Computer Aided Drug Design
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture - 16
Molecular Mechanics / Force Field

Hello everyone, welcome to the course on computer aided drug design. We will talk more on concept of molecular mechanics and force field. In the previous class I introduced these concepts to you. There are 2 different major computational methods for calculating structures, 3 dimensional structures as well as some properties, okay, one is called the quantum mechanics approach.

So in the quantum mechanics the nucleus and electrons in the molecules are separately constituted, that means you look at electrons separately and nuclei separately. There are 2 approaches here one is called the ab initio method, here, the other one is called the semi-empirical method. Okay, the ab initio method is very, very rigorous, it requires lot of calculations.

There are no stored parameters, it takes a long time, it can be used only for small molecules, okay, 10s of atoms. The other one is the semi-empirical method, it is faster so lot of parameterization has been done, so it calculates certain features using some regression relationship, it is less accurate it can be used for large molecules also, that means 100s of atoms, okay.

These quantum mechanics can be used if you want to calculate molecular orbital energy, if you want to calculate partial charges, if you want to calculate electrostatic potentials, dipole moments, if you want to calculate what is the energy required for breaking bonds, energy required for forming bonds and so on actually. The other method is called the molecular mechanics method, okay, here this is the simpler version of quantum mechanics.

(Refer Slide Time: 02:04)

Two major computational methods for the Calculation of Structure and Property

1. Quantum mechanics

  ✓ Nuclei and electrons of the molecules are seperately considered
  ✓ Two approaches
        Ab initio - more rigorous, from first principles (no stored parameters or
              data), takes a long time, restricted to small molecules
        Semi-empirical - faster, but less accurate, can be used on larger molecules
  ✓ Useful for MO energies, partial charges, electrostatic potentials, dipole moments

2. Molecular mechanics

  ✓ Nuclei and electrons are lumped into atom-like particles.
  ✓ Interactions are based on springs and classical potentials.
  ✓ calculate structure and dynamics, total energies, entropies, free energies, and
    diffusion

Here you do not differentiate between the nucleus and electrons, okay, so it is lumped together as atom like particles, okay, it is like balls, then these balls are connected by springs okay, they are all connected by springs, so you use classical potentials like Hooke's law for calculating energy, okay, so we can calculate structure, dynamics, total energy, entropy, free energy and diffusion.

So we can go up to 1000s of atoms in molecules, so if you are doing docking of various ligands to a protein then we always use these molecular mechanics method, okay, so it is widely used for docking calculation, it is widely used for molecular dynamic studies, so most of computer aided drug design uses molecular mechanics and little bit of semi-empirical method also is used for calculating partial charges, electrostatic potential and so on actually.

(Refer Slide Time: 03:14)

Components of a Force Field

Energy =
    Stretching Energy +
    Bending Energy +
    Torsion Energy +
    Non-Bonded Interaction Energy

Okay, so we will go more into the molecular mechanics initially, so the energy is made up of 4 different components, the stretching energy, okay, stretching energy, then the bending energy, the torsion energy, the non-bonded interaction okay, so the stretching.

(Refer Slide Time: 03:26)
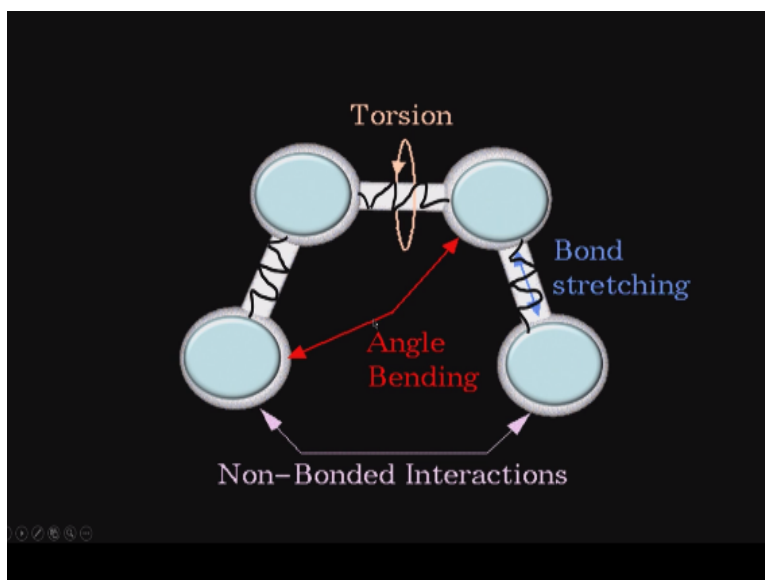
So if you look at particular molecule assume it has got 4 atoms connected, or 4 balls connected in 3 bond formation, okay, so this is called the bond stretching, okay, 2 atoms are stretched or conversely maybe pushed in or compressed, that is called the bond stretching, then we have the angle bending, suppose you have 3 atoms okay, then the 2 bonds can bend, inwards or outwards, the third is the torsion energy, okay.

This is like a twist in the third dimension, that is the torsion energy, the 4th one is called the non-bonded interactions, that means look at these 2 atoms, they are not connected, but there could be some interactions between these 2 because of electrostatic forces, van der Waals forces, or even hydrogen bond formation, if the atoms are favorable, so they are called non-bonded interaction.

So you have energy terms for each one of them, you add up and then you get the total energy of this particular molecule, now let us look at each one of them independently.

(Refer Slide Time: 04:39)
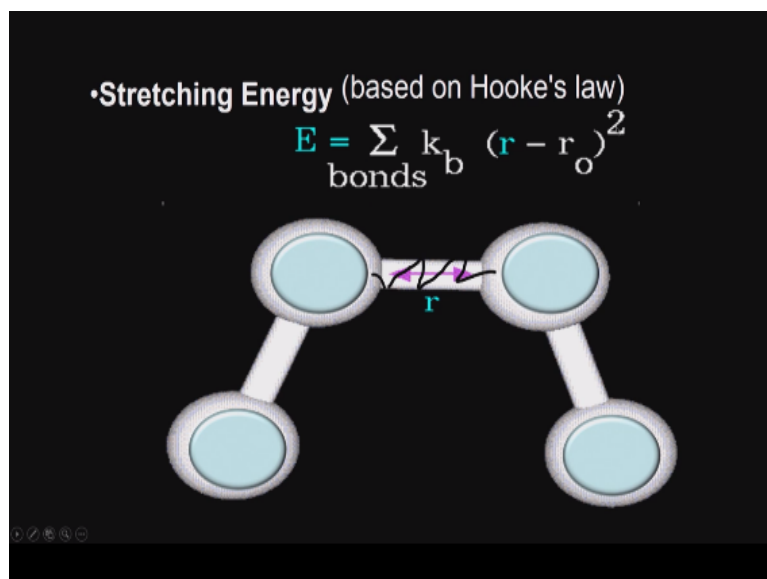
Stretching energy, E = summation of kb r - r0 whole square so r is the distance here between these 2 atoms, r0 could be the equilibrium, r – r0 is the distance you have stretched, or how much it has been elongated or even compressed. Hooke's law as you know is delta x square, that is why you have a square term coming in, so kb is a constant. You are doing summation because there could be many bonds.

So there could be many stretching happening. So that all adds up to this energy, so we have parameters kb, we have r0. We need to know them. So depending upon the type of atom this one and this one, the r0 will change, kb will change, okay, it could be carbon-carbon and carbon-nitrogen and carbon-oxygen and so on and between carbon-carbon also there could be single bond sp3, double bond sp2, triple bond sp or aromatic and so on actually.

So you need to have that data for the kb as well as for r0 for various systems which is stored, okay, they are called parameter, so we need lot of parameters, so depending upon the type of software the number of parameters may vary and you may have large number of parameters in some software and some software may have less number, so between software you will always get a difference in the energy calculation.

Okay, so if you look at the graph because it is r – r0 square it will appear like a parabola, so you stretch, the energy will increase.

(Refer Slide Time: 06:15)
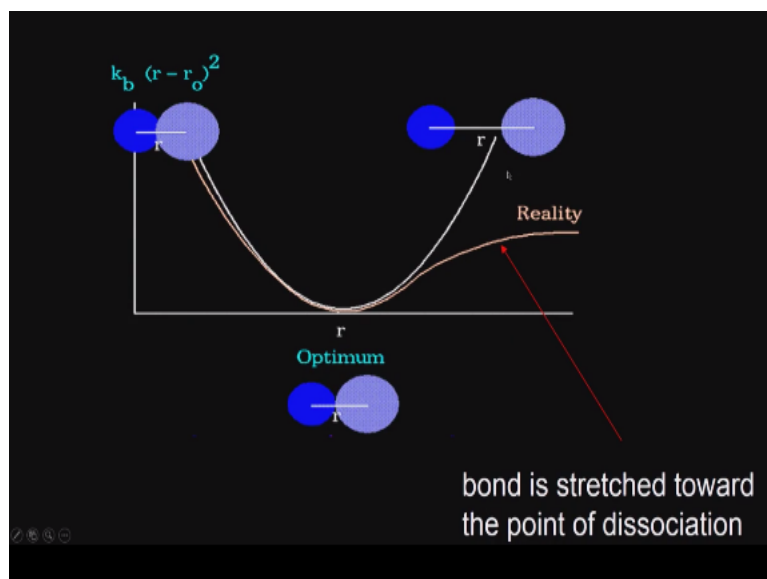
You pull them together or push them together, energy will increase, so this is the optimum, okay, optimum distance, okay, if you push them inwards energy will increase, if you pull them out energy will increase of course you cannot pull beyond certain distance as you know the bond can break, bond is stretched towards the point of dissociation, okay, so this is the optimum.

And ideally you would like to have that sort of optimum value, okay now let us go to bond stretch again, generally we use a quadratic term as I showed you, but the energy can be very large for very elongated bonds, for example as you can see here the energy goes up like a parabola, so the energy numbers go up even if there is a small change in the $r - r0$, so quadratic has the problem okay, especially for elongated balls.

There is something called Morse potential which is more accurate, the term for that is like this, E stretching = summation of D 1 – exponent – beta r – r0 whole square okay, so now this has got 3 different parameters or constants, but this is more accurate than this, especially for elongated bonds because the energy values will look very high because this is a quadratic term, so many software try to use this rather than (()) (07:50).

Cubic is not used because it gives you a wrong asymptomatic form, but sometimes quartic is also used, that is r – r0 or 4, okay, so you may have r – r0 square, r – r0 cube, r – r0 power 4, so you have many, many, many constants coming here so if you look here a quadratic you will have only 2 constants, if you go to a quartic you are going to have 1, 2, 3, 4 constants, okay, it also fits very well.

Morse potential if you go that is also quite accurate here you have 3 constants, okay, so you have to note that r0 is not the same as the equilibrium bond length because of non-bonded contribution because there could be attractions because of electrostatic when you consider a molecule with many, many atoms.

(Refer Slide Time: 08:35)

Bond Stretch Term

- force fields use just a quadratic term, but the energy is too large for very elongated bonds

$$E_{str} = \Sigma \, k_i \, (r - r_0)^2$$

- Morse potential is more accurate,

$$E_{str} = \Sigma \, D_e \, [1\text{-}exp(\text{-}\beta(r - r_0))]^2$$

- a cubic polynomial has wrong asymptotic form

- quartic polynomial is a good fit for bond length

$$E_{str} = \Sigma \, \{ \, k_i \, (r - r_0)^2 + k'_i \, (r - r_0)^3 + k''_i \, (r - r_0)^4 \, \}$$

- Note: The reference bond length, r_0, not the same as the equilibrium bond length, because of non-bonded contributions

So r0 will not be the same as the r0 when we have just CC type of bond or CO type of bond, but once you add many other atoms, the r0 changes because of the other interactions that is happening, okay, so if you look at it this is the quadratic, this is the 4 quadric and this is the Morse and this is the actual exact.

(Refer Slide Time: 09:10)

Comparison of Potential Energy Functions for Bond Stretch
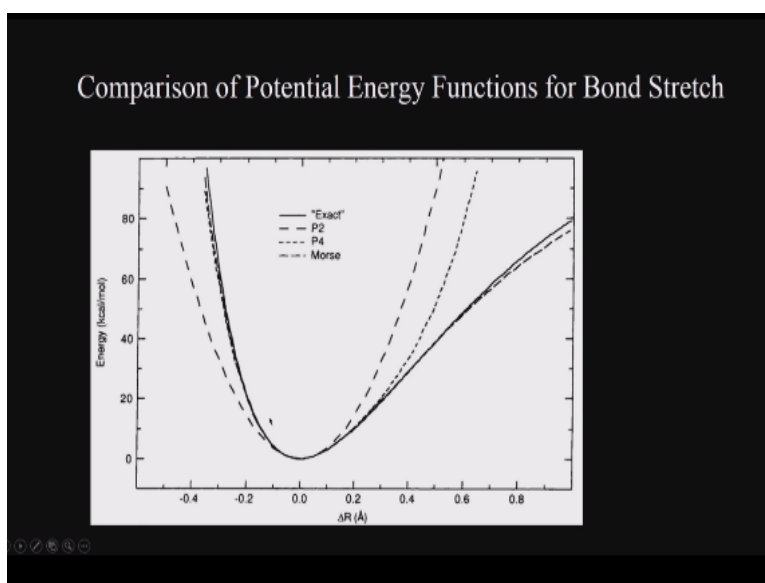
So as you can see many times Morse fits much better than quadratic or quadric. Let us go to bending energy now.
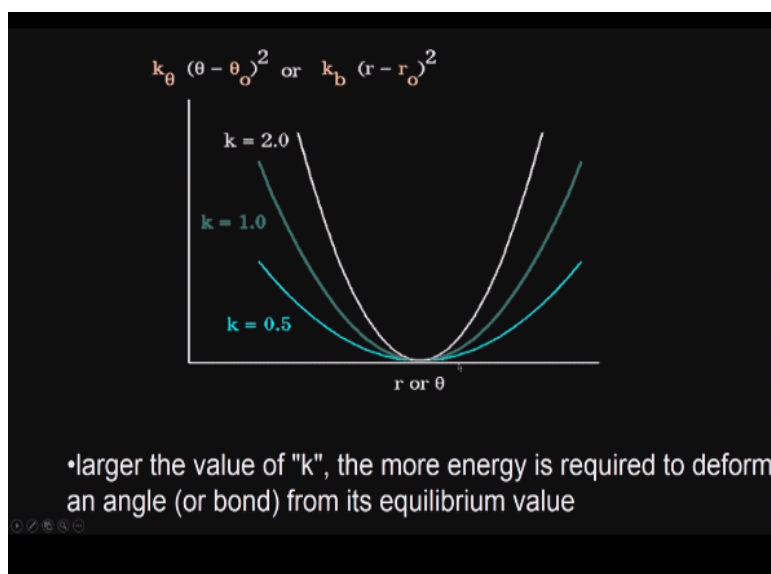
(Refer Slide Time: 09:24)
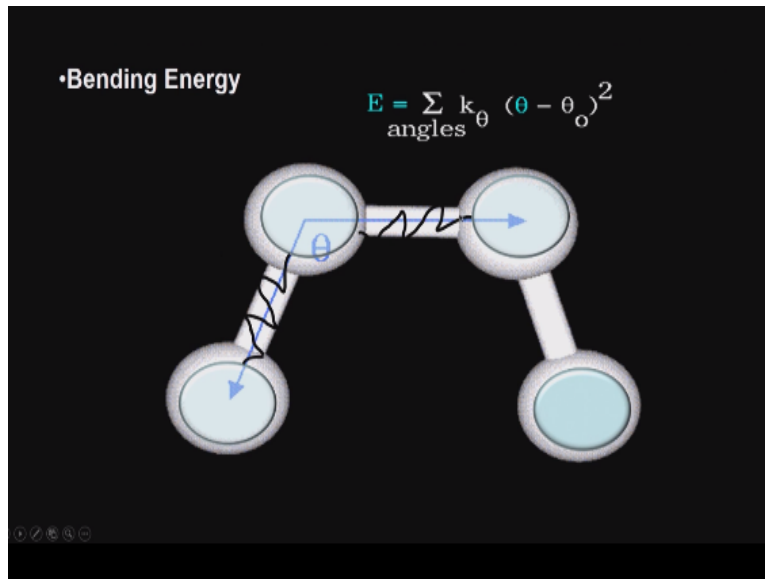
Bending energy, you have 3 atoms and there are 2 bonds so the angle could be less or it could be extended or expanding or compressing, again you go for a quadratic form like this E = summation of K theta, theta – theta not square. So you have 2 parameters, theta not and k theta and again you do a summation because there could be many bonds where there could be a bending happening.

(Refer Slide Time: 09:55)
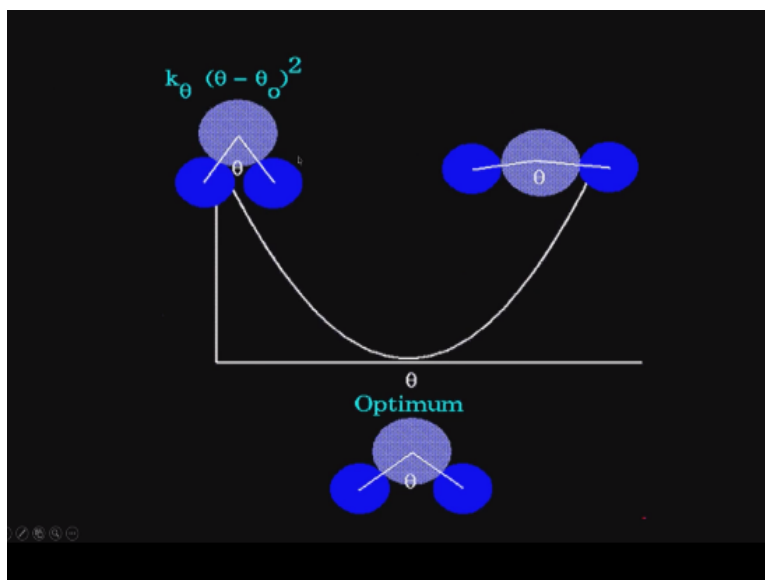
Okay, now again you are going to have a parabolic type of relation so if you expand the bond again it may be going up energy if you make it acute again the energy may be going out, so obtuse, acute both times energy will be going up in a parabolic form this is the optimum.

$$k_\theta \; (\theta - \theta_o)^2 \quad \text{or} \quad k_b \; (r - r_o)^2$$

k = 2.0

k = 1.0

k = 0.5

r or θ

•larger the value of "k", the more energy is required to deform an angle (or bond) from its equilibrium value

You could also change the k, so you may have shallow looking quadratic term like this or like this okay, so you can have different types of quadratic, larger the value of k, the more energy is required to deform an angle form it is equilibrium okay, because the energy becomes very high, whereas this type of shallow looking things you do not need very high energy to deform from it is theta not.

Angle Bend Term

• quadratic polynomial

$$E_{bend} = \Sigma \; k_i \; (\theta - \theta_0)^2_i$$

• for very strained systems (e.g. cyclopropane) a higher polynomial is better

$$E_{bend} = \Sigma \; k_i \; (\theta - \theta_0)^2 + k'_i \; (\theta - \theta_0)^3 + k''_i \; (\theta - \theta_0)^4 + \cdots$$

• special atom types may be used for very strained atoms

Angle bend term, again you can have a, this is the quadratic form, okay, you can have even go for quartic form, cyclopropane like, as you know cyclopropane is like this right, it is cyclic you have here CH2, CH2, CH2. It is a very strained system, so you may have to go for this, special atom types may be used for very strained atoms, this is a very strained system, remember that, okay.
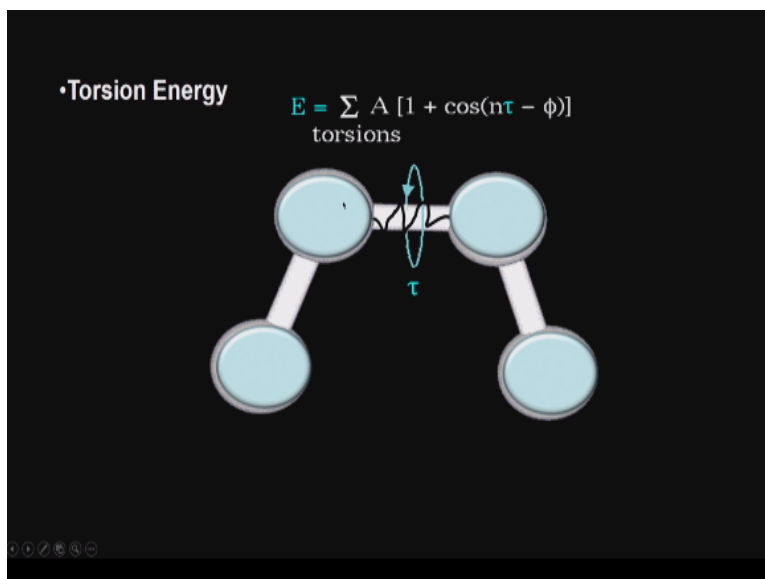
Then we go to torsion, torsion the expression for the energy looks slightly different, so you have torsion is the twist as we can say, it is the twist in the third dimension, so we have terms like A 1 + cos n tau – phi so we have constants these are the parameters so we will call it, okay, these are the parameters which need to be stored in your database. So torsion energy, for example if you take a butane like system, okay.
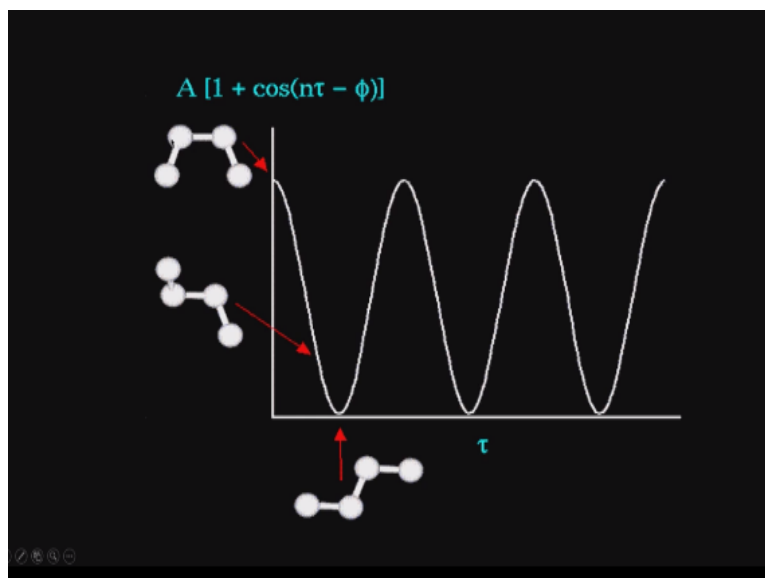
This is cis-butane like, so the energy is going to be very high, if this is a trans-butane energy is going to be lower, as we might have studied, okay, in between various configurations of these 4 we may have different energies, so as the tau increases you see this cyclic, but the energy numbers is very high for a cis, energy value is very low for a trans configuring.

Torsional Term

- single cosine with appropriate barrier multiplicity, n

$$E_{tors} = \Sigma\, V_i \cos[n(\theta - \theta_0)]$$

- some use a sum of cosines for 1-fold (dipole), 2-fold (conjugation) and 3-fold (steric) contributions

$$E_{tors} = \Sigma\, \{\, V_i \cos[(\theta - \theta_0)] + V'_i \cos[2(\theta - \theta_0)] + V''_i \cos[3(\theta - \theta_0)]\, \}$$

So you can have different (()) (12:16) forms of the equation, so we can have a single cosine with appropriate barrier multiplicity that is Vi and n, or we can even go for these type of relation cos theta – theta not, cos 2 theta – theta not, cos 3 theta – theta not, so if you have dipole type you can put 1 fold, 2 fold, conjugation, 3-fold steady contributions and so on actually okay.

So we can go down to 3 terms here or we can stick to one also, this is the called the torsion energy.
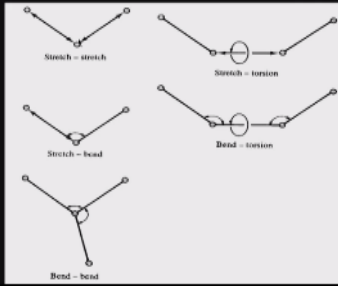
(Refer Slide Time: 12:48)

Cross Terms
interaction terms between stretch, bend and torsion

$$E_{str-str} = \Sigma\, k_{ij}\, (r_i - r_{i0})(r_j - r_{j0})$$
$$E_{str-bend} = \Sigma\, k_{ij}\, (r_i - r_{i0})(\theta_j - \theta_{j0})$$
$$E_{bend-bend} = \Sigma\, k_{ij}\, (\theta_i - \theta_{i0})(\theta_j - \theta_{j0})$$
$$E_{bend-bend-tors} = \Sigma\, V_{ij}\, (\theta_i - \theta_{i0})(\theta_j - \theta_{j0}) \cos[n(\theta_{ij} - \theta_{ij0})]$$

Now we can also have cross term imagine that just because you are stretching a particular bond, the other bonds and other angles remains constant, that is not true, there is always going to be some interaction, if I bend a system of course the bond length may change so that

is why we have call it cross terms, for example you can even stretch 2 atoms, this stretch maybe get disturbed for example if I stretch maybe this bend will get disturbed.

So there could be an interaction between stretch and bend, okay, if you bend one angle I am sure another angle may get disturbed, so there could be an interaction between bend and bend and so on, so you could have similarly when we stretch maybe the torsion gets disturbed, when I bend also torsion get disturbed. So, there is always interaction and the terms are not just independent of each other so some software consider cross terms.
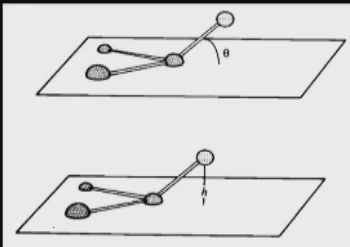
Interaction terms have been stretch, bend and torsion, for example stretch-stretch r i – r i not, r j – r j not, stretch bend, so you see that the terms come in here multiplying each other bend-bend, you can have bend-bend torsion, you can even have 3 term coming in to the picture, okay, some software consider the cross term depending upon the complexity you want to put in, realistic factors you want to put in, you can have more terms.

And of course more the terms I need to have more parameters known, for example this is the new parameter k_{ij} which talks about interaction between 2 different stretches, i and j, so I need to know that parameter value. In addition, computational time also will go up because you have extra terms into your energy calculations okay.

(Refer Slide Time: 14:47)

So out-of-plane bending, imagine you have a molecule with 4 atoms okay, 3 on the plane and one is up, that is called out of plane bending term, angle-to-plane or distance to plane can be used for the out-of-plane bending, so angle or distance that is called improper torsions for

out-of-plane bending, chirality constraints are required in united atom force fields. So we can bring in those aspects also.

Out of plane bending energy related to that so that could be based on either distance or that could be based on the angle.
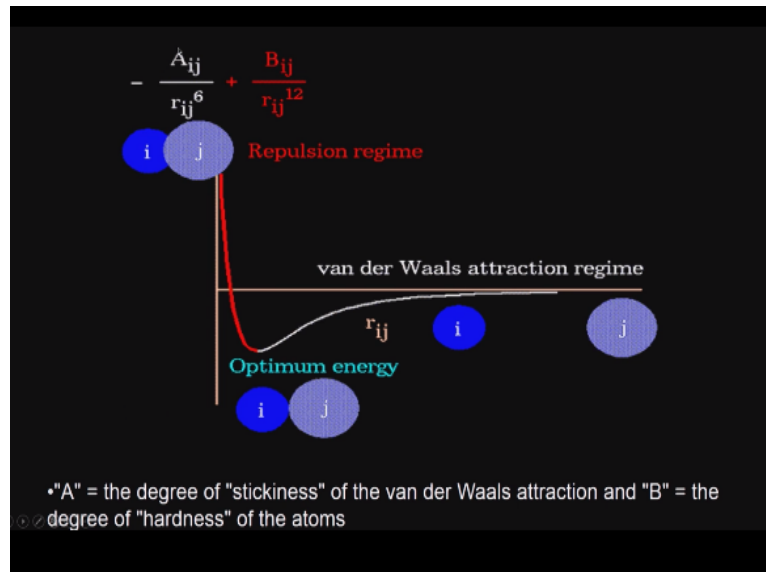
(Refer Slide Time: 15:20)
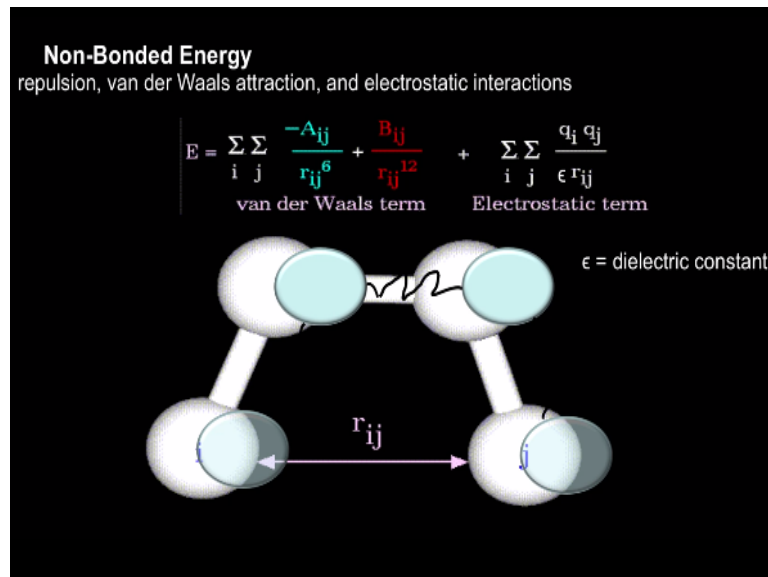
Then we have non-bonded interactions and the important ones van der Waals and electrostatic, okay, what is electrostatic. Imagine there is charge on this atom qi, there is charge on this atom qj, qi qj/distance rij that is the electrostatic term, okay, and van der Waals term looks like this aij rij raise to the power 6 + Bij rij raise to the power 12, okay, that is called the van der Waals term.

So here the denominator is rij whereas here you have 2 denominators rij 6, rij raise to the power 12, okay, so this is the van der Waals, okay this contributes negatively, this contributes positively towards the energy term, A and B are constant, okay, and there are variations to this, there are variations to this, we will look at some of them, okay, so we have the repulsion, we have the attraction here electrostatic interaction, these are non-bonded.

And there are many non-bonded interactions like hydrogen bond dipole-dipole and so on, we will look at some of them later, okay, so, how will the van der Waal look like, so if they are far away there is going to be attraction, if they come very, very close there is going to be repulsion, okay, that is why we have these 2 terms okay, far away 2 atoms there is going to be attraction and when they are closer there is going to be repulsion.

So the A is called the degree of stickiness and B is called the degree of hardness, okay B is called the degree of hardness, A is called the degree of stickiness.

(Refer Slide Time: 17:08)

Okay, you can modify by changing this term, we can have qi qj/D rij okay, D is the dielectric constant, so the D change is depending upon r, okay, so we bring in this extra term in the denominator for the electrostatics. Normally if we ignore that it will be qi qj/rij, but we bring in distance dependent dielectric constant here, so that means when they come closer you may have different dielectric constant when they are far away you can have different dielectric constant.

(Refer Slide Time: 17:50)

➢Partial charges are created due to the asymmetric distribution of electrons in chemical bonds. (Mulliken charge)

➢polar covalent bond like HCl, the shared electron oscillates between the bonded atoms

➢Chemical bonds between atoms within a molecule consist of one or more electron pairs distributed among the connected atoms.

➢Bonding electrons are not equally distributed among the atoms that have different electronegativity

https://docs.chemaxon.com/display/docs/Calculation+of+partial+charge+distribution

Okay, so what happens there are partial charges are created due to the asymmetric distribution of electrons in the chemical bonds right, so if you have there are electronegative atoms and electropositive atoms so there is always going to be some partial charges created okay, O means you may have negative slightly, they are all delta, delta negative and so on, so that is called Mulliken charge, okay.

Polar covalent bond like HCl, the shared electron oscillates between the bonded atoms. Whereas chemical bonds between atoms within molecule consists of one or more electron pairs distributed among the connected atoms and bonding electrons are not equally distributed atoms that have different electronegativity like I said O – you may have, okay, may be carbon may be different positive.

So small charge which leads to all these electrostatic forces okay, now let us look at benzene for example, if you look at benzene, so we have carbon, carbon, carbon, carbon so the H being positive you have a small delta positive charge so the carbon has to take in small negative charge because the whole molecule per se is neutral so these carbons will have slightly negative charge.

Of course in some cased carbons can have positive also, we will look at it. Look at this benzoic acid, we have C double bond O, OH, as you know H so it has got a positive charge, okay, this O you can see has got negative charge and this O has got a negative charge, this O has a higher negative charge than this O, see this, carbon has a positive charge here, whereas in the benzene you see carbons have negative charge, okay.

But then this carbon has positive charge, but all the carbons in the aromatic have negative charge, you can see and they are not same, they are all very different because of this benzoic acid, okay, they are symmetric here, but this COOH is closer, so these carbons have very low negative charge, okay, whereas these carbons have higher negative charge so it is different because of this particular COOH group here, okay.

Look at this NH2, okay, so we have benzene connected with the NH2 here so nitrogen – 0.156 and again all these carbons on the benzene ring and again as you can see it is not same unlike here because of the NH2 group okay, these carbons have different charges so these 2 carbons have same negative charge this carbon has less negative charge and this carbon has higher negative charge okay, between this and this if you see benzoic acid and aniline.

The carbon here are quite different as far as the charges are concerned okay, you can see lot of difference, that is why the reactivity changes, that is why the electrostatic values change and so on actually, okay there are software for example this particular software it can calculate the charge distribution if I put in a molecule okay. We will look at it at some point but this particular software called ChemAxon, it can give you the charge distribution shared.

(Refer Slide Time: 21:27)
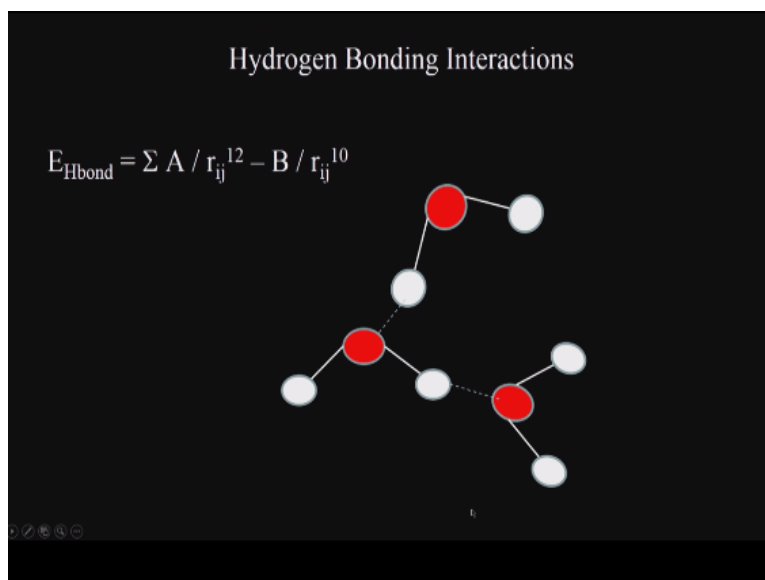
Okay, hydrogen bond that is the very, very important term because many ligand protein interactions are based on hydrogen bond, they have a different form of equation, okay if you look at van der Waal, we have r power 6 and r power 12, okay, here you see r power 12 and r

power 10, okay, A/rij 12 what is this hydrogen bond, so this oxygen here becomes hydrogen bond acceptor, okay.

This hydrogen here it is hydrogen bond donor here, okay, so we can have all the 3 may be involved, okay, these 2 maybe hydrogen bond donors, this maybe hydrogen bond acceptors and you can have a very ne2rk of hydrogen bond formation in water molecule, okay, so they have different terms as you can see with respect to van der Waal.

(Refer Slide Time: 22:26)

So non-bonded interactions can be made up of van der Waals, electrostatic and hydrogen bonding. The repulsive part of van der Waals potential due to overlap of electron distribution, it is called Pauli's exclusion principle, it rises very steeply because of steric repulsion, okay, attractive part of van der Waals potential due to London or dispersion forces, it is dipole induced dipole interaction.

It is proportional to r power 6, okay, as I showed you here it is proportional to r power 6, okay, that is called the London or dispersion forces, okay.

(Refer Slide Time: 23:09)

Non-Bonded Interaction
- Lennard-Jones potential
  - $E_{vdW} = \Sigma\, 4\,\varepsilon_{ij}\, (\,(\sigma_{ij}/r_{ij})^{12} - (\sigma_{ij}/r_{ij})^{6}\,)$
  - easy to compute, but r^{-12} rises too rapidly
- Buckingham potential
  - $E_{vdW} = \Sigma\, A\,\exp(-B\,r_{ij}) - C\,r_{ij}^{-6}$
  - QM suggests exponential repulsion better, but is harder to compute
- tabulate σ and ε for each atom
  - obtain mixed terms as arithmetic and geometric means
  - $\sigma_{AB} = (\sigma_{AA} + \sigma_{BB})/2;\quad \varepsilon_{AB} = (\varepsilon_{AA}\,\varepsilon_{BB})^{1/2}$

So different types for forms for non-bonded interactions is called the Lennard-Jones potential, we have summation of 4 epsilon sigma rij raise to the power 12 – sigma rij raise to the power 6, so r – 12 rises 2 rapidly, it is easy to compute, we have another one called Buckingham potential A exponent – B rij – C rij – 6, but it is slightly harder to compute we have to calculate this.

Or we can tabulate sigma and epsilon for each atom obtain mixed terms as arithmetic and geometric means like this okay, sigma = sigma + sigma AA BB/2 so sigma AB could be geometric means or it can be even arithmetic means for sigma and geometric means for epsilon, okay. So different approaches by which we can calculate these terms for AB or ij.

(Refer Slide Time: 24:15)

Interaction between Drug and Receptor

•Covalent bond : ΔG_o = -40 to –110 kcals/mol

•Ionic or electrostatic interactions: -5 to –10 kcals/mol

•Ion-Dipole or Dipole-Dipole interaction: -1 to – 7 kcals/mol

•Hydrogen bond: -3 to –5 kcals/mol

•Charge-Transfer complex:

•Hydrophobic interaction

So when you have a drug and it is going and binding to a receptor that is your protein and there are lot of bonded interactions, non-bond interactions, for example your covalent bond is bonded there is a strong bond, you can look at the energies 40 to 110 kilocals/mol whereas ionic or electrostatic interactions, these are all non-bonded, it is only -5 to -10 kilocals/mol, ion dipole or dipole-dipole -1 to -7.

Hydrogen bond -3 to -5 kilocals/mol, so you see the term they are all quite small when compared to the bonded, all these non-bonded interaction or energies are very, very small when compared to this, but in ligand protein binding only these play a very important role because there is no bond formed, there are only non-bonded interactions, electrostatic, van der Waals, bi pi and charge salt bridge and so on actually.

(Refer Slide Time: 25:25)
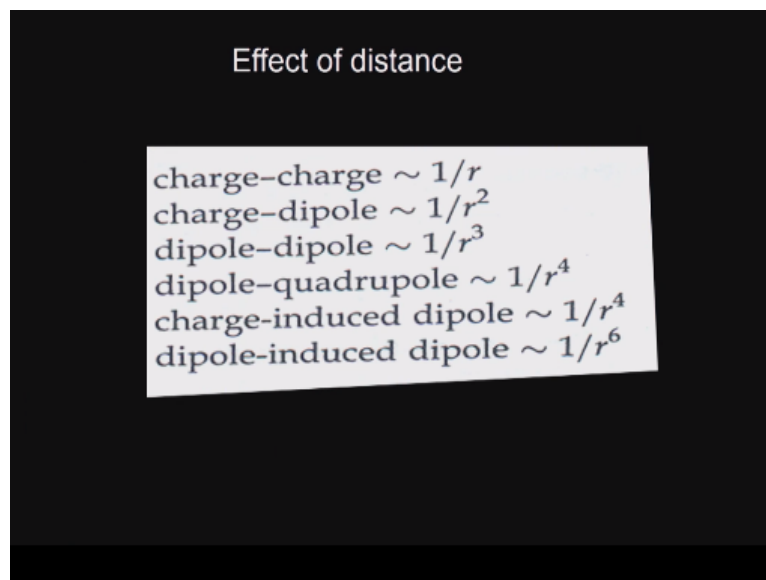
If you look at the effect of distance, as you know electrostatic we have qi qj/rij so it is proportional to 1/r, okay whereas if you look at charge dipole type of interaction it is related to 1/r2 okay, dipole-dipole it is related to 1/r3, dipole-quadrupole 1/r power 4, charge-induced dipole 1/r power 4, dipole-induced dipole 1/r power 6. So we call this as the long term interaction, because all these other terms will rapidly decrease as the distance increases.

Whereas this decreases only linearly, whereas this maybe decreasing, all these decreased nonlinearly, quadratically raise to the power 4, raise to the power 6 decrease, so they decrease very rapidly whereas this decreases only linearly so we call these as the long term interaction, so in the long distance when the ligand and protein are there initially we start having only electrostatic forces.

But as I come closer and closer all other forces start taking place and they become important in judging how the ligand binds into the active site what type of conformation it takes and so on actually, okay, so we will continue more.

(Refer Slide Time: 26:55)

There are many parameters we saw right, so many different parameters, so many different parameters, constants, k values in the multiplication, r0, theta not so many parameters, okay so all these force field or molecular mechanics software need to have a database of parameters, so where do they get that from, they get from literature, experimental data, sometimes they do very exact calculations that is called ab initio calculations.

So much of the data is from small molecule so I collect data from small molecules, but when I go to a larger molecule I assume those parameters do not change, do you understand, so if I have a small molecule which has got a CO bond, I get the parameters for that, but when I go to a large molecule and it has a CO bond, I assume that parameters do not change, that is a very important assumption I make, okay.
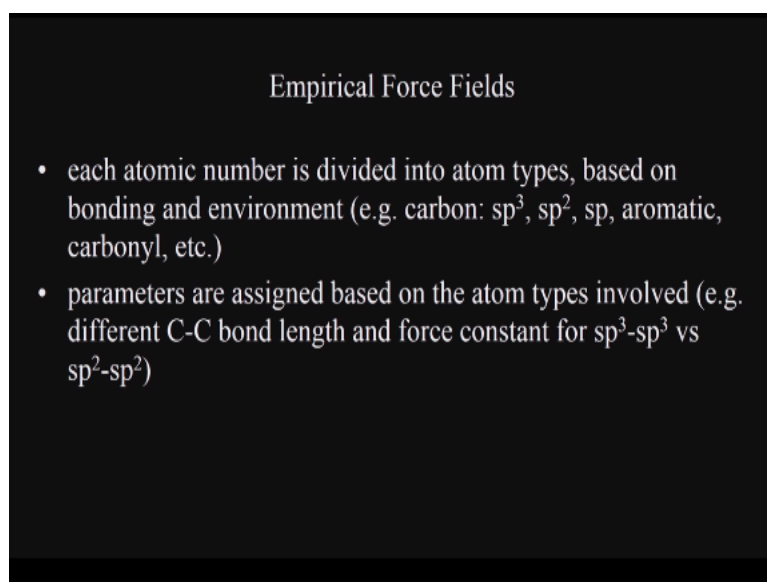
So parameters for fragments of proteins are appropriate for that fragment in different context also, okay, in different context also that is a very important point we need to, so whether it is in a small molecule or in a large molecule we assume the same parameters that is very important parameter approximation, so much of the data is from small molecule. Okay, we cannot get all these parameters from a big peptide.

But using small molecule like 200, 250, 300 molecular weight we get from crystal structure because crystal structures when you crystalize the organic molecule we can get the length, angle, non-bonded coefficients so we can run some molecular mechanics calculations and then fit the findings to the experimental data and get the parameters, we can get from vibrational spectroscopy, okay, some of the vibrational spect Raman or IR.

We can also do ab initio quantum mechanics calculation, colorimetric studies, we can find out heat of the formation, we can get delta G okay, using some colorimeter, we can get from, so from thermodynamic measurements we can get some, so all these experimental data can be used to get many of the parameter and one important assumption is small molecule parameters obtained from small molecule can be extrapolated even if it is a fragment in a large molecule also.

Sometimes they are also got from ab initio quantum mechanics calculation, as I said ab initio is a very accurate study and it perform such very accurate experimental studies using very high and quantum mechanics ab initio calculation, okay, so this is how the parameters are got actually, there are many empirical force field, okay.

(Refer Slide Time: 29:57)

Lot number of force field software are there, okay, we will look at some of them and their form, but all of them will have basically the terms connected to bond stretching, bond angle, torsion, non-bonded interaction and sometimes cross terms okay, and some software will have lot of parameters, some software will have less parameters, some software are used predominantly for organic molecules.

Some can be sued for inorganic and so on actually and some may even have metals, so parameters for metals, iron for example, zinc, copper and so on actually, okay, so we will look at all these more in detail how the parameters vary between software, how the equations look like between software and so on actually, okay. Thank you very much for your time.