT$.

We conclude that this model has two major limitations. Firstly, the implicit water model does not capture the subtle differences in solvation of cis and trans peptides of HH. To verify this, we performed explicit water solvation calculations (SI, Figure S1) using multi-state

hamiltonian replica exchange simulations and evaluated the free energy of solvation. The ΔG_{s}(cis) and ΔG_{s}(trans) were $-25.8\ k_BT$ and $-23.2\ k_BT$ with Amber-ff14sb force-field and explicit waters (SI, Table S1) compared to $-17.7\ k_BT$ and $-11\ k_BT$ with implicit water calculations (Figure 5). Part of this discrepancy emerges from the differential arrangement of water within the inner solvation shell of the peptide, a feature that is absent in implicit solvation models. Secondly, the geometry optimizations led to the peptide backbone (ϕ,ψ) adopting a local minimum structure. Furthermore, a harmonic approximation only treats small excursions about that energy minima. Thus, the variety of orientations available for the proline backbone as seen in the Ramachandran map[42,43] and CMAP[44,45] were not reflected here. These shortcomings were addressed naturally in our MD calculations (section 3.3) that include explicit waters and exhaustive sampling (SI, Figure S3) of the peptide backbone .

3.3 Corrected force-field captures experimental conformational equilibria

Ring conformational equilibrium:

The populations of endo and exo pucker states obtained from MD simulations are shown in Table 1. As expected, the ring pucker corrected force-field ($\mathrm{FF_C}$) performed much better than the uncorrected one (FF). The correction was particularly significant for obtaining the correct behavior for RF compared to experiment.

Table 1: (Endo% : Exo%) Pucker populations in trans dipeptides

	HH	RF	SF
Amber-ff14sb	60 : 40	-	-
FF [a]	-	53 : 47	85 : 15
$\mathrm{FF_C}$ [b]	-	2 : 98	95 : 5
Experiment[39]	66 : 34	14 : 86	95 : 5

[a] Force-field without pucker correction; [b] with correction;

In light of the successful treatment of ring conformational equilibrium, we next take up the case of cis-trans isomerism. This is a significant check for our force-field model,

as the correction was applied only along ring pucker coordinates. This will also test the capability of our force-field to capture the coupling between ring puckering effects (induced by fluorination) and the cis-trans conformational equilibrium.

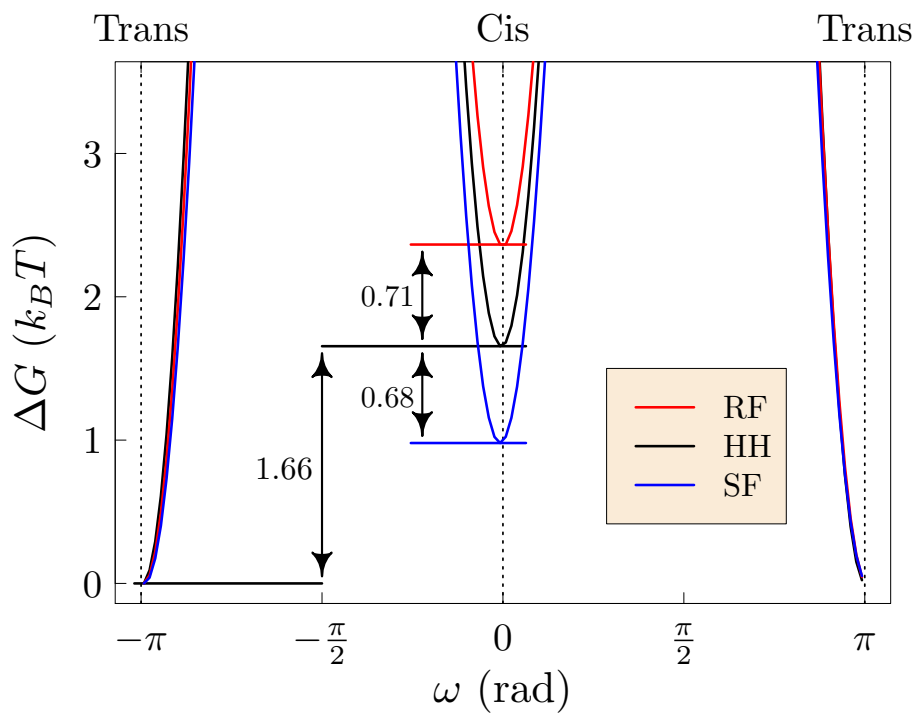


Figure 6: Free energy change with ω, the torsion angle along the cis→trans coordinate. Here, the simulations treat a single dipeptide (RF/HH/SF) in bulk water at $T = 298$ K.

Cis-trans conformational equilibrium:

A quantitative analysis of the cis-trans conformational equilibrium for HH, RF and SF dipeptide cases were obtained from potential of mean force calculations along the ω torsion angle (Figure 6). The free energy of the trans isomer was shifted to zero and those of the cis isomer are compared. For the case of proline dipeptide (HH), the free energy change was 1.66 $k_B T$. In terms of relative populations, this amounted to 84% and 16% for trans and cis respectively. As a side note, we add that the electronic $n \rightarrow \pi^*$ interactions are not explicitly modeled in classical force-fields like Amber[28] but captured through compensating electro-

statics, van der waals and intramolecular torsion potentials.[46] For the fluorinated proline dipeptides (RF and SF), the free energy changes differed by $\approx 0.7\ k_BT$ and sandwiched the value for HH. For the R stereo-isomer (RF), the trans state was stabilized, whereas the S stereo-isomer (SF) stabilized the cis isomer.

Table 2: Relative populations (Trans% : Cis%) estimated from figure 6

	RF[a]	HH[b]	SF[a]
Experiment[41]	87 : 13	82 : 18	71 : 29
MD	91 : 9	84 : 16	73 : 27

[a] Corrected force-field (FF_C); [b] Amber-ff14sb.

These results are in good quantitative agreement with experimental populations of trans and cis states (Table 2) available for a dipeptide model, however, with a slightly different blocking group. The main difference between our model (Ace-Pro-NHMe) and the experimental (Ace-Pro-OMe) system is the presence of a NH instead of O in the terminal cap. The NH group is capable of forming hydrogen bonds with carbonyl oxygens and water, accounting for part of the small deviations from experiment.

4 Discussion

Rapid progress in experimental capability and synthesis have led to the testing of a variety of novel substituted prolines such as trifluoromethylated prolines and poly-fluorinated prolines.[41] These have further extended the tunability of cis and trans populations. Such methods of imposing stereo-electronic effects by pre-organizing amino acids have been shown to impart proteins with extraordinary conformational stability without perturbing its native structure. Replacing multiple residues in collagen strands with 4R-fluoroprolines have been shown[47] to hyperstabilize triple helices with an increase in thermal stability upto $\approx 90^{\circ}C$. Such improvements can be beneficial for cold chain storage and tackling transport problems associated with biologics.

The inclusion of puckering effects are essential for computations involving such novel

ring substituted proline residues. Specifically, introduction of a fluorine leads to puckering effects (for RF and SF) that are not inherently captured in classical force-fields. This is due to a shift in puckering trend resulting from a solvent induced electronic structure response (Section 3.2, Fig. 4). A literature survey revealed that this effect is unknown in the context of prolines, however, a similar effect was recently reported in the context of furanose rings[18] in carbohydrates. That effect was attributed to a flux of atomic charges in the ring induced by solvent.

Since derivation of atomic partial charges for classical force-fields are typically done in gas phase, the electronic response of the ring due to solvent is not captured correctly. In this work, we outline a theoretical construct (Section 2.2.1, Eq. 4-8) for using implicit water calculations to obtain an intramolecular ring puckering correction. Adding this correction fixes a fundamental shortcoming in the treatment of ring puckering in classical force-fields, and in principle, can be extended to other substituents (besides fluorine) or rings (furanoses). This would allow the continued use of vastly developed biomolecular force-fields for studying puckering related phenomena.

We demonstrate the utility of our corrected force-field in MD calculations discussed above. For simple proline dipeptide (HH), the trans(aq.) state was relatively more stable than cis(aq.) state by 1.66 k_BT. Using the corrected force-field for fluorinated prolines, we find that switching the stereo chemistry of the fluorine at the C_γ site reversed the puckering preference and ultimately perturbed the cis-trans isomerism (by $\approx 0.7\ k_BT$). This is because the exo pucker results in a favorable conformation between adjacent carbonyl groups (in the trans state) leading to a stronger interaction.[11] Our force-field model quantitatively predicts these coupling effects and can be used to model and guide new substitutions in similar systems.

5 Conclusions

We compared the quantum mechanical potential energy surface along ring pucker coordinates for proline and 4-fluorinated proline dipeptides. Electronic structure calculations showed a significant shift in puckering trends from gas phase to implicit solvent phase. This is due to the ring's electronic response to solvent, an effect previously unknown for prolines but recently observed in the context of furanoses in carbohydrates. The implicit water calculations qualitatively agreed with experimental trends and were used for correcting classical force-fields. We demonstrated the utility of our corrected force-field for capturing subtle shifts in the cis-trans isomerism due to coupling with ring puckering behavior. The evaluated free energies were in excellent agreement with experiment. This model can be extended to other rings and substituents to guide new chemistries in similar systems.

Acknowledgement

The authors thank Professor Qiang Cui for helpful discussions. This research was performed using the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. This research was also supported in part by UW-Madison Department of Chemistry PHOENIX research cluster through National Science Foundation Grant CHE-0840494. AM is a Hirschfelder fellow.

Supporting Information Available

6 Comparison between implicit and explicit hydration

[a] Amber-ff14sb; [b] MP2/6-31G(d,p).

Conformer	TI[a]	BAR[a]	MBAR[a]	PCM[b]
Cis	-25.9	-25.7	-25.75	-17.7
Trans	-23.3	-23.2	-23.2	-11

Table S1: Free energy of solvation of proline dipeptide (k_BT units) evaluated using multi-state simulations with hamiltonian replica exchange (Figure S1). The results for Thermodynamic integration (TI), Bennet acceptance ratio (BAR) and Multi-state Bennet acceptance Ratio (MBAR) were calculated using the alchemical analysis package.[48] The polarizable continuum model[23] (PCM) was used to perform implicit water calculations at the MP2/6-31G(d,p) level.

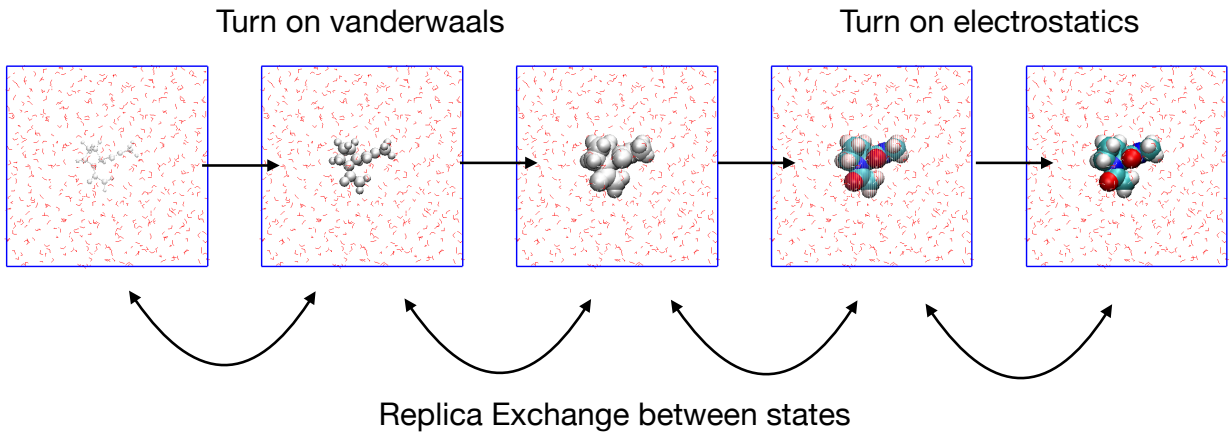


Figure S1: Multistate simulations with hamiltonian replica exchange for the evaluation of hydration free energy of proline dipeptide (HH). The solute-solvent interactions were gradually added through 11 intermediate states (only 5 shown here). The Van der Waals and electrostatics interactions were individually turned on in steps of $\Delta\lambda = 0.2$ ($\lambda = 0$ represents no interaction and $\lambda = 1$ represents full interaction). Replica exchange attempts were made every 1ps with an observed success rate of $\approx 10\%$. Equilibration and production molecular dynamics runs were 1ns and 5ns respectively using Gromacs simulation package. A single peptide was solvated in 500 waters and are described using Amber-ff14sb and TIP3P force-field models.

7 PES scan with force-field

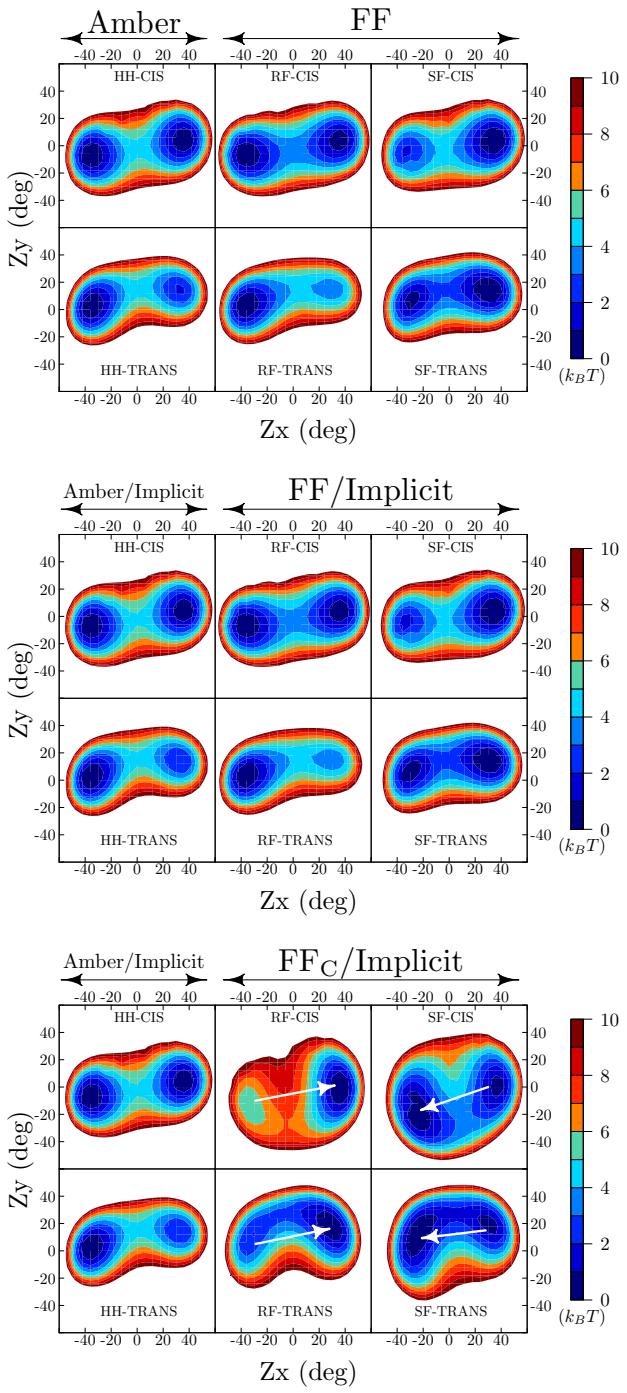


Figure S2: Energy scans for cis and trans dipeptides of proline (HH), 4R-fluoroproline (RF) and 4S-fluoroproline (SF) along collective variables Zx and Zy . The corrected Amber force-field (FF_C) recovers the correct ring puckering trend when compared to experiment.

8 Backbone (ϕ, ψ) sampling in MD simulations

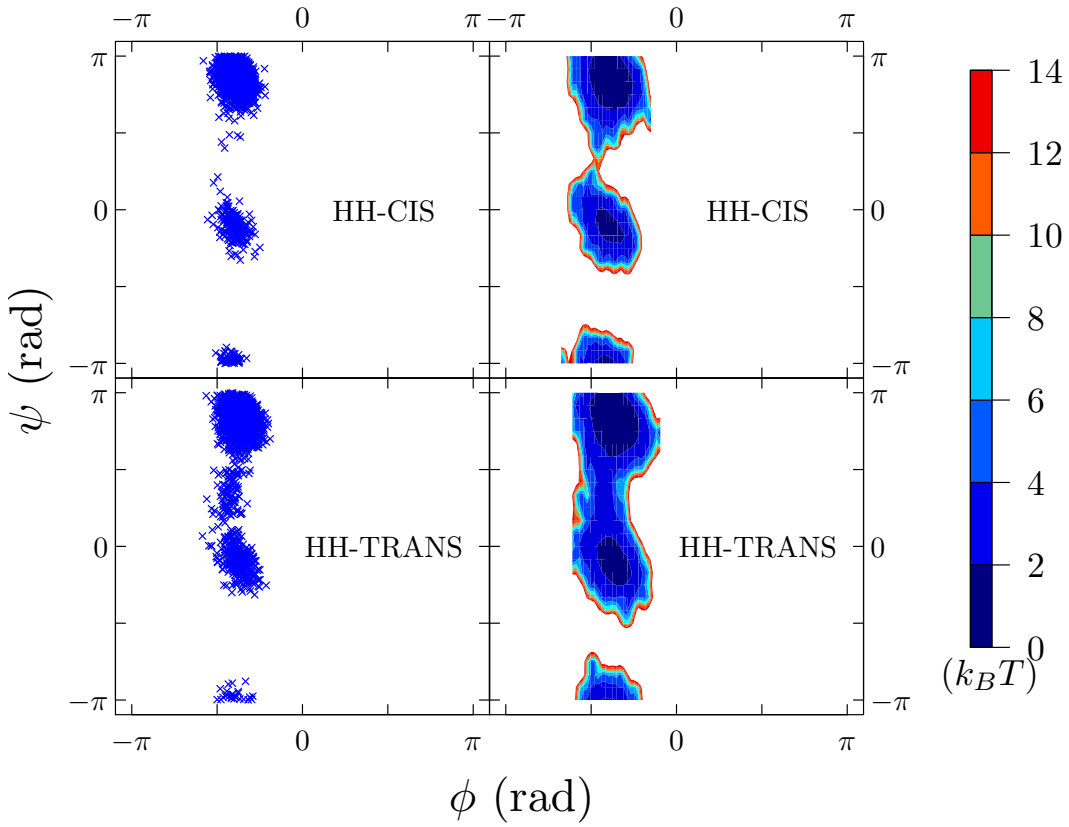
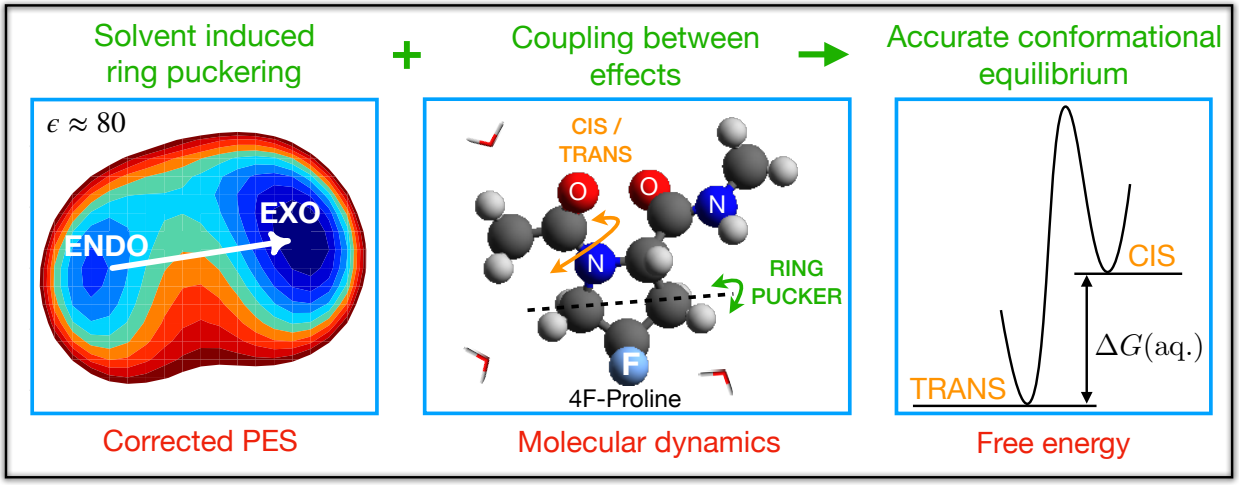


Figure S3: Ramachandran (ϕ,ψ) plot and the corresponding free energies, $-k_B T \ \ln p(\phi,\psi)$, for proline dipeptide (HH) from MD simulations at $T = 298\ K$. The trajectory was obtained from umbrella sampling runs with windows centered at $\omega = -\pi$ (trans) and $\omega = 0$ (cis).

9 Table Of Contents - graphic



References

(1) Link, A. J.; Mock, M. L.; Tirrell, D. A. Non-canonical amino acids in protein engineering. Current opinion in biotechnology **2003**, 14, 603–609.

(2) Renfrew, P. D.; Choi, E. J.; Bonneau, R.; Kuhlman, B. Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design. PLoS One **2012**, 7.

(3) Das, R.; Baker, D. Macromolecular modeling with rosetta. Annu. Rev. Biochem. **2008**, 77, 363–382.

(4) Berger, A. A.; Völler, J.-S.; Budisa, N.; Koksch, B. Deciphering the Fluorine Code-The Many Hats Fluorine Wears in a Protein Environment. Accounts of chemical research **2017**, 50, 2093–2103.

(5) Buer, B. C.; Marsh, E. N. G. Fluorine: a new element in protein design. Protein Science **2012**, 21, 453–462.

(6) Arnold, U.; Raines, R. T. Replacing a single atom accelerates the folding of a protein and increases its thermostability. Organic & biomolecular chemistry **2016**, 14, 6780–6785.

(7) Holzberger, B.; Obeid, S.; Welte, W.; Diederichs, K.; Marx, A. Structural insights into the potential of 4-fluoroproline to modulate biophysical properties of proteins. Chemical Science **2012**, 3, 2924–2931.

(8) Xia, J.; Levy, R. M. Molecular dynamics of the proline switch and its role in Crk signaling. The Journal of Physical Chemistry B **2014**, 118, 4535–4545.

(9) Bai, H.; Xue, R.; Bao, H.; Zhang, L.; Yethiraj, A.; Cui, Q.; Chapman, E. R. Different states of synaptotagmin regulate evoked versus spontaneous release. Nature communications **2016**, 7, 1–9.

(10) Pastorino, L.; Sun, A.; Lu, P.-J.; Zhou, X. Z.; Balastik, M.; Finn, G.; Wulf, G.; Lim, J.; Li, S.-H.; Li, X., et al. The prolyl isomerase Pin1 regulates amyloid precursor protein processing and amyloid-β production. Nature **2006**, 440, 528–534.

(11) Newberry, R. W.; Raines, R. T. The $n \rightarrow \pi^*$ Interaction. Accounts of chemical research **2017**, 50, 1838–1846.

(12) Bartlett, G. J.; Choudhary, A.; Raines, R. T.; Woolfson, D. N. $n \rightarrow \pi^*$ interactions in proteins. Nature chemical biology **2010**, 6, 615.

(13) Newberry, R. W.; Raines, R. T. Peptidomimetics I; Springer, 2016; pp 1–25.

(14) Improta, R.; Benzi, C.; Barone, V. Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. Journal of the American Chemical Society **2001**, 123, 12568–12577.

(15) Fischer, S.; Dunbrack Jr, R. L.; Karplus, M. Cis-trans imide isomerization of the proline dipeptide. Journal of the American Chemical Society **1994**, 116, 11931–11937.

(16) Kang, Y. K. Ab initio and DFT conformational study of proline dipeptide. Journal of Molecular Structure: THEOCHEM **2004**, 675, 37–45.

(17) Ramachandran, G.; Bansal, M.; Bhatnagar, R. A hypothesis on the role of hydroxyproline in stabilizing collagen structure. Biochimica et Biophysica Acta (BBA)-Protein Structure **1973**, 322, 166–171.

(18) Gaweda, K.; Plazinski, W. The systematic influence of solvent on the conformational features of furanosides. Organic & biomolecular chemistry **2019**, 17, 2479–2485.

(19) Awada, H.; Grison, C. M.; Charnay-Pouget, F.; Baltaze, J.-P.; Brisset, F.; Guillot, R.; Robin, S.; Hachem, A.; Jaber, N.; Naoufal, D., et al. Conformational Effects through

Hydrogen Bonding in a Constrained γ-Peptide Template: From Intraresidue Seven-Membered Rings to a Gel-Forming Sheet Structure. The Journal of organic chemistry **2017**, 82, 4819–4828.

(20) Huang, M.; Giese, T. J.; Lee, T.-S.; York, D. M. Improvement of DNA and RNA sugar pucker profiles from semiempirical quantum methods. Journal of chemical theory and computation **2014**, 10, 1538–1545.

(21) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., et al. Gaussian 16. 2016.

(22) Sato, T. Another method for specifying furanose ring puckering. Nucleic acids research **1983**, 11, 4933–4938.

(23) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. Chem. Rev. **2005**, 105, 2999–3093.

(24) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of chemical theory and computation **2015**, 11, 3696–3713.

(25) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. Journal of computational chemistry **2004**, 25, 1157–1174.

(26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. Journal of the American Chemical Society **1995**, 117, 5179–5197.

(27) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule

and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. Journal of Computational Chemistry **1995**, 16, 1357–1377.

(28) Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M., et al. AMBER 18; 2018. University of California, San Francisco

(29) Robalo, J. R.; Huhmann, S.; Koksch, B.; Verde, A. V. The multiple origins of the hydrophobicity of fluorinated apolar amino acids. Chem **2017**, 3, 881–897.

(30) Robalo, J. R.; Verde, A. V. Unexpected trends in the hydrophobicity of fluorinated amino acids reflect competing changes in polarity and conformation. Physical Chemistry Chemical Physics **2019**, 21, 2029–2038.

(31) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the generalized Born model suitable for macromolecules. The Journal of Physical Chemistry B **2000**, 104, 3712–3720.

(32) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A., et al. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. Computer Physics Communications **2009**, 180, 1961–1972.

(33) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. Computer Physics Communications **2014**, 185, 604–613.

(34) Bonomi, M.; Bussi, G.; Camilloni, C.; Tribello, G. A.; Banáš, P.; Barducci, A.; Bernetti, M.; Bolhuis, P. G.; Bottaro, S.; Branduardi, D., et al. Promoting transparency and reproducibility in enhanced molecular simulations. Nature methods **2019**, 16, 670–673.

(35) Nosé, S. A Molecular Dynamics Method for Simulations in the Canonical Ensemble. Mol. Phys. **1984**, 52, 255–268.

(36) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied physics **1981**, 52, 7182–7190.

(37) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: fast, flexible, and free. Journal of computational chemistry **2005**, 26, 1701–1718.

(38) Grossfield, A. An implementation of WHAM: The Weighted Histogram Analysis Method, Version 2.0.9. Available at membrane.urmc.rochester.edu/content/wham **2014**,

(39) DeRider, M. L.; Wilkens, S. J.; Waddell, M. J.; Bretscher, L. E.; Weinhold, F.; Raines, R. T.; Markley, J. L. Collagen stability: insights from NMR spectroscopic and hybrid density functional computational investigations of the effect of electronegative substituents on prolyl ring conformations. Journal of the American Chemical Society **2002**, 124, 2497–2505.

(40) Eberhardt, E. S.; Panasik, N.; Raines, R. T. Inductive effects on the energetics of prolyl peptide bond isomerization: implications for collagen folding and stability. Journal of the American Chemical Society **1996**, 118, 12261–12266.

(41) Verhoork, S. J.; Killoran, P. M.; Coxon, C. R. Fluorinated prolines as conformational tools and reporters for peptide and protein chemistry. Biochemistry **2018**, 57, 6132–6143.

(42) Lovell, S.; Davis, I.; Arendall 3rd, W.; Word, P. de, Bakker JM, Prisant MG, Richardson JS, et al. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins **2003**, 50, 437–450.

(43) Ting, D.; Wang, G.; Shapovalov, M.; Mitra, R.; Jordan, M. I.; Dunbrack Jr, R. L. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. PLoS computational biology **2010**, 6.

(44) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. Journal of chemical theory and computation **2013**, 9, 4046–4063.

(45) Kar, P.; Gopal, S. M.; Cheng, Y.-M.; Predeus, A.; Feig, M. PRIMO: a transferable coarse-grained force field for proteins. Journal of chemical theory and computation **2013**, 9, 3769–3788.

(46) Doshi, U.; Hamelberg, D. Reoptimization of the AMBER force field parameters for peptide bond (Omega) torsions using accelerated molecular dynamics. The Journal of Physical Chemistry B **2009**, 113, 16590–16595.

(47) Holmgren, S. K.; Bretscher, L. E.; Taylor, K. M.; Raines, R. T. A hyperstable collagen mimic. Chemistry & biology **1999**, 6, 63–70.

(48) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. J Comput Aided Mol Des **2015**, 29, 397–411.