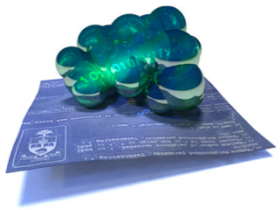
A
Bioinformatics
Course

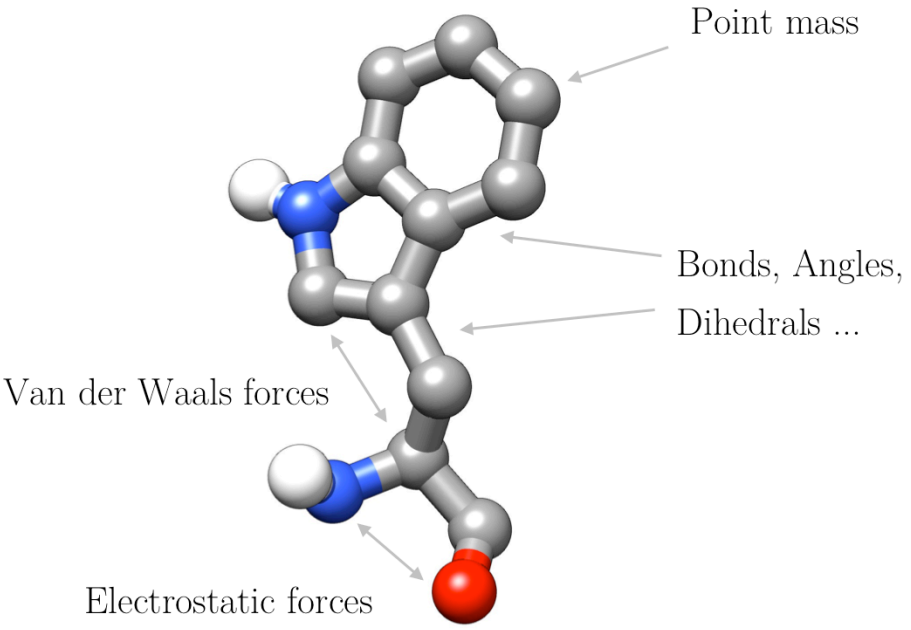
Molecular Forcefields



Boris Steipe

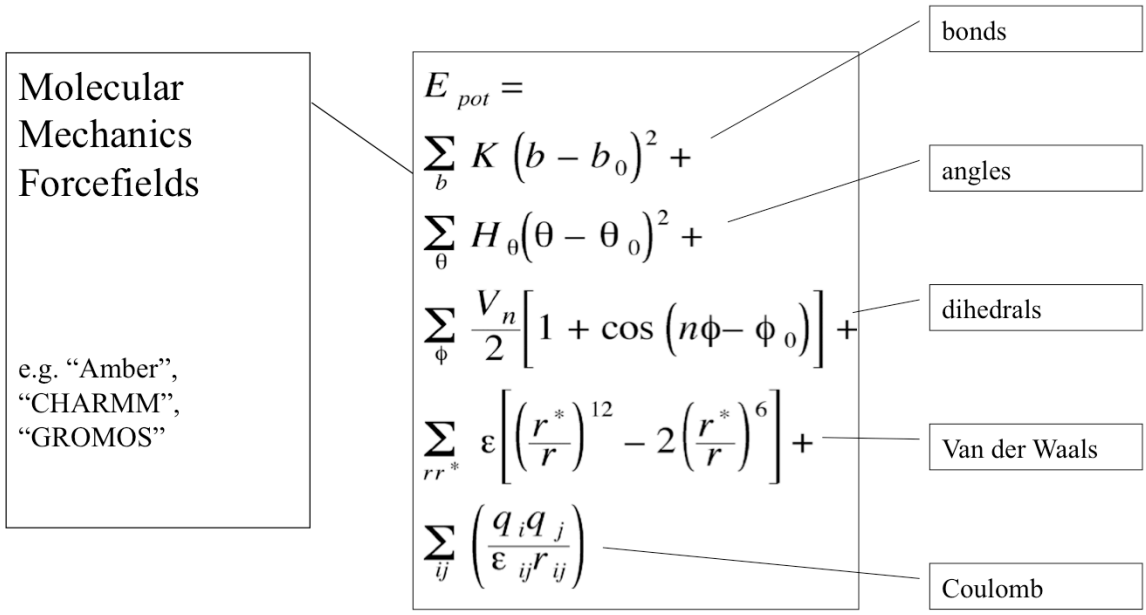
department of biochemistry – department of molecular genetics
University of Toronto

Forcefield: functional Form



To quantify the energy of particular conformations of biomolecules, we approximate the atoms as point-masses and calculate the forces that act between them. This includes bonded interactions that capture the geometry of the molecule, and non-bonded interactions that capture attractive or rpulsive forces that act through space.

Forcefield: functional Form



These interactions are summarized in a "molecular mechanics potential".

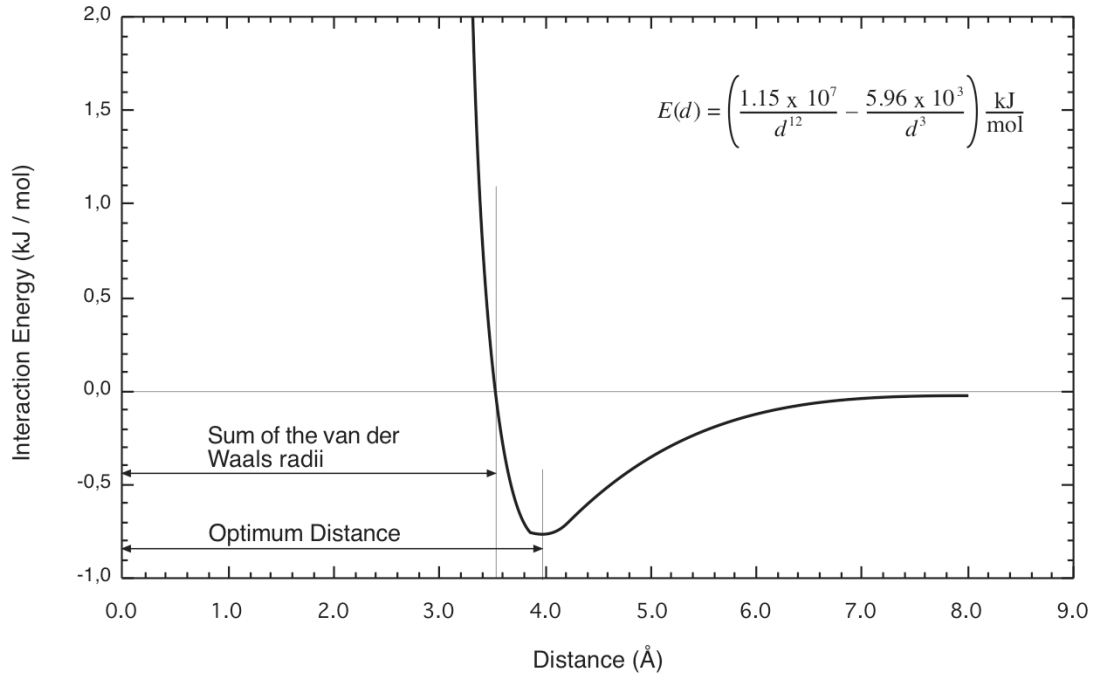
The formula looks complicated, but is actually only a simple sum of rather trivial functions. In general, a force is calculated as a deviation from idealized bond, angle and dihedral terms (the *bonded interactions*) plus the *non-bonded interaction terms* for Van der Waals forces and electrostatic interactions. The latter two add considerable computational requirements to the calculation, since they have to be considered over distances that potentially cover a large number of atoms.

Note that entropy is *not* explicitly considered in this function!

To evaluate the potential energy of a momentary configuration, consider all atoms in turn, look up idealized bond-lengths, and stiffnesses in a large dictionary of values that have been calculated from first principles and adjusted to reproduce biophysical observables. Then evaluate the expressions and sum the individual terms.

Inaccuracies can come from errors in the forcefield parameters and from multi-body interactions that are not explicitly taken into account. For example some bonds are polarized in an electrostatic field, and that changes the charge distribution.

VanDer Waals Forces

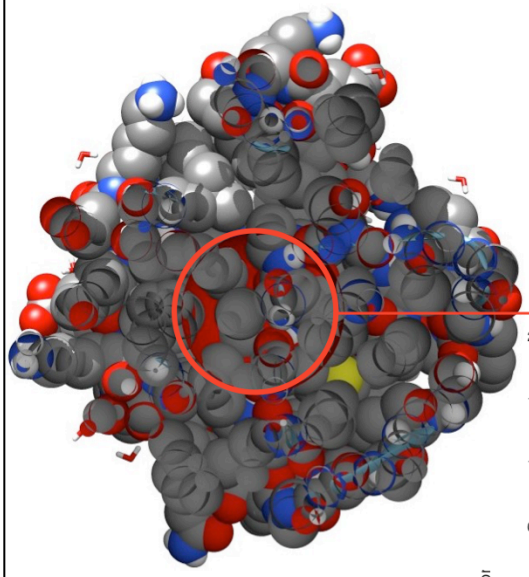


Carbon-Carbon interaction energy as a function of distance for two atom centers, modelled as a Lennard-Jones potential function.

Van der Waals forces (also called London dispersion forces) are significant over distances that can include dozens of atoms. Thus, even though the individual pair interactions are weak, they sum to significant amounts. The potential function is zero when the two atoms are separated by the sum of their Van der Waals radii, i.e. the atoms just "touch" each other – in fact that is how the Van der Waals radii are defined in the first place. If the atom centres are moved closer together, the energy rises rapidly. However at a slightly larger separation, the interaction energy is negative: atoms are "sticky".

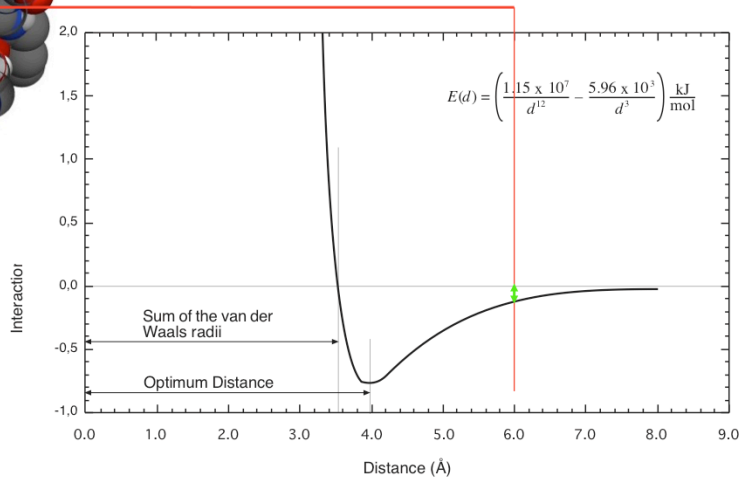
Protein stability is determined by the free energy difference between the folded and the unfolded state. Naïvely, one would argue that the unfolded protein – being immersed in liquid water – would make just as many atomic contacts as the folded protein. That is however not the case: side-chain packing in natural proteins is highly complementary, and the void volume in the core of a folded protein is significantly smaller than that of a liquid. The precise shape of the protein's turns, strands, helices and convolutions matters. A lot. Thus Van der Waals interactions provide a part of the driving forces behind protein folding.

VanDer Waals Forces



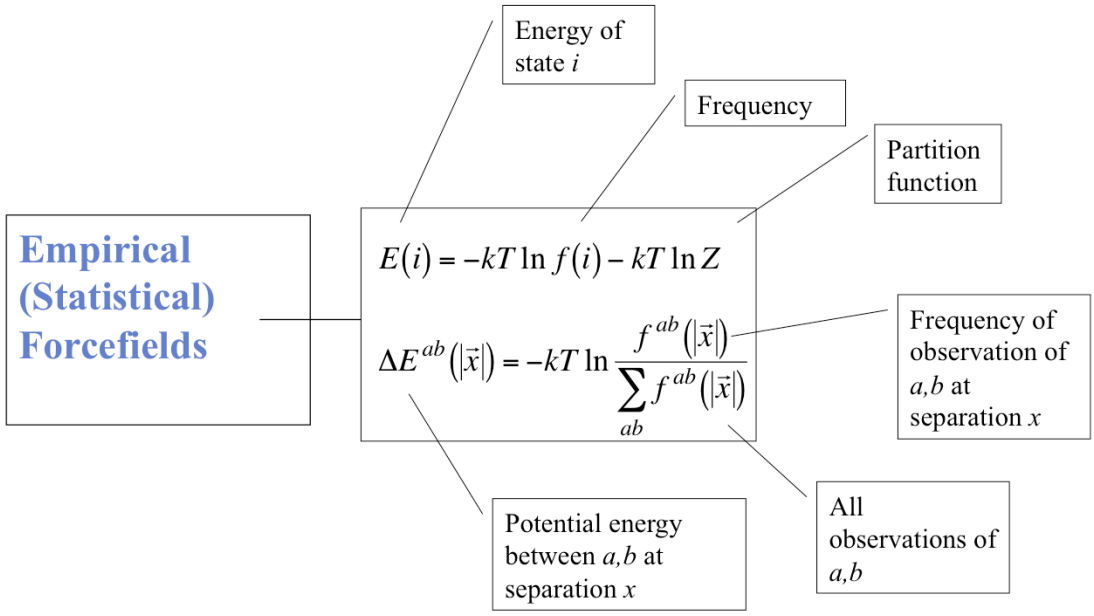
The core of the 1BM8 structure, with a 6.0Å diameter circle.

Carbon-Carbon interaction energy as a function of distance for two atom centers, modelled as a Lennard-Jones potential function.



6Å diameter circle drawn in the core of 1BM8 to illustrate the number of atoms that contribute to the Vander Waals forces on the atom in the centre.

Forcefield: functional Form



Empirical free energy functions are based on Boltzmann's formula: they relate the frequency of observed states to the free energy difference between the states. For example, think of a "state" as two amino acids being observed at a particular distance in protein structures in the PDB. From comparing how much more often the two amino acids are found in close proximity (or contact) than all other pairs of amino acids, one can deduce the free energy of their interaction. This can be converted into a potential.

The big advantage is that temperature, solvation, multi-body effects and entropy are all *implicitly* being accounted for, to the degree that they influence the distribution of distances in actual proteins.

In order to get make these statistical potential accurate, a very large number of structures needs to be considered. Remember, for amino acids – we need data for $20^2 = 400$ possible combinations, some of which are intrinsicly rare, over many resolved distances. You might think that with over 120,000 structures in the PDB that would not be a problem. However, the actual, usable number is much smaller since we need to remove bias from oversampling homologous structures, and the number of distinct folds in the PDB is only on the order of 1,300 – a number that that has not increased by much over the last decade (we may have caught them all by now).

Just as in molecular mechanics potentials, multi-body interactions are not considered, only pairwise interactions.

In practice, current potential functions often combine molecular mechanics for local detail and statistical potentials for solvent and long-range interactions.

http://steipe.biochemistry.utoronto.ca/abc

B O R I S . S T E I P E @ U T O R O N T O . C A

DEPARTMENT OF BIOCHEMISTRY & DEPARTMENT OF MOLECULAR GENETICS
UNIVERSITY OF TORONTO, CANADA