Supplementary Information

Mugnai & Elber, "Thermodynamic Cycle Without Turning Off Self Interaction… "

Some insights on alchemical substitution of angles and torsions

ANGLES

Bond-angle potential restrains the angle θ between three particles (say, i, j and k) near an equilibrium value. In molecular mechanics force field like OPLS-AAL[1] such a potential is defined as a spring in angular space:

$$U_{ang}(\theta) = k_{ang}\left(\theta - \theta_{eq}\right)^2 \tag{1}$$

Here, k_{ang} is the spring constant and θ_{eq} is the equilibrium angle. Given the Cartesian coordinates of three particles, the angle is computed in MD simulations as:

$$\theta_{MD} = \arccos\left[\frac{\vec{r}_{ij}\cdot\vec{r}_{jk}}{r_{ij}r_{jk}}\right] = \arccos\left[\cos(\theta)\right] \tag{2}$$

Here, the vectors $\vec{r}_{ij}$ and $\vec{r}_{jk}$ connect particle j to the other two particles. The bond-angle potential used in MD is then slightly different from Eq. 1:

$$U_{ang}(\theta_{MD}) = k_{ang}\left(\theta_{MD} - \theta_{eq}\right)^2 = k_{ang}\left\{\arccos\left[\cos(\theta)\right] - \theta_{eq}\right\}^2 \tag{3}$$

The inverse function arccosine is defined in the set $[0,\pi]$. Therefore, only in this set are the potentials in Eq. 1 and in Eq. 3 the same. Figure 1 shows the bond-angle potential (Eq. 1) in blue and the one that we use in MD (Eq. 3) in red.

FIGURE 1

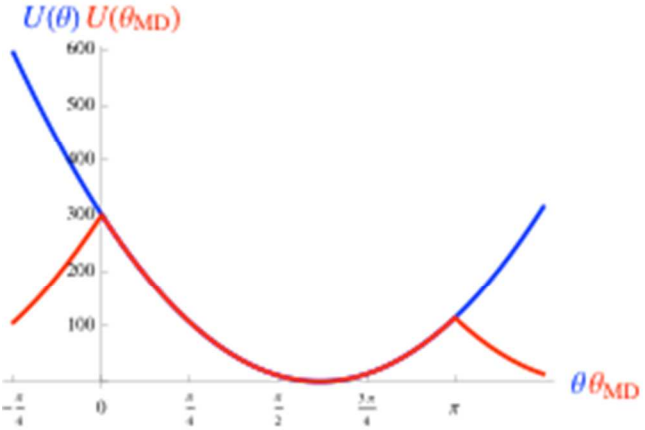

Figure 1: The bond-angle potential (Eq.1) is reported in blue, while the potential obtained in MD (Eq.3) is reported in red. The equilibrium value (θ_{eq}=1.941) and the spring constant

(k_{ang}=80kcal/mol) are those for the N-C_α-C backbone angle in OPLS-AAL force field[1]. The figure was made with Mathematica[2].

It is clear that if the angle θ exits the range $[0,\pi]$ in a MD simulation, we will encounter problems. The derivative of the potential in Eq. 1 is:

$$\frac{dU_{ang}(\theta)}{d\theta} = 2k_{ang}\left(\theta - \theta_{eq}\right) \tag{4}$$

In MD, according with Eq. 2, we actually compute:

$$\frac{dU_{ang}(\theta_{MD})}{d\theta_{MD}} = 2k_{ang}\left(\theta_{MD} - \theta_{eq}\right) = 2k_{ang}\left\{\arccos\left[\cos(\theta)\right] - \theta_{eq}\right\} \tag{5}.$$

Figure 2 reports the derivative of the potential in Eq.4 in blue, and the derivative of the potential in Eq. 5 in red.

FIGURE 2

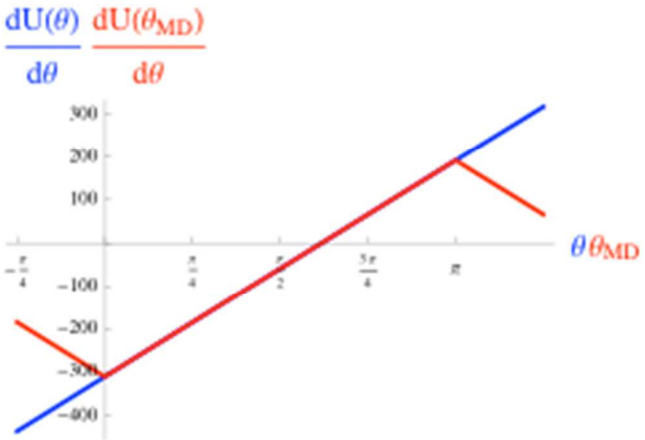

Figure 2: The derivative of the bond-angle potential (Eq.4) is reported in blue, while the derivative of the potential obtained in MD (Eq.5) is reported in red. The equilibrium value (θ_{eq}=1.941) and the spring constant (k_{ang}=80kcal/mol) are those for the N-C_α-C backbone angle in OPLS-AAL force field[1]. The figure was made with Mathematica[2].

The function in Eq. 5 is continuous in $\theta=0$ and in $\theta=\pi$, but it has the wrong sign in $\theta<0$ and $\theta>\pi$. While, indeed, the potential in Eq. 3 is increasing (decreasing) for $\theta<0$ ($\theta>\pi$) (see Fig. 1), the derivative that we compute is negative (positive) for $\theta<0$ ($\theta>\pi$) (see Fig. 2). This will make the sign of the force opposite to the correct one outside the range $[0,\pi]$. Our energy (and so free energy) calculation is then going to be inconsistent with the forces governing the evolution (and so the sampling) of configurations.

In a regular MD simulation at 300K, the spring constant is stiff enough to restrain the fluctuations of the angle within the range of a few degrees. The problem just highlighted is never encountered, since the energy barrier to reach the configurations $\theta<0$ and $\theta>\pi$ is too high to overcome.

In an alchemical substitution, the potential in Eq. 1 (or Eq. 3) is multiplied by the switching parameter λ. The effective spring constant is then weakened, and the lower the value of λ the larger is the angular space that can be sampled. Eventually, the energy barrier to reach the configurations $\theta<0$ and $\theta>\pi$ will be so low that such

configurations may actually be sampled. Therefore, we need to find a way to restrain the angle between three particles such that if annihilated by alchemical methods is well behaved at $\theta<0$ and $\theta>\pi$.

The solution that we propose is to use Urey-Bradley bonds to restrain the angle around the equilibrium value for all those angles that are going to be annihilated along the alchemical pathway (e.g. the angles that connect the protein to the mutated part). Given the three particles *i*, *j* and *k*, the Urey-Bradley bond restrains the fluctuations of θ around the equilibrium angle by adding a spring in Cartesian space between particles *i* and *k*:

$$U_{UB}\left(r_{ik}\right)= k_{UB}\left(r_{ik}-r_{ik,eq}\right)^2 \tag{6}$$

Here, of course

$$r_{ik}=\sqrt{r_{jk}^2+r_{ij}^2-2r_{jk}r_{ij}\cos\left(\theta\right)} \tag{7}$$

The two parameters $r_{ik,eq}$ and θ_{eq} are chosen such that the minimum of the potential corresponds to the equilibrium position according to bond-angle potential and the small fluctuations around the minimum are the same. Therefore, we obtain:

$$r_{ik,eq}=\sqrt{r_{jk,eq}^2+r_{ij,eq}^2-2r_{jk,eq}r_{ij,eq}\cos\left(\theta_{eq}\right)} \tag{8}$$

$$k_{UB}=\frac{1+2\dfrac{r_{ij,eq}^2 r_{jk,eq}^2}{r_{ij,eq}^2+r_{jk,eq}^2}\cos\left(\theta_{eq}\right)}{\dfrac{r_{ij,eq}^2 r_{jk,eq}^2}{r_{ij,eq}^2+r_{jk,eq}^2}\sin^2\left(\theta_{eq}\right)}k_{ang} \tag{9}$$

The choice of the alchemical pathway to annihilate/create the Urey-Bradley bonds is arbitrary. Therefore, we select a scheme (and a particular λ-scaling) such that the free energy difference is computed numerically with higher accuracy. For this purpose, we consider a "toy" problem in which we can evaluate the free energy analytically. We compute the free energy difference due to the annihilation of a Urey-Bradley bond for a triatomic molecule. The alchemical potential of such simple system is given by two chemical bonds and a Urey-Bradley bond scaled by the switching parameter. The parameter α is the target for optimization:

$$U\left(r,r',r_{UB};\lambda\right)= k\left(r-r_{eq}\right)^2+k'\left(r'-r'_{eq}\right)^2+\lambda^{\alpha}k_{UB}\left(r_{UB}-r_{UB,eq}\right)^2 \tag{10}$$

The configurational partition function for this potential is[3]:

$$Z\left(N,V,T;\lambda\right)= 8\pi^2 V\int r^2 r'^2\sin\left(\theta\right)\exp\left[-\beta U\left(r,r',r_{UB};\lambda\right)\right]dr\,dr'\,d\theta \tag{11}$$

Here, we changed the coordinates to a polar system and integrated the external degrees of freedom. Let us make the assumption that the spring constant for the chemical bond is so stiff that in the Jacobian in Eq. 11 and in the Urey-Bradley bond (see Eq. 7) the bond lengths *r* and *r'* can be substituted by their equilibrium values. This yields:

$$Z\left(N,V,T;\lambda\right)= 8\pi^2 V r_{eq}^2 r'^2_{eq}\sqrt{\frac{2\pi}{k\beta}}\sqrt{\frac{2\pi}{k'\beta}}\int\sin\left(\theta\right)\exp\left[-\beta\lambda^{\alpha}\left(r_{UB}-r_{UB,eq}\right)^2\right]d\theta \tag{12}$$

Let us further assume that $k=k'$, $r_{eq}=r'_{eq}$ and that $\theta_{eq}=0$. In this case the Urey-Bradley potential is (see Eq. 6-8):

$$U_{UB}(\theta) = 2k_{UB}r_{eq}^2\left[1-\cos(\theta)\right] \quad (13).$$

The free energy difference computed by TI is:

$$\frac{\partial F}{\partial \lambda} = \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda} = 2\alpha\lambda^{\alpha-1}k_{UB}r_{eq}^2 \frac{\int \sin(\theta)\left[1-\cos(\theta)\right]\exp\left\{-2\beta\lambda^{\alpha}k_{UB}r_{eq}^2\left[1-\cos(\theta)\right]\right\}d\theta}{\int \sin(\theta)\exp\left\{-2\beta\lambda^{\alpha}k_{UB}r_{eq}^2\left[1-\cos(\theta)\right]\right\}d\theta} \quad (14)$$

The integral can be solved numerically as a function of λ using Mathematica[2]. With $\alpha=1$ we get the blue line in Fig. 3, with $\alpha=2$ the red line.

FIGURE 3

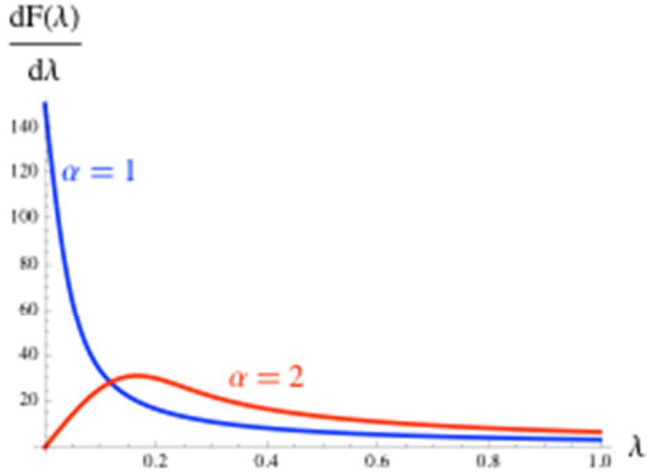

Figure 3: The derivative of the alchemical free energy with respect to λ (Eq. 14) is reported. In blue we show α=1, and in red α=2. The rest of the parameters are: r_{eq}=1Å, k_{UB}=75 kcal/mol/ Å² and β=0.6 kcal/mol. The figure was made with Mathematica[2].

We noticed (unpublished results) that the larger relative errors in the evaluation of $dF/d\lambda$ are found when λ approaches *0*. A possible rationale is that the closer the system is to the decoupling point, the larger is the configurational space to sample. To reduce the impact of these terms on the overall free energy difference, we would like $dF/d\lambda$ to be small at $\lambda \cong 0$. Therefore, we adopted $\alpha=2$ in our simulations.

TORSIONS

Given four particles (say, *i*, *j*, *k* and *l*), the torsion potential restrains the dihedral angle φ between the plane identified by the particles *i*, *j* and *k* and the plane identified by the particles *j*, *k* and *l* (see Fig. 4). The functional form used in molecular mechanics force fields is periodic in the angle and does not suffer the same problems as the angular potential. On the other hand, if the weakening of the angular interaction allows particle *l* to collapse over particle *j*, the plane identified by the particles *j*, *k* and *l* is not defined, and so the dihedral angle itself is not defined. To avoid this issue it is enough to remove torsions and improper torsions before the angles.

FIGURE 4

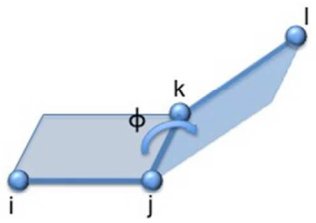

Figure 4: An example of torsion dihedral φ is reported.

Free energy contribution of substituting angular potential with Urey-Bradley potential.

In the thermodynamic cycle that we compute, we remove some angular interactions from ILE sidechain analog (between *"P"* and *"N"* particles) and create some angular interactions in GLN sidechain analog (between *"P"* and *"M"* particles) in water. Then we do the opposite in vacuum. The list of the angles that are substituted (i.e. removed or created) is reported in the following Table.

TABLE 1

Created/annihilated angles in ILE side chain analog				Created/annihilated angles in GLN side chain analog			
P		N		P		M	
HS1	CS	CI1		HS1	CS	CQ1	
HS2	CS	CI1		HS2	CS	CQ1	
HS3	CS	CI1		HS3	CS	CQ1	
CS		CI1	HI11	CS		CQ1	HQ11
CS		CI1	HI12	CS		CQ1	HQ12
CS		CI1	CI2	CS		CQ1	CQ2

Table 1: This table reports the list of angles that are removed/created in the mutations in Fig. 2 of the main text (horizontal arrows). These are all the angles that include *"P"* and *"N"* particles or *"P"* and *"M"* particles. The names of the atoms can be found in a molecular sketch in Fig. 3 of the main text.

As already stated, it is convenient to remove/add Urey-Bradley bonds instead of regular angular interactions. Clearly, the free energy difference of removing/creating Urey-Bradley bonds is different from the free energy difference of removing/creating bond angle interactions since the functions are different. According to our protocol as discussed in Table 1 and Table 2, the mutation simulations (horizontal arrows in Fig. 2 of main text) use Urey-Bradley bonds. In the simulations of aqueous solutions (vertical arrows in Fig. 2 of main text) the sampling of configurations was performed according to the regular bond-angle potential. Therefore, we need to correct the free energy for the use of different functional forms. Potential switches occur at the four corners of Fig. 2 of the main text. In the top left corner we change from the regular angular potential to the Urey-Bradley potential for all the substituted angles in ILE sidechain analog (Table 1). This free energy difference is defined $\Delta F_{I,solv}^{A\to UB}$. We then reach the top right corner with a Urey-Bradley potential for the substituted angles in GLN sidechain analog (Table 1), so we need to compute the free energy difference of removing them and

creating the regular angular potential. We refer to this term as $-\Delta F_{Q,solv}^{A\to UB}$. In the bottom right corner of Fig. 2 in the main text we have regular bond-angle potential, but the mutation is performed according to Urey-Bradley terms. Therefore we need to compute the free energy difference of substituting the regular bond angle term to Urey-Bradley terms for all the substituted angles in GLN sidechain analog in vacuum (Table 1). We refer to this term as $\Delta F_{Q}^{A\to UB}$. Finally, in the bottom left corner of Fig. 2 of the main text all the substituted angles in ILE sidechain analog (Table 1) are described according to a Urey-Bradley potential, but in the solvated part we use the regular bond-angle term. The free energy difference of performing this substitution in vacuum is $-\Delta F_{I}^{A\to UB}$. Overall, the total correction to the cycle due to our inconsistent use of different angle potentials is:

$$\oint dF^{A\to UB} = \Delta F_{I,solv}^{A\to UB} - \Delta F_{I}^{A\to UB} - \Delta F_{Q,solv}^{A\to UB} + \Delta F_{Q}^{A\to UB} \tag{15}$$

Each free energy difference in equation (18) was computed using the Bennett Acceptance Ratio (BAR) method[4]. According to this method, the free energy difference associated with our change in the force field is computed using the following formula:

$$\Delta F^{A\to UB} = -k_B T \ln \frac{\left\langle \frac{1}{1+\exp\left[\beta\left(U_{UB}-U_A-C\right)\right]} \right\rangle_A}{\left\langle \frac{1}{1+\exp\left[-\beta\left(U_{UB}-U_A-C\right)\right]} \right\rangle_{UB}} + C \tag{16}$$

Here, U_A represents the angular potential for all those angles that are alchemically removed (see Table 1), U_{UB} is instead the potential for the Urey-Bradley bond that substitutes the regular angular potential. The symbol $\langle \ldots \rangle_A$ refers to an average performed over an ensemble of structures sampled when the angular potential is used. The symbol $\langle \ldots \rangle_{UB}$ refers to an average performed over an ensemble of structures sampled when the Urey-Bradley bonds are used. The value of C is determined according to the following formula:

$$C = \Delta F^{A\to UB} + k_B T \ln \frac{n_{UB}}{n_A} \tag{17}$$

where n_{UB} is the number of structures in the sample performed with the Urey-Bradley bonds, and n_A is the number of structures in the sample performed with the regular angular potential. The two equations 16 and 17 can be used iteratively to obtain the value of the free energy difference.

The computation of the variance of the free energy associated with BAR method was performed according to the following formula[4,5]:

$$\sigma^2_{BAR} = \frac{1}{n_A\beta^2}\left[\frac{\left\langle\left(\frac{1}{1+\exp\left[\beta\left(U_{UB}-U_A-C\right)\right]}\right)^2\right\rangle_A}{\left\langle\frac{1}{1+\exp\left[\beta\left(U_{UB}-U_A-C\right)\right]}\right\rangle_A^2}\right] + \frac{1}{n_{UB}\beta^2}\left[\frac{\left\langle\left(\frac{1}{1+\exp\left[-\beta\left(U_{UB}-U_A-C\right)\right]}\right)^2\right\rangle_{UB}}{\left\langle\frac{1}{1+\exp\left[-\beta\left(U_{UB}-U_A-C\right)\right]}\right\rangle_{UB}^2}\right] \quad (18)$$

The simulations are carried out using the same systems as those of the alchemical substitutions (see text for further details). All the simulations were 2ns long. The results are in the following table:

Table 2

$\Delta F_{I,solv}^{A->UB}$	(0.1421±0.0088)kcal/mol
$\Delta F_{I}^{A->UB}$	(0.2007±0.0047)kcal/mol
$\Delta F_{Q,solv}^{A->UB}$	(0.1505±0.0010)kcal/mol
$\Delta F_{Q}^{A->UAB}$	(0.2038±0.0048)kcal/Mol

Table 2: The results of the free energy difference upon substitution of the regular angular interactions with the Urey-Bradley interactions are reported with their errors.

Hence the free energy of changing the potential of the angles between the fragment and the molecule from the regular bond angle potential to the Urey-Bradley potential is small for each individual term. In the context of a comparison with experiment it is one order of magnitude smaller than the expected systematic errors (~1-2kcal/mol in the case of ligand binding[6]). In the present context of testing numerical accuracy we exploit our knowledge that the entire cycle must be zero. Therefore, we consider only the contribution to the entire cycle. According to Eq. 15 this number is:

$$\oint dF^{A\to UB} = \Delta F_{I,solv}^{A\to UB} - \Delta F_{I}^{A\to UB} - \Delta F_{Q,solv}^{A\to UB} + \Delta F_{Q}^{A\to UB} = \left(-0.005 \pm 0.011\right) kcal/mol \quad (19)$$

It turns out that this correction due to the changes of the angular potential is negligible.

References

(1) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.

(2) *Wolfram Research, Inc., Mathematica, Version 8.0* Champaign, IL, 2010.

(3) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 10.

(4) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245.

(5) Pohorille, A.; Jarzynski, C.; Chipot, C. *J. Phys. Chem. B* **2010**, *114*, 10235.
(6) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struc. Biol.* **2011**, *21*, 150.