Nomenclature for Bayesian sampling driven classical mechanics force field parameterization paper

Bryce C. Manubay

March 13, 2017

I. Molecules

Index	SMILES	C_count	O_count	AlkEthOH_id	IUPAC_names
0	C	1	1	AlkEthOH_c0	methane
15	CC	2	1	AlkEthOH_c38	ETHANE
339	CC(C)O	3	2	AlkEthOH_c488	Propan-2-ol
603	CCC	3	1	AlkEthOH_c901	PROPANE
789	CCCO	3	2	AlkEthOH_c11…	Propan-1-ol
804	CCO	2	2	AlkEthOH_c11…	ethanol
805	CCOC	3	2	AlkEthOH_c11…	Methoxyethane
896	CO	1	2	AlkEthOH_c12…	methanol
897	COC	2	2	AlkEthOH_c12…	Methoxymetha…
912	O	0	2	AlkEthOH_c13…	oxidane
105	C1COC1	3	2	AlkEthOH_r131	Oxetane

Figure 1: The molecules being used as the test set for this initial parameterization. Each row entry includes the SMILES string, C and O composition, ID from the AlkEthOH set and the IUPAC name for a given molecule.

II. Parameters **(Names given by the forcefield, will indicate differently in paper. See Nomenclature section)

A. Bond:
   1. ‘k‘: the bonded force constant $\left(\frac{kcal}{mol \cdot ^2}\right)$
   2. ‘length‘: the equilibrium bond length $()$

B. Angle:
   1. ‘k‘: the angular force constant $\left(\frac{kcal}{mol \cdot rad^2}\right)$
   2. ‘length‘: the equilibrium bond angle (rad)

C. Angle:
   1. ‘idivfi‘: barrier height divisor for torsional term i
   2. ‘ki‘: force constant for torsional term i $\left(\frac{kcal}{mol}\right)$
   3. ‘phasei‘: phase angle for torsional term i $(\deg)$
   4. ‘periodicityi‘: periodicity of torsional term i

III. **How many parameters?**

A. Bonds (total of 5 unique SMIRKS -> 10 parameters)
   1. $[\#6X4:1]-[\#6X4:2]$ (C-C)
   2. $[\#8:1]-[\#1:2]$ (O-H)
   3. $[\#6X4:1]-[\#8X2:2]$ (C-O)
   4. $[\#6X4:1]-[\#1:2]$ (C-H)
   5. $[\#6X4:1]-[\#8X2H0:2]$ (C-O where the O is bonded to another C, so ether type C-O)

B. Angles (total of 3 unique SMIRKS -> 6 parameters)
   1. $[\#1:1]-[\#6X4:2]-[\#1:3]$
   2. $[*:1]-[\#8:2]-[*:3]$
   3. $[*:1]\,[\#6X4:2]-[*:3]$

C. Torsions (torsion parameters need to be counted on a individual SMIRKS basis)
   1. $[\#1:1]-[\#6X4:2]-[\#6X4:3]-[\#6X4:4]$
      a. 1 term
      b. 4 total parameters
   2. $[\#6X4:1]-[\#6X4:2]-[\#8X2H0:3]-[\#6X4:4]$
      a. 2 terms
      b. 8 total parameters
   3. $[*:1]-[\#6X4:2]-[\#6X4:3]-[*:4]$
      a. 1 term
      b. 4 total parameters
   4. $[\#1:1]-[\#6X4:2]-[\#6X4:3]-[\#1:4]$
      a. 1 term
      b. 4 total parameters
   5. $[*:1]\,[*:2]\,[*:3]\,[*:4]$

a. 1 term
b. 4 total parameters

6. $[\#1:1]-[\#6X4:2]-[\#6X4:3]-[\#8X2:4]$
   a. 2 terms
   b. 8 total parameters
7. $[*:1]-[\#6X4:2]-[\#8X2:3]-[\#1:4]$
   a. 1 term
   b. 4 total parameters
8. $[\#6X4:1]-[\#6X4:2]-[\#8X2H1:3]-[\#1:4]$
   a. 2 terms
   b. 8 total parameters

D. 60 total parameters

IV. **Nomenclature:**

A. Probability Terms:
   1. $\hat{\Theta}$: denotes some set of parameters
   2. $Pr\left(\hat{\Theta}|O\right)$: posterior probability distribution; prob of some parameter set given observables
   3. $L\left(O|\hat{\Theta}\right)$: Likelihood function; prob of observing some observable given parameters
   4. $pr\left(\hat{\Theta}\right)$: prior probability; probability of parameter set based on prior knowledge

B. Parameters:
   1. Bonds:
      a. k_B: the bonded force constant $\left(\frac{kcal}{mol\cdot^2}\right)$
      b. x_0: the equilibrium bond length $()$
   2. Angles:
      a. k_θ: the angular force constant $\left(\frac{kcal}{mol\cdot rad^2}\right)$
      b. θ_0: the equilibrium bond angle (deg)
   3. Torsions:
      a. $idivf_i$: barrier height divisor for torsional term i
      b. k_i: force constant for torsional term i $\left(\frac{kcal}{mol}\right)$
      c. ψ_i: phase angle for torsional term i (deg)
      d. τ_i: periodicity of torsional term i

C. Observables:
   1. Bonds:
      a. μ_B: mean bond length of a bond distribution $()$
      b. σ_B^2: variance in bond length of a bond distribution $\left(^2\right)$

2. Angles:
    a. μ_θ: mean bond angle of an angle distribution (rad)
    b. σ_θ^2: variance in bond angle of an angle distribution $\left(rad^2\right)$
3. Torsions:
    a. A_i: Fourier coefficient of the i^{th} term
    b. ψ_i: phase angle of the i^{th} term
    c. τ_i: periodicity of the i^{th} term