Electronic Supplementary Material (ESI) for Chemical Society Reviews.

THE OPEP COARSE-GRAINED FORCE FIELD.

1. From OPEPv1 to OPEPv5

In this section we provide the analytical formulation of the OPEP Hamiltonian. The details of how parameters have been optimized are given in the chart flow in the main text. The OPEP Hamiltonian consists of three main terms:

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

where 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:

$$\begin{aligned} V_{local} &= w_b \sum_{bonds} K_b (r - r_0)^2 + w_a \sum_{angles} K_\alpha (\alpha - \alpha_0)^2 \\ &+ w_t \sum_{torsions} K_\phi (1 + cos(n\phi - \phi_0)) + w_a \sum_{imp} K_\omega (\omega - \omega_0)^2 \\ &+ w_{\phi,\psi} (\sum_{\phi} V_\phi + \sum_{\psi} V_\psi) \end{aligned} \tag{2}$$

where each contribution w_x represents a weighting factor opportunely optimized. The first three terms describe bond, angle and torsion fluctuations. The fourth term for improper dihedrals is used to maintain the cis or trans conformation of the peptide bond and the chirality of the amino acid. The final terms concern specifically the torsions along the backbone and were added in OPEPv2 and kept in the following versions to confine the dihedrals in the Ramachandran plots via harmonic barriers:

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

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:

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

where the first contribution is short range in nature and accounts for all 1-4 interactions along each torsional degree of freedom. The second block of terms is long-range with respect to sequence position and involve (M',M') interactions and mixed interactions (M',Sc),(M',C_α) and (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 (C_α,C_α) $j > i$+3 interactions and (Sc,Sc) $j > i$+1 interactions, where i and j are two amino acids in the sequence. These interactions have been recently parametrized introducing a new LJ-like function, see below the description of the term V^2_{VdW}.

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

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

where 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} in OPEPv1, v2 and v3 reads as follows:

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

where $H(\epsilon)$ indicates the Heavyside function, and equals to 1 for $\epsilon \geq 0$ (attractive/repulsive attractions) and 0 otherwise (purely repulsive interactions).

In OPEPv4, the second potential function V^2_{VdW} reads as follows:

$$
V^{2}_{\mathrm{VdW}} = V_{\mathrm{AR}} H(\epsilon_{i,j}) - \epsilon_{i,j} \left(\frac{r^0_{ij}}{r_{ij}} \right)^{8} H(-\epsilon_{i,j})
\tag{7}
$$

with the attractive/repulsive interactions described by :

$$V_{\mathrm{AR}} = \epsilon_{i,j} \left(\left(\frac{G(r^0_{ij})}{r_{ij}} \right)^6 \mathrm{e}^{-2r_{ij}} + A_0 \tanh[2\left(r_{ij} - r^0_{ij} - 0.5\right) - 1] \right) \tag{8}$$

where $G(r^0_{ij})$ is determined by imposing the value of the potential at a specific location r^0_{ij}. For $r_{ij} > r^0_{ij}$ the potential is controlled by the second part of Eq. 8, which parameters ensure a steeper profile with respect to a 12-6 Lennard-Jones potential. The constant A_0 is set to 0.6563701. In OPEPv4, we also changed the $\epsilon_{i,j}$ coupling constants of some interaction pairs present in α-helices. Indeed, using a statistical analysis on 150 PDB structures, we identified several $(i,\ i+3)$ and $(i,\ i+4)$ Sc-Sc contacts frequently observed in α-helices that were not distinguished in OPEPv3, which used the same coupling constant for each side-chain pair independently of the i,j separation. We have identified 11 interactions which have now a different coupling term if they are of $(i,\ i+3)$ and $(i,\ i+4)$ types or $(i,\ i+2)$ and $(i,\ i>4)$ types. The new $(i,\ i+3)$ and $(i,\ i+4)$ interactions include Lys-Glu, Lys-Asp, Glu-Arg, and Asp-Arg. For $(i,\ i+4)$, we also identified the following Lys-Gln, Lys-Leu, Ala-Arg, Ala-Gln, Ala-Glu, Leu-Glu, and Ile-Lys interactions. All these 11 interactions are no longer considered repulsive in OPEPv4 and OPEPv5 as in OPEPv1,v2 and v3 but attractive-repulsive.

The final contribution V_{HB} in all OPEP versions adds together a two-body ($V^{(2)}_{HB}$) and a four-body ($V^{(4)}_{HB}$) contribution:

$$V_{HB} = V^{(2)}_{HB} + V^{(4)}_{HB} \tag{9}$$

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

$$V^{(2)}_{HB} = w^{HB}_{1,4} \sum_{i,j,i=j+4} \epsilon^{HB}_{1,4} \mu(r_{ij})\nu(\alpha_{ij}) + w^{HB}_{1>4} \sum_{i,j,j>i+4} \epsilon^{HB}_{1>4} \mu(r_{ij})\nu(\alpha_{ij}) \tag{10}$$

where

$$\mu \quad (r_{ij}) = 5(\frac{\sigma}{r_{ij}})^{12} - 6(\frac{\sigma}{r_{ij}})^{10} \tag{11}$$

$$\nu \quad (\alpha_{ij}) = \begin{cases} cos^2(\alpha_{ij}), & \alpha_{ij} > 90 \\ 0, & otherwise \end{cases}$$

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 NHO 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:

$$\begin{aligned} V_{HB}^{(4)} &= w_{\alpha}^{HB} \sum \varepsilon_{\alpha}^{HB} \exp(-(r_{ij}-\sigma)^2/2)\ \exp(-(r_{kl}-\sigma)^2/2)\ \Delta(ijkl) \\ &\quad + w^{HB}\beta \sum \varepsilon_{\beta}^{HB} \exp(-(r_{ij}-\sigma)^2/2)\ \exp(-(r_{kl}-\sigma)^2/2)\ \Delta'(ijkl) \end{aligned} \tag{12}$$

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 $ijkl$.

In OPEPv5, the ionic interactions between side chains Lys/Asp, Lys-Glu, Arg/Asp and Arg/Glu are not described by Eq 7 but replaced by *ad hoc* pair potentials $V^{eff}(r)$ derived by using the iterative Boltzmann inversion (IBI) procedure and targeting the radial distribution functions between the center of mass of the side chains obtained by extended atomistic MD simulations of the ionic pairs in explicit solvent. The potentials are presently tabulated as a function of the separation distance on a grid of resolution $\delta r = 0.1$ Å. For the pairs Arg/Asp and Arg/Lys the potential depends also on the relative orientation between the two side-chains as a result of the different hydrogen bond patterns between the ionic groups. For each pair two possible orientations exist (head-to-head and parallel configuration), and for each of them a distance dependent potential $V_{1(2)}^{eff}(r)$ is constructed via IBI. The orientation dependency is handled parametrically by considering a mixing rule:

$$V_{Arg/x}^{eff} = a_1^{\theta} * V_1^{eff} + a_2^{\theta} * V_2^{eff} \tag{13}$$

where x=Asp,Glu, and the coefficients $a_{1(2)}^{\theta}$ are weights that depend on the side chains relative orientation. In the present implementation the weights $a_{1(2)}$ are parameters and do not contribute directly to the force via their angular dependence.

Since the absolute energy contribution from all the others OPEP terms depend on the weigh factors w_x also the ion pair potentials are weighted by an optimal scaling factor. The optimal range of values was individuated by extensive empirical tests on target systems.

2. OPEP ACCESSIBILITY

All the parameters of OPEPv3 and OPEPv4 are publicly available. They can be obtained via the web server **http://opep.galaxy.ibpc.fr**, which returns for your sequence or your all-atom pdb file, a CG OPEP pdb file, the scale.dat file which contains the numerical values of the weights, a list file containing the information for the last two terms of Eq. (4) and a topology file. The OPEP force field is presently implemented in an *in-house* developed code available upon request for cooperative collaborations. In the near future, an homogenous version of the code with a manual will be released.