

Biophysical Journal, Volume 96

Supporting Material

A free-energy approach for all-atom protein simulation

Abhinav Verma and Wolfgang Wenzel

PFF02: An Overview

All atom force field (except apolar CH_n groups)

No vibration terms

All bond lengths and peptide planes are kept fixed

Dihedral angles of main chain and side chain rotate

Protein Model

The atoms in the polypeptide chain are classified according to their chemical characteristics and are then used to obtain different force field parameters.

Residue/ Chain Part	Atom Type	Potential Type	Group Type	Index (ρ)	Charge (φ)
main chain, N-terminus	NT	n3	NTM	7	
	HEN1	h3	NTM	7	
	HEN2	h3	NTM	7	
	HEN3	h3	NTM	7	
	CA	cb	NTM	7	
	C	cd	CO	2	+0.38
	O	odd	CO	2	-0.38
main chain	N	n	NH	1	-0.28
	HN	hn	NH	1	+0.28
	CA	caa	CA	7	
	C	cd	CO	2	+0.38
	O	odd	CO	2	-0.38
main chain, C-terminus	N	n	NH	1	-0.28
	HN	hn	NH	1	+0.28
	CA	caa	END	7	
	C	cd	END	7	
	O	odd	END	7	
	OT	o-	END	7	
ALA	CB	cme	CB	7	
ASP	CB	cme	ASP	6	-0.20

Residue/ Chain Part	Atom Type	Potential Type	Group Type	Index (ρ)	Charge (q)
ASP	CG	c-	ASP	6	+0.34
ASP	OD1	o-	ASP	6	-0.57
ASP	OD2	o-	ASP	6	-0.57
ASN	CB	cme	CB	7	
ASN	CG	c-	COG	5	+0.38
ASN	OD1	od	COG	5	-0.38
ASN	ND2	n2	ND2	4	-0.56
ASN	HNA	h2	ND2	4	+0.28
ASN	HNB	h2	ND2	4	+0.28
ARG	CB	cme	CB	7	
ARG	CG	cme	CG	7	
ARG	CD	cme	ARG	6	+0.19
ARG	NE	na	ARG	6	-0.50
ARG	HNE	h1	ARG	6	+0.37
ARG	CZ	cb	ARG	6	+0.46
ARG	NH1	n+	ARG	6	-0.50
ARG	NH2	n+	ARG	6	-0.50
ARG	HHA	h+	ARG	6	+0.37
ARG	HHB	h+	ARG	6	+0.37
ARG	HHC	h+	ARG	6	+0.37
ARG	HHD	h+	ARG	6	+0.37
CYS	CB	cme	CB	7	
CYS	SG	s	SG	7	
GLU	CB	cme	CB	7	
GLU	CG	cme	GLU	6	-0.20
GLU	CD	c-	GLU	6	+0.34
GLU	OE1	o-	GLU	6	-0.57
GLU	OE2	o-	GLU	6	-0.57
GLN	CB	cme	CB	7	
GLN	CG	cme	CG	7	
GLN	CD	c-	COD	5	+0.38
GLN	OE1	od	COD	5	-0.38

Residue/ Chain Part	Atom Type	Potential Type	Group Type	Index (ρ)	Charge (q)
GLN	NE2	n2	NE2	4	-0.56
GLN	HNA	h2	NE2	4	+0.28
GLN	HNB	h2	NE2	4	+0.28
HIS	CB	cme	HIS	6	+0.11
HIS	CG	cr	HIS	6	+0.17
HIS	ND1	n1	HIS	6	-0.50
HIS	HD1	hp	HIS	6	+0.37
HIS	CD2	cr	HIS	6	+0.33
HIS	CE1	cr	HIS	6	+0.65
HIS	NE2	n1	HIS	6	-0.50
HIS	HE2	hp	HIS	6	+0.37
ILE	CB	cme	CB	7	
ILE	CG1	cme	CG1	7	
ILE	CG2	cme	CG2	7	
ILE	CD1	cme	CD1	7	
LEU	CB	cme	CB	7	
LEU	CG	cme	CG	7	
LEU	CD1	cme	CD1	7	
LEU	CD2	cme	CD2	7	
LYS	CB	cme	CB	7	
LYS	CG	cme	CG	7	
LYS	CD	cme	LYS	6	+0.12
LYS	CE	cb	LYS	6	+0.30
LYS	NZ	n3	LYS	6	-0.50
LYS	HZA	h3	LYS	6	+0.36
LYS	HZB	h3	LYS	6	+0.36
LYS	HZC	h3	LYS	6	+0.36
MET	CB	cme	CB	7	
MET	CG	cme	CG	7	
MET	SD	s	SD	7	
MET	CE	cme	CE	7	
PHE	CB	cme	CB	7	

Residue/ Chain Part	Atom Type	Potential Type	Group Type	Index (ρ)	Charge (q)
PHE	CG	cr	CG	7	
PHE	CD1	cr	CD1	7	
PHE	CD2	cr	CD2	7	
PHE	CE1	cr	CE1	7	
PHE	CE2	cr	CE2	7	
PHE	CZ	cr	CZ	7	
PRO (main chain)	N	n	PRO	7	
	CA	caa	PRO	7	
	C	cd	CO	2	+0.38
	O	odd	CO	2	-0.38
PRO	CB	cme	CB	7	
PRO	CG	cme	CG	7	
PRO	CD	cme	PRO	7	
SER	CB	cme	OGS	3	+0.03
SER	OG	os	OGS	3	-0.38
SER	HOG	hd	OGS	3	+0.35
THR	CB	cme	OGT	3	+0.03
THR	OG1	os	OGT	3	-0.38
THR	HOG	hd	OGT	3	+0.35
THR	CG2	cme	CG2	7	
TRP	CB	cme	CB	7	
TRP	CG	cr	CG	7	
TRP	CD1	cr	CD1	7	
TRP	CD2	cr	CD2	7	
TRP	NE1	nw	NEH	6	-0.28
TRP	HNE	h1	NEH	6	+0.28
TRP	CE2	cr	CE2	7	
TRP	CE3	cr	CE3	7	
TRP	CZ2	cr	CZ2	7	
TRP	CZ3	cr	CZ3	7	
TRP	CH2	cr	CH2	7	
TYR	CB	cme	CB	7	

Residue/ Chain Part	Atom Type	Potential Type	Group Type	Index (g)	Charge (q)
TYR	CG	cr	CG	7	
TYR	CD1	cr	CD1	7	
TYR	CD2	cr	CD2	7	
TYR	CE1	cr	CE1	7	
TYR	CE2	cr	CE2	7	
TYR	CZ	cr	OHH	3	+0.03
TYR	OH	oy	OHH	3	-0.38
TYR	HOH	hd	OHH	3	+0.35
VAL	CB	cme	CB	7	
VAL	CG1	cme	CG1	7	
VAL	CG2	cme	CG2	7	

Table 1. Types of atoms and groups for different residues. The meaning of the electrostatic parameters (g and q) are explained below in the text.

Interaction Terms

1. Lennard-Jones interaction

The van der Waals interactions are included in the force field as a Lennard-Jones 6-12 potential:

$$V_{\text{L}}(r_{ij}) = V_{\text{L}}^{(0)} \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^6 \right]$$

with further renormalization

$$V_{\text{L}}(r_{ij}) = \frac{C}{(C + V_{\text{L}}^{(0)}) / V_{\text{L}}(r_{ij}) + 1}$$

Here i, j represent the atoms included in the force field, r_{ij} is the distance between these atoms and R_{ij} is the Lennard-Jones radius $R_{ij} = \sqrt{R_i R_j}$; $V_{\text{L}}^{(0)} = 0.01$ kcal/mole; $C = 30$ kcal/mole. The parameters for the Lennard-Jones potential were derived from a potential of mean force approach to experimental data by fitting short range (2Å-5Å) radial distributions to a set of 138 proteins which represent a wide span

of different protein folds.

Potential Type	R_{ij} , Å	σ_j , cal/mol Å ²
CAA	4.10	+60
CME	4.10	+84
CR	3.49	+93
CB C-	4.10	-6
CD	4.10	+15
N N1 NW NA N+	3.55	-30
N2	3.55	+15
N3	3.55	-45
OS OY ODD	3.10	-30
O- OD	3.10	-15
S	3.80	+84
H1 H2 H3 H+ HD HP	1.95	as by bound partner
HN	2.15	as by bound partner

Table 2: Lennard-Jones radii and the solvation enthalpies for different potential types.

2. Electrostatics

The electrostatics is included using Coulomb potential. This interaction can be further splitted into main chain and side chain contributions. The dielectric constants used here are group-specific dielectric constants. The interaction potential between two groups a and b is given by

$$V_{\text{el}}^{(a,b)} = \frac{e^2 N_A}{4\pi\epsilon_0} \frac{s_a s_b}{g(a,g(b))} \sum_{\substack{i \in a \\ j \in b}} \frac{q_i q_j}{r_{ij}},$$

where i, j represent the atoms included in the force field, q_i and q_j are the corresponding partial charges, r_{ij} is the distance between these atoms, $g(a, g(b))$ are group-specific dielectric constants, e is the electron charge, N_A is the Avogadro number, ϵ_0 is the electric constant. The group specific parameter s_a is equal to 1 for the uncharged groups and is proportional to the relative solvent accessible surface for charged groups:

$$s_a = \frac{4}{A_{a,\max}} \sum_{i \in a} A_i,$$

where A_i is the solvent accessible surface area of atom i and the values of the parameter $A_{a,\max}$ are given in Table 3.

Group type	ARG	ASP	GLU	HIS	LYS
$A_{a,\max}, \text{\AA}^2$	168	105	117	143	135

Table 3. Parameters $A_{a,\max}$ for the charged groups.

The group specific dielectric constants are chosen according to different types of electrostatic interaction. This represents the characteristics of the atoms as being part of different amino acids and takes their specific partial charges, orientation or accessibility to the solvent into account. This is a strong approximation to the real environment, as only the interacting amino acids and not the complete environment is taken into consideration. The group indices $g(a)$ and $g(b)$ are given in Table 1, the parameters $\epsilon_{g(a)g(b)}^{-1} = \epsilon_{g(b)g(a)}^{-1}$ are given in Table 4. This parameterization excludes some parts or even complete side chains (like PHE, GLY, MET, PRO) from contributions to the electrostatics.

	$g(b) = 1$	$g(b) = 2$	$g(b) = 3$	$g(b) = 4$	$g(b) = 5$	$g(b) = 6$
$g(a) = 1$	0.103	0.103	0.000	0.287	0.287	0.086
$g(a) = 2$		0.103	0.324	0.287	0.287	0.062
$g(a) = 3$			0.000	0.000	0.324	0.091
$g(a) = 4$				0.287	0.287	0.086
$g(a) = 5$					0.287	0.062
$g(a) = 6$						0.050

Table 4: Parameters for the inverse group specific dielectric constants $\epsilon_{g(a)g(b)}^{-1} = \epsilon_{g(b)g(a)}^{-1}$. For $a, b = 7$, $\epsilon_{g(a)g(b)}^{-1} = 0$. For CO and NH groups belonging to the same peptide bond, $\epsilon^{-1} = 0$.

The parameters for $g(i) = 1, 2$ are used to describe the hydrogen bonding for the main chain as dipole-dipole interaction and constitute the biggest contribution from electrostatics. $g(i) = 3, 4, 5$ describe interactions of the partially charged OH, CO and NH₂ groups of the (ASN, GLN, SER, THR, TRP)-sidechains, which are smaller in their contributions. The interaction of the charged COO⁻ and NH⁽⁺⁾ x of (ASP, GLU, ARG, LYS, HIS, TRP) are the smallest contributions to the electrostatic interaction. The electrostatics of the sidechains contribute only in minor quantities to the total free energy of the protein.

The total electrostatic energy is the sum over all pairs of groups:

$$V_{\text{el}} = \sum_{a,b} V_{\text{el}}^{(a,b)}.$$

3. Hydrogen bonding

The interaction partners are the CO and NH groups of the backbone (not the groups END, NTM and PRO). NH can have only one partner. CO can have two partners on the opposite sides of the plane C CO. The groups belonging to neighboring (and the same) residues cannot be partners.

In addition to the pure electrostatic term identical to that described in the previous section, the interaction between partners is characterized by another term, which approximates the chemical energy of a hydrogen bond. It is defined by the following parameters: θ , the angle between the vectors $\vec{\text{HN}}$ and $\vec{\text{OH}}$ (degrees); α , the angle between the vectors $\vec{\text{CO}}$ and $\vec{\text{HN}}$ (degrees); r ; the distance between H and O (Å). The interaction energy for two partners is given by

$$V_{\text{hb}}^{(p)} = V_{\text{hb}}^{(0)} \tanh \left[1.1 s_{2.4;0.075}(r) s_{45;5}(\alpha) s_{40;5}(\beta) s_{1;0.05} \left(\sqrt{\left(\frac{\alpha}{45}\right)^2 + \left(\frac{\beta}{35}\right)^2} \right) \right]$$

where $V_{\text{hb}}^{(0)} = -2.89$ kcal/mole and

$$s_{A,B}(x) = \frac{1}{2} \left(1 - \tanh \left(\frac{x-A}{B} \right) \right).$$

The groups CO and NH are considered to be partners if the distance r does not exceed 3.3 Å. In case when more group pairs than allowed satisfy this criterion, the pair with the lowest interaction energy has the preference. The total hydrogen bonding energy is the sum of the energies of all pairs of partners.

4. Solvation Effect

In PFF02 solvent interaction is modelled by calculating the solvent accessible surface area (SASA) of each atom of the protein. The contribution from solvent is then given by

$$V_{solvent} = \sum_j k_j \text{PT}(j) A_j$$

with

$$A_j = 4\pi \left(\frac{R_{ij}}{2} + r_w \right)^2 p_j,$$

where, i counts all atoms, $\text{PT}(j)$ is the potential type of atom j , $\text{PT}(j)$ gives the atomic solvent parameter (ASP) according to potential type and A_j gives the solvent accessible surface area (SASA) of atom j . The parameter p_j is the part of the extended surface of atom j that is not within the extended surface of any other atom. "Extended" means that the atom radius $R_{ij}/2$ is extended by the "water radius" $r_w = 1.4 \text{ \AA}$.

The coefficient k_j is a scaling factor defined as follows. Its "basic" value is 1.0. Then it is multiplied:

- by 0.70 for the atoms of the terminal residues;
- additionally by 0.70 for the atoms in the groups NTM and END;
- by 0.80 for the side chain atoms of ALA, ILE, PRO, VAL;
- by 0.64 for the side chain atoms of LEU, MET;
- by 0.96 for the side chain atoms of PHE;
- by 0.60 for the side chain atoms of TRP;
- by 0.80 for the side chain atoms of type *cme*;
- by 2.00 for the side chain atoms of type *mw*;
- by 0.8 for all backbone atoms.

The parameters PT are calculated using the above equation to fit the experiment. The parameters R_{ij} and PT are given in Table 2.

5. Local Electrostatics

Local electrostatics is defined as the electrostatic energy of the main chain CO and NH groups of a residue n arising from the interactions with main chain CO and NH groups within that residue and with the adjoining peptide groups (for the residue PRO the group PRO is taken instead of NH). Thus for $(\text{NH})_n$ interactions are calculated for $(\text{CO})_{n-2}$, $(\text{NH})_{n-1}$, $(\text{CO})_n$ and $(\text{NH})_{n+1}$ and for $(\text{CO})_n$ the interactions are calculated for $(\text{CO})_{n-1}$, $(\text{NH})_n$, $(\text{CO})_{n+1}$ and $(\text{NH})_{n+2}$. For group a belonging to residue n the interaction energy is

$$V_{\text{loc}}^{(a)} = \frac{e^2 N_A}{4\pi\epsilon_0} \frac{1}{2} \sum_b \sum_{\substack{j \in a \\ j \in b}} \frac{q_i q_j}{r_{ij}}$$

where q_i is the charge of atom i , r_{ij} is the distance between the atoms i and j , and the parameter n is an

amino acid specific parameter given in Table 5. The summation \sum_b is performed over all partners of group a , except PRO. The total energy of local interaction is (summation over all groups CO, NH, PRO)

$$V_{\text{loc}} = \sum_a V_{\text{loc}}^{(a)}$$

Residue		Residue		Residue		Residue	
ALA	0.17	GLU	0.11	LEU	0.29	SER	0.17
ASP	-0.01	GLN	0.21	LYS	0.19	THR	0.18
ASN	0.11	GLY	0.12	MET	0.34	TRP	0.21
ARG	0.22	HIS	0.21	PHE	0.37	TYR	0.28
CYS	0.23	ILE	0.43	PRO	0.00	VAL	0.40

Table 5: Amino acid specific parameters for local electrostatic interaction.

Dihedral Angle Term

The dihedral angle term for residue n is defined as

$$V_{\text{tor}}^{(n)} = -0.6 \exp \left[-0.005(\phi_n + 110)^2 - 0.00125(\phi_n - 130)^2 \right], \text{ kcal/mole .}$$

Here ϕ_n and ψ_n are the backbone dihedral angles (degrees). For GLY, PRO and terminal residues, $V_{\text{tor}}^{(n)} = 0$. The entire term for the whole chain is (summation over all residues)

$$V_{\text{tor}} = \sum_a V_{\text{tor}}^{(a)}$$