IMPORTANCE OF THE ION-PAIR INTERACTIONS IN THE OPEP COARSE-GRAINED FORCE FIELD: PARAMETRIZATION AND VALIDATION.

F. STERPONE, P. NGUYEN, M. KALIMERI, P. DERREUMAUX

1. The OPEP v.4

In this section we provide the analytical formulation of the Hamiltonian of the OPEP force field. The details of how parameters have been optimized are given in[1] and[2]. The OPEP Hamiltonian consists of three main terms:

$$V = V_{local} + V_{non-local} + V_{HB} \tag{1}$$

the first one, V_{local}, casts together all the local interactions that are related to the molecular connectivity, the second one, $V_{non-local}$, includes all the non-bonded interaction terms and the third one,V_{HB}, accounts for the Hydrogen-Bond propensity of backbone atoms.

The local contributions are given by:

$$\begin{aligned} V_{local} &= w_b \sum_{bonds} K_b(r-r_0)^2 + w_a \sum_{angles} K_\alpha(\alpha-\alpha_0)^2 + \\ &+ w_t \sum_{torsions} K_\phi(1+cos(n\phi-\phi_0)) + w_{\phi,\psi}(\sum_{\phi} V_\phi + \sum_{\psi} V_\psi) \end{aligned} \tag{2}$$

where each contribution w_x represents a weighting factor opportunely optimized. The first three terms describe bond, angle and torsion fluctuations. The final terms concern specifically the torsions along the backbone and attempt to confine dihedrals in the Ramachandran plots via Harmonic barriers[1]:

$$\begin{aligned} V_\phi &= K_{\phi\psi}(\phi-\phi_0)^2 \\ V_\psi &= K_{\phi\psi}(\psi-\psi_0)^2 \end{aligned} \tag{3}$$

where $\phi_0 = \phi$ if the torsion is found in the interval $[\phi_{lower}, \phi_{upper}]$ otherwise $\phi_0 = min(\phi - \phi_{lower}, \phi - \phi_{upper})$; the confined region is bound by $\phi_{lower} = -160^\circ$ and $\phi_{upper} = -60^\circ$. For the torsion ψ the confinement is ensured by $\psi_{lower} = -60^\circ$ and $\psi_{upper} = 160^\circ$.

The non-bonded interactions involve backbone and side-chain (Sc) center of forces. The C_αs are hereby considered separately from the rest of backbone atoms N,C',O and H,

these latter indicated below by M'. The potential energy is then fruitfully decomposed in the following contributions:

$$
\begin{aligned}
V_{non-local} &= w_{1,4} \sum_{1,4} V^1_{VdW} + \\
&+ w_{1>4}\Big(\sum_{M',M'} V^1_{VdW} + \sum_{M',C_\alpha} V^1_{VdW} + \sum_{M',Sc} V^1_{VdW} + \sum_{C_\alpha,Sc} V^1_{VdW}\Big) + \\
&+ w_{C_\alpha,C_\alpha} \sum_{C_\alpha,C_\alpha} V^2_{VdW} + w_{Sc,Sc} \sum_{Sc,Sc} V^2_{VdW}
\end{aligned} \tag{4}
$$

the first contribution is short range in nature and accounts for all kind of interactions 1-4 not excluded. The second block of terms consists of long-range interactions with respect to sequence position involving main chain atoms M' and mixed interactions (M',Sc),(M',C_α),(C_α,Sc). They share the same weighting factor $w_{1>4}$ and the functional form VdW^1 that will be described below. The last two terms specifically account for interactions involving (C_α,C_α) and (Sc,Sc) pairs. These interactions have been recently parametrized introducing a new LJ-like function[2], see below the description of the term V^2_{VdW}.

The VdW potential used in OPEP v.4 has two distinguished forms depending on the center of forces involved. The first type V^1_{VdW} is a standard 12-6 potential:

$$
V^1_{\mathrm{VdW}} = \epsilon_{i,j} \left(\left(\frac{r^0_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^0_{ij}}{r_{ij}} \right)^{6} \right) \tag{5}
$$

r_{ij} is the distance between particles i and j, r^0_{ij} = *(r^0_i + r^0_j)/2* with r^0_i and $r^0_j$the Van der Waals radius of particle i and j, and $\epsilon_{i,j}$ is the coupling contant.

The second potential function V^2_{VdW} reads as follows:

$$
V^2_{\mathrm{VdW}} = V_{\mathrm{AR}} - \epsilon_{i,j} \left(\frac{r^0_{ij}}{r_{ij}} \right)^{8} H(-\epsilon_{i,j}) \tag{6}
$$

where the repulsive final term holds for $\epsilon_{i,j} < 0$ while the attractive/repulsive interactions are described by the potential term:

$$
V_{\mathrm{AR}} = \epsilon_{i,j} \left(\left(\frac{G(r^0_{ij})}{r_{ij}} \right)^{6} \mathrm{e}^{-2r_{ij}} + A_0 \tanh[2\left(r_{ij} - r^0_{ij} - 0.5\right) - 1] \right) H(\epsilon_{i,j}) \tag{7}
$$

where *G(r^0_{ij})* is determined by imposing the value of the potential at a specific location r^0_{ij}[3]. The $H(\epsilon)$ indicates the Heavyside function, and equals to 1 for $\epsilon \geq 0$ and 0 otherwise. For $r_{ij} > r^0_{ij}$ the potential is controlled by the second part of Eq 7, which parameters ensure a

steeper profile with respect to a 12-6 Lennard-Jones potential (see Figure 3 of Ref.[3]). The constant $A_0 = 0.6563701$. The potentials for LyS/Glu and Arg/Asp pairs are reported in Figure 1.

The final contribution V_{HB} adds together a two-body ($V_{HB}^{(2)}$) and a four-body ($V_{HB}^{(4)}$) contribution:

$$V_{HB} = V_{HB}^{(2)} + V_{HB}^{(4)} \tag{8}$$

The two body term $V_{HB}^{(2)}$ is decomposed in short and long range HBs with respect to the position along the sequence:

$$V_{HB}^{(2)} = w_{1,4}^{HB} \sum_{i,j,i=j+4} \epsilon_{1,4}^{HB} \mu(r_{ij}) \nu(\alpha_{ij}) + w_{1>4}^{HB} \sum_{i,j,j>i+4} \epsilon_{1>4}^{HB} \mu(r_{ij}) \nu(\alpha_{ij}) \tag{9}$$

where

$$\begin{aligned} \mu \quad (r_{ij}) &= 5(\frac{\sigma}{r_{ij}})^{12} - 6(\frac{\sigma}{r_{ij}})^{10} \\ \nu \quad (\alpha_{ij}) &= \begin{cases} cos^2(\alpha_{ij}), & \alpha_{ij} > 90 \\ 0, & otherwise \end{cases} \end{aligned} \tag{10}$$

The sum runs over all residues separated by $j \geq i + 4$, r_{ij} is the $O..H$ distance between the carbonyl oxygen and the amide hydrogen in the backbone, α_{ij} is the angle $N\hat{H}O$ and σ is the equilibrium distance of the HB.

The four-body term takes the form of sum of weighted products of Gaussian functions each monitoring the existence of an HB on the basis of distance criteria. The Gaussian functions are calculated for each possible HB pairs and they give a contribution to the four-body term only if tight conditions on sequence-separation are verified:

$$\begin{aligned} V_{HB}^{(4)} &= w_{\alpha}^{HB} \sum \varepsilon_{\alpha}^{HB} \exp(-(r_{ij} - \sigma)^2/2) \ \exp(-(r_{kl} - \sigma)^2/2) \ \Delta(ijkl) \\ &\quad + w^{HB} \beta \sum \varepsilon_{\beta}^{HB} \exp(-(r_{ij} - \sigma)^2/2) \ \exp(-(r_{kl} - \sigma)^2/2) \ \Delta'(ijkl) \end{aligned} \tag{11}$$

The parameter $\Delta(ijkl)$ is set to 1 if residues (k, l) = (i+1, j+1), otherwise $\Delta(ijkl) = 0$. $\Delta'(ijkl)$ =1 if k and l satisfy either conditions: (k, l) = (i+2, j−2) or (i+2, j+2); otherwise $\Delta'(ijkl) = 0$. Thus these conditions help stabilize α-helices and β-sheets, independently of the (ϕ,ψ) dihedral angles, but also any segment satisfying the conditions on $ijkl$.

2. All atoms simulations

In this section we detail the protocols of the all-atom simulations that have been performed to complement the analysis of the CG modeling.

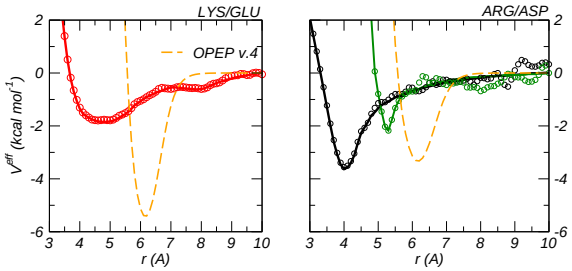

FIGURE 1. Two representative examples of the new potentials for ion-pair. The outcome of the IBI procedure (symbols) is smoothed to remove noise in the attractive and repulsive parts of the potential. The original potentials OPEP v.4 scaled by their own weights are reported in orange dashed lines.

2.1. **The amino-acid pairs.** The simulated amino-acid pairs are Arg/Asp, Arg/Glu, Lys/Asp, Lys/Glu capped by NH2 and COOH groups. We used the OPLS all-atom force field[4] to model the amino-acids and the TIP3P model to describe the water. The initial configurations of the pairs were obtained by placing each amino-acid randomly in a periodic octahedral boxes containing 1500 water molecules. The solvated systems were then minimized using the steepest descent method and were equilibrated for 1 ns at constant pressure (1 atm) and temperature (T = 300 K), respectively, using the Berendsen coupling[5] and velocity rescaling methods[6]. Subsequently, each system was simulated at constant temperature (T=300 K and T=360 K) and constant volume for 200-600 ns using GROMACS[7]. The longest simulation (600 ns) was carried out to sample correctly the tail of the Arg/Asp pair interaction. The equations of motion were integrated by using a leap-frog algorithm with a time step of 2 fs. Covalent bond lengths were constrained via the SHAKE[8] procedure. We used the particle-mesh Ewald method to treat the long-range electrostatics interactions[9]. The non-bonded interaction pair-lists were updated every 5 fs, using a cutoff of 1.2 nm. Data were collected every 2 ps.

2.2. **The REMD simulations of Peptide-C and β-Hairpin.** The 12-residue helical peptide of protein C (peptide C) and the C-terminal end of GB1 domain (β-hairpin) were simulated by REMD. We used the OPLS all-atom force field[4] to model the peptides and the TIP3P model to describe the water; for the β-hairpin we also tested the Amber and Gromos force fields. The initial configurations of the peptides were extracted from the crystallographic structure of the proteins (PDBcodes 5RSA and 2GB1, respectively) and

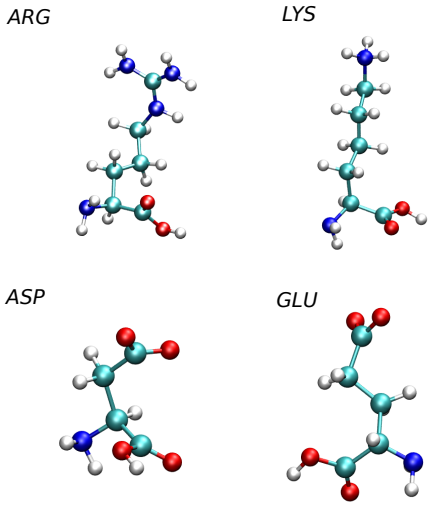

FIGURE 2. Positively and negatively charged amino-acids used in this work. The peptide terminals are capped with COOH and NH_2 groups. Color codes: nitrogen atoms are colored in blue, carbon atoms in cyan, oxygen atoms in red and hydrogen atoms in white.

placed in octahedral boxes containing 2500 water molecules. For the the C-peptide three single point mutations were done at K1A/E9L/M13A using Pymol tools. For REMD simulations, given the lowest (280 K) and highest (410 K) temperatures and requesting an acceptance ratio of $\approx$ 20 %, the temperatures of replicas were determined by using the method recently proposed by Patriksson and van der Spoel[10]. This resulted in 48 replicas for both the systems. Exchanges between replicas were attempted every 1.5 ps, large enough compared to the coupling time of the heat bath. Each replica was run for 50 ns and the data were collected every 2 ps.

2.3. **The protein HPr.** All atom simulations were done using CHARMM22[11] for protein and TIP3P model for water using the NAMD package. The system (Pdbcode 1CM2) was solvated in a rectangular box of 4202 water molecules and neutralized using 3 sodium atoms. The solvated neutralized structure was relaxed with 1000 steps of energy minimization. A short equilibration (1 ns with a timestep of 1 fs) was performed at 300 K before the production phase. The system was run at T=300K and p=1 atm for about 100 ns with a timestep of 2 fs. Electrostatic was handled by PME[9] and bonds involving hydrogens were kept rigid[8].

3. TWO STATE MODEL FOR ARGININE

In Fig. 3 we report the probability distributions $p(r, cos(\theta))$ for the amino-acid pairs Arg/Asp, Arg/Glu from all-atoms simulation, see Sec. 2.1. The distributions show an approximate two states localization as a function of the order parameter $cos(\theta) = \mathbf{n_i} \cdot \mathbf{n_j}$

where $\mathbf{n_i}$ and $\mathbf{n_j}$ represent the vectors connecting the C_α to the side-chain grain for the residues i and j respectively.

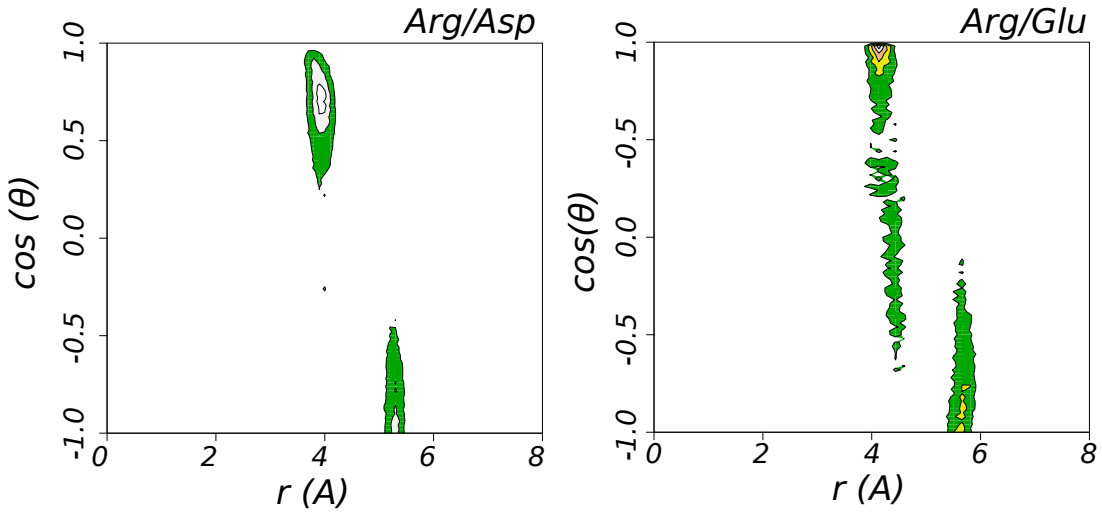

FIGURE 3. Probability distribution $p(r, cos(\theta))$ for the two pairs Arg/Asp and Arg/Glu.

4. Temperature effects

In Fig. 4 we report the Boltzmann inverted radial distribution functions computed for the four ion-pairs at T=300 K (black) and at T=360 K (red). Temperature increase is found to stabilizes -as expected- ion-pairs, the stability gain obtained for our atomistic model is of about 0.2 Kcal/mol.

5. The ion-pair distances

In Figs. 5 we report the distributions of distances between the charged amino acids in the β-hairpin comparing CG (top figure) and all atom (bottom figure) simulations.

6. The protein HPr

In top panel of Fig. 6 we report the the free-energy landscapes reconstructed considering the RMSD with respect to the crystallographic structure at several temperatures. The low-rmsd configuration superimposed to the crystallographic structure is represented in the lower panel. The fraction of the folded state p_F is obtained calculating the number of configurations of RMSD<6 A.

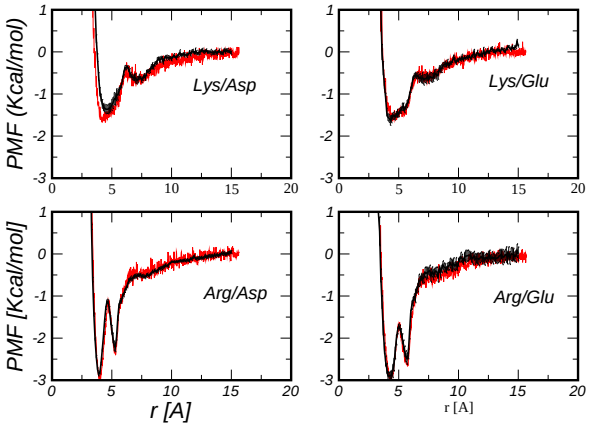

FIGURE 4. Potential of mean forces computed for the four ion-pairs at T=300 K (black) and T=360 K (red).

REFERENCES

[1] Maupetit, J.; Tuffery, P.; Derreumaux, P. *Proteins* **2007**, *69*, 394–408.
[2] Chebaro, Y.; Pasquali, S.; Derreumaux, P. *J Phys Chem B* **2012**, *116*(30), 8741–8752.
[3] Pasquali, S.; Derreumaux, P. *J Phys Chem B* **2010**, *114*(37), 11957–11966.
[4] Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J Phys Chem B* **2001**, *105*(28), 6474–6487.
[5] Berendsen, H.; Postma, J.; van Gunsteren, W.; Nola, A. D.; Haak, J. *J Chem Phys* **1984**, *81*, 3684–3690.
[6] Bussi, G.; Donadio, D.; Parrinello, M. *J Chem Phys* **2007**, *126*, 014101.
[7] Berendsen, H.; van der Spoel, D.; van Drunen, R. *Comp Phys Comm* **1995**, *91*, 43–56.
[8] Ryckaert, J.-P.; Ciccotti, G.; HJC, H. B. *J Comput Phys* **1977**, *23*, 327–341.
[9] Darden, T.; York, D.; Pedersen, L. *J Chem Phys* **1993**, *98*, 10089–10092.
[10] Patriksson, A.; van der Spoel, D. *Phys Chem Chem Phys* **2008**, *10*, 2073–2077.
[11] MacKerell, A. D.; Feig, M.; Brooks(III), C. L. *J Comput Chem* **2004**, *25*, 1400–1415.

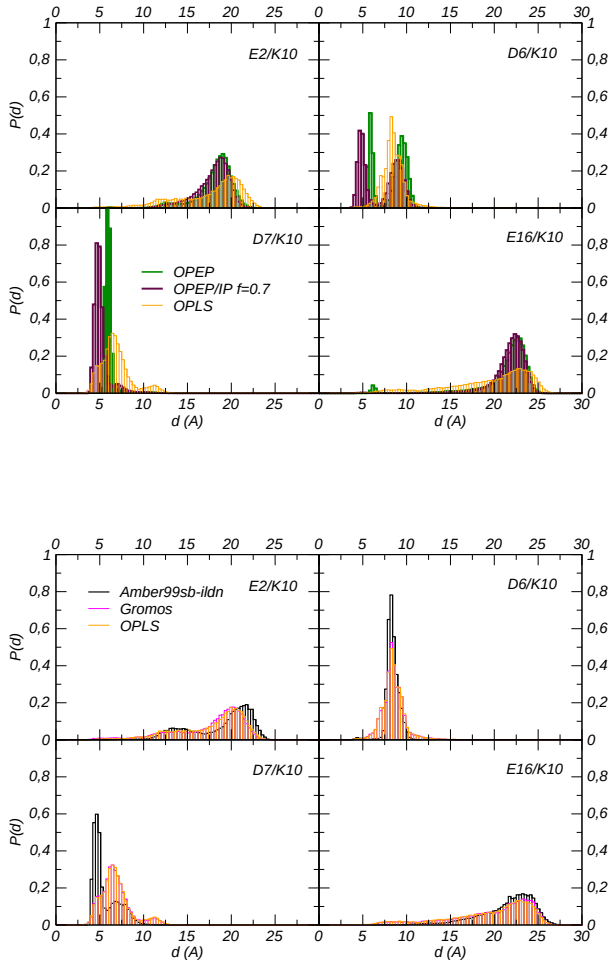

FIGURE 5. β-hairpin. Probability distribution of the distances between oppositely charged amino acid from OPEP v.4 (green) and OPEP/IP (maroon) and all atom OPLS (orange) REMD simulations. In the bottom panels we compare the results from Amber99sb-ildn, Gromos and OPLS REMD simulations. Data refer to T=300K.

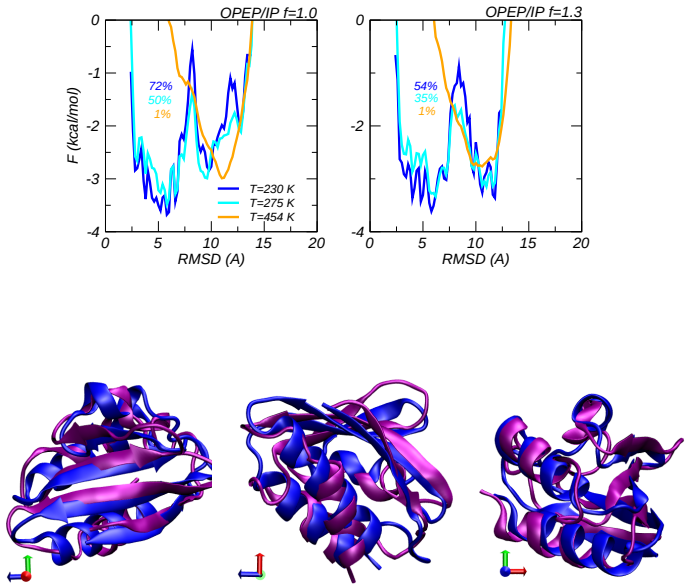

FIGURE 6. Top panel: Free energy surface along the RMSD reaction coordinate computed for OPEP/IP f=1.0 and f=1.3. In the lower panel we compare a representative configuration of the low-rmsd state (blue) with the crystallographic structure (purple)