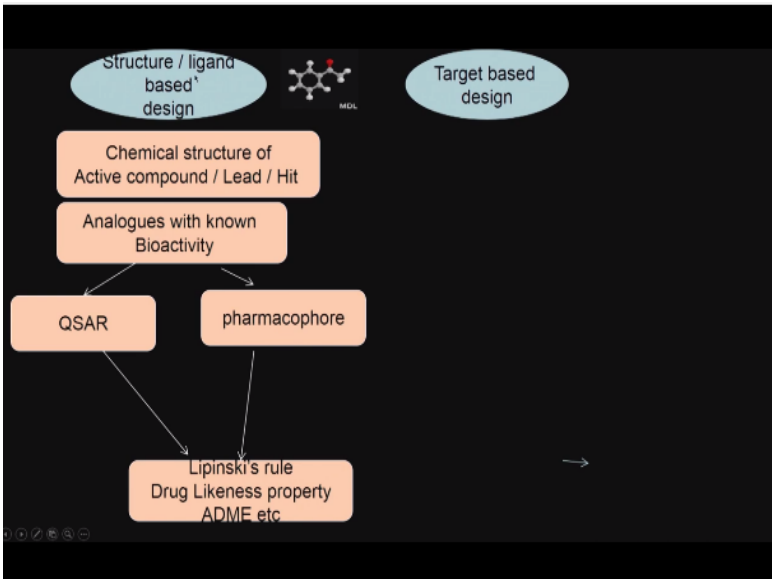
**Computer Aided Drug Design**
**Prof. Mukesh Doble**
**Department of Biotechnology**
**Indian Institute of Technology – Madras**

**Lecture - 15**
**Molecular Modelling**

Hello everyone, welcome to the course on computer aided drug design, today we are going to start a new topic that is called molecular modelling. Molecular modelling involves modelling the 3 dimensional structure of the molecule because ultimately the molecule has a particular confirmation when it goes and binds to the protein, so we need to understand how to model that and how molecules attain different types of confirmation based on the energies, okay.

So there are 2 approaches by which we can do the drug design, one is called the structure or ligand based design.

**(Refer Slide Time: 00:51)**




That means based on the drug structure, based on the ligand structure, okay, the other is called the target based design that means, I know target protein where the drug will go and bind, so I will design molecules based on the target structure, so this is based on the drug structure that means I do not know any knowledge, any information about my target whereas this is based on the target 3-dimensional active site okay, so 2 different approaches.

So this is called structure ligand based design. In ancient days there was no knowledge about getting the 3-dimensional structure of your target because x-ray, crystallography, proteomics

was not there, so generally the tested different compounds and they found some compounds to be very active then they started designing compounds with similar structure, okay, the 3-dimensional structure of the protein was not known.

So it was predominantly the ligand based approach or structure based approach, but off late with the evolving research and technology development in the area of proteomics, x-ray crystallography, LCMS, 2D gel and so on, the target based approach became very very fashionable or important because I will know which target protein or enzyme I am going to inhibit by designing a drug, so all the drugs started being designed based on this.

You can also combine both, we can do some screening using the ligand based approach and then we shift to the target based approach, okay, in the ligand based approach, so I know some structures, I know the chemical structures of some active compounds, we call it leads or hits and then so I design synthesize new analogs with known bioactivity, okay.

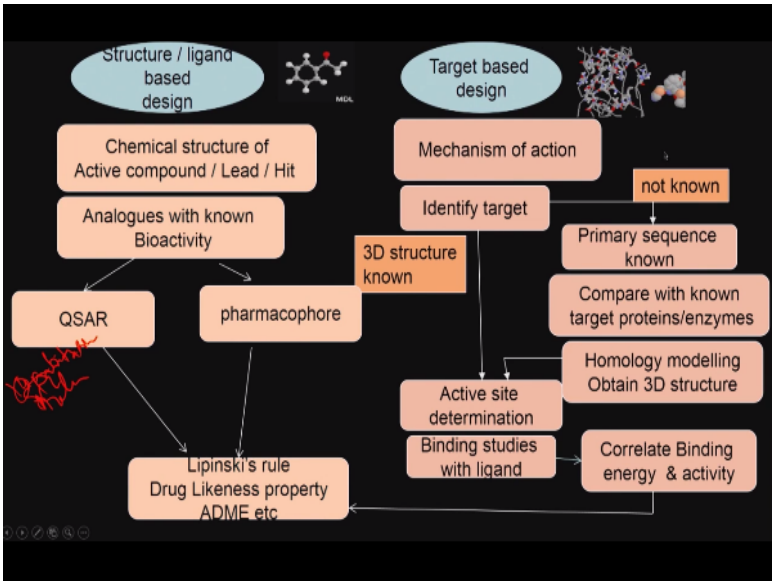
So and then you do something called quantitative structure activity relationship, that means what are the structural features required for the activity, or I do something called a pharmacophore modelling that means I know what are the functional groups responsible for certain activity so I try to retain all these, okay, so, I design new molecules by retaining certain important functional properties, I design structure activity relationship.

And then of course later on, I use drug likeness property, Lipinski's rule, ADME property, all these to shortlist molecule, okay, this is called the ligand based approach, okay, so in the past many classes we did look at some of these things right, how to screen based on Lipinski's, how to screen based on drug likeness and then how do you identify the structural similarities, but we are going to spend lot of time later on QSAR.

That is quantitative structure activity relationship, the other is pharmacophore based approach, so this is called the ligand based design, okay, as you can see here, I do not know anything about the target protein into which the particular ligand, okay, which you have got here is going to bind to, I have absolutely no knowledge about it, now the other approach is called the target based approach.

That means I have some knowledge about the 3-dimensional structure of the target, so I designed molecules so that they go and bind very effectively into the target, forms good interactions with amino acids present and so on, so that is called the target based approach, okay, so what does that involve, it involves many things.

**(Refer Slide Time: 04:51)**




First I need to know the mechanism of action, that means, I need to know how the drug goes and acts, what are the various pathways, what are the various enzymes involved and all, then from there I identify a target, so I will say, there might be many enzymes, but I many decide on only one target, okay, that is identifying the target, now I know the 3D structure of the target, or I do not know the 3D structure of the target.

For example, if you take inflammation there are enzymes like cyclooxygenase 2 is to be inhibited to prevent certain inflammation. You may have the 3 dimensional structure of cyclooxygenase 2 for see some mammals but not for human, okay, so that sort of situation can happen, so I do not have the details for human, but I may have for some animals.

So if I have details 3-dimensional structures of these enzyme, 3-dimensional structure is known, there is database called PDB, protein data bank, it contains 3-dimensional structure of large number of proteins, so you go to that particular database and see whether the protein of your interest is there. So the 3 dimensional structure is known, what do we do, life if very simple.

So I have to identify the active site, like if you see in this picture the drug goes and bind to the active site, so I should know something about the active site, okay, either from literature, I will know the active site or when they crystalize and put the 3 dimensional structure of the protein they may have already some component site or ligand inside during the crystallization, so I will consider that to be my active site, okay.

Otherwise it is a complex process, either I go to literature and find out or I look at various sites in the protein to determine what should be the active site, okay, imagine, if the active site is known what do I do is I bind the ligand to particular protein and I see what is the binding energy, okay, that is called the binding energy, BE, so based on the binding energy I will say the interaction is very good or the interaction is very poor.

Okay, so based on 3 dimensional structure, based on the active site, I will bind the ligand into the protein and calculate the binding energy and I will say if the binding energy is highly negative, then it is very good, if the binding energy is poor, I will say the binding is not excellent, this route if I know the 3-dimensional structure of the protein of interest.

Suppose I do not have idea about 3 dimensional structure of the protein, for example, I do not have the 3 dimensional structure of the cyclooxygenase 2 enzyme for human, but I may have for mouse or rat, okay, then I perform something called the homology modelling, okay, this is very very important, I perform, so what I will do is I will have the primary sequence data of the protein, then I will compare it with other proteins in the PDB.

And if there is some similarity I take that protein and I develop a homology model using that protein as my template, okay, that means I build the 3-dimensional structure of the protein of my interest based on the 3 dimensional structure of another protein, which is similar in the sequence, okay, 30%, 35% similarity in the sequence, once I do that I come here, I will look at the active site.

I decide on the active site like this picture and then I will bind my ligands to the protein, I find out what is the binding energy, so whether the binding energy is very good or the binding energy is poor. If the binding energy is very good, I will say the ligand may have high activity and vice versa, okay, this is how you do a target based design, okay, these are the 2 different approaches.

One is called the structure and ligand based design, another is called the target based design, okay, the structure based or ligand based design, I have no idea about my target, so I will look at molecules, which I have shown good activity towards a particular disease then I will develop analogs okay, known analogs, I may synthesise them, I may test their in vitro activity.

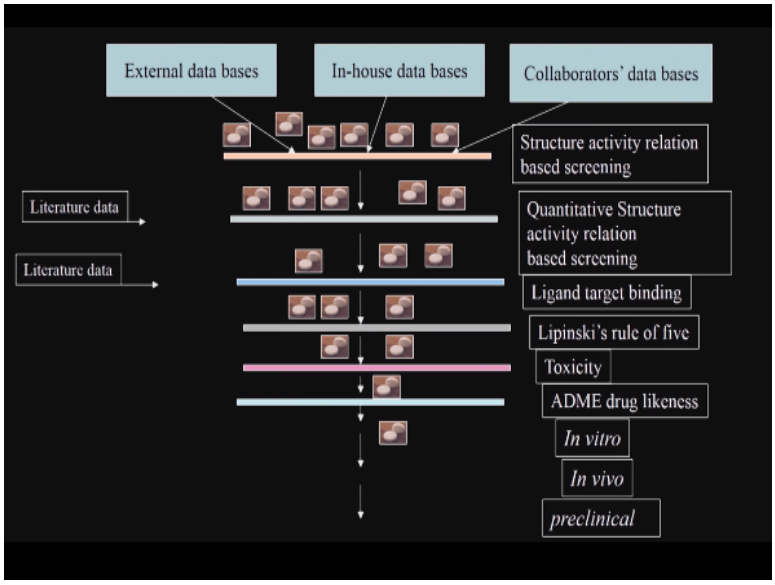
I may develop a quantitative structure activity relationship or I may even develop a pharmacophore model, okay, and that is how I come up with new structures and then I will of course use Lipinski's rule, drug likeness rules, ADME rules which we have seen in the previous many classes and shortlist possible compounds, so that is how the structure based works.

Whereas in the target based either I know the 3 dimensional structure of my target as here whereas I do not know the 3-dimensional structure of the target like here. If I know the 3-dimensional structure of the target, if I know the active site, I bind many ligands to it, and calculate the binding energy and then I will correlate binding energy versus activity, so if the binding is very good.

I will say it will have good activity, if the binding is poor I will say it will have poor activity, so if the 3-dimensional structure of the protein of my interest is not there, but similar protein structure are available like I said, for human I might not have the 3D structure of that protein, but I may have the 3D structure analogues to a rat or a mouse then what I do is, I do a homology model.

That means, I will have built the 3D of protein of my interest using the template of that protein, okay, that is called homology model. Once I do that then again this is called a modelled protein, then I will bind my molecules and then see how the binding energy correlates with the activity. We can combine both these structure based approach and target based approach also for coming up with you possible candidates. I will show you how it is in the next slide.

**(Refer Slide Time: 11:53)**




Okay, so here this particular picture shows how we can combine lot of in silico, in vitro information, okay, so imaging I have lot of structures based on external databases like zinc and drug DB, I showed you so many different data inhouse database, as a pharma company I have structures of maybe millions of compounds. I will maybe get on payment basis databases from other collaborators or vendors.

So I will start looking at all these millions of molecules okay, what do I do I may develop a structure activity relation based screening so I will look at what are the structural features that gives you activity, so I will neglect compounds which does not have those structural features okay, I may take less number of compounds, after that if I know the target then I will bind all the molecules to the target.

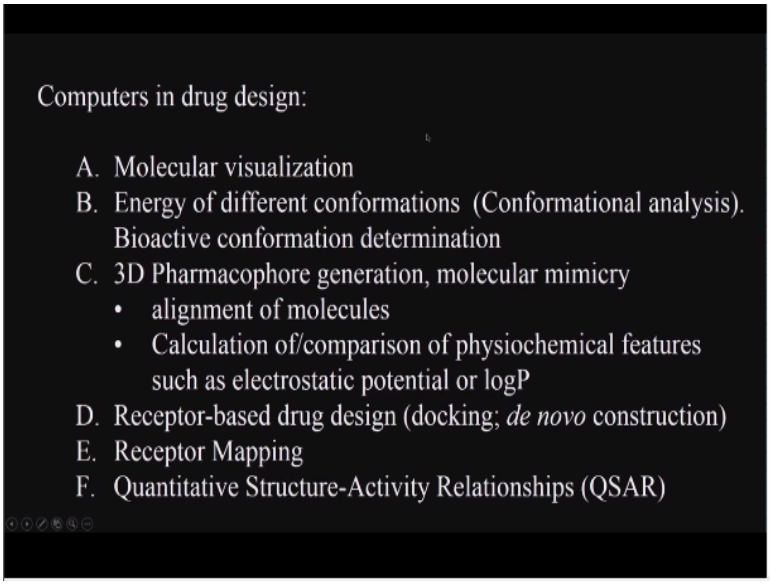
Okay and that is the target binding and see which molecules bind well, which does not bind, so I will take those which binds very well, then I will apply Lipinski's rule of 5, I will apply toxicity rules, I will apply ADME drug likeness rules, and then keep removing compounds which do not seem to satisfy and then finally come up with maybe 100s of molecules.

So I may start with billions then with structure activity based relationship I eliminate some, by binding I eliminate some more then I apply the Lipinski's rule, I will apply toxicity rule, I will apply ADME rules, drug likeness rules, BBB, PGB, HERG, biotransformation stability, protein plasma binding, so many different rules and then I come up with a small number maybe 100s.

I can do a high throughput screening, I come up with the best candidates Hits we will call it. Then I may go to animal models, then I may go to human trials, so this is how we can combine in silico techniques with in vitro, in vivo preclinical study, okay, so we need to have lot of literature data initially, we will need to have lot of databases initially to start screening.

So when I use in silico methods I can use millions of molecules because I just need a computational resource to do this job whereas real experimental may start from here, okay, much lower, whereas these are computational tools so we can start with millions of molecules and apply different rules, different techniques to shortlist maybe 1000s or 100s okay, so computers are extremely useful in this area for molecular visualization.

**(Refer Slide Time: 14:50)**



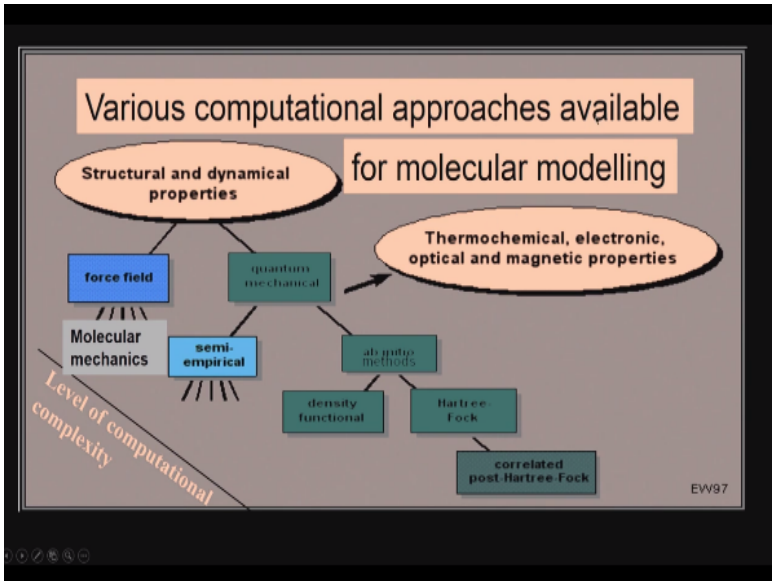

We can look at molecules, draw the structures, how they look like, we can look at how the energy of the molecule changes based on their conformation, okay, we can determine the minimum energy conformation as we call it, that is called the minimum energy conformation and then what is the bioactive conformation? That means the conformation which gives you the bioactive.

We can look at the 3-dimensional pharmacophore generation, we can get the 3D structure of the molecules and then we can say these functional groups are important pharmacophores then we can do molecular mimicry, we can align molecules suppose if there are 20 molecules which I have shown good activity we can align them together and then we can see what is the similarity.

We can compare their physicochemical features such as electrostatic potential polar surface area and so on. And then we can do docking like I said, dock the drug to the enzyme receptor and see how they dock, we can look at the receptor mapping then we can do quantitative structure activity relationship studies that means what are the structural features that gives you activity, can I get a mathematical relation between the structural features.

And activity that sort of studies, we are going to talk more about all these in detail. Okay, so many things can be done using these type of techniques, okay, so many things can be done using computational tools, okay, so we are going to look at all of them in the forthcoming classes, so there are different computational approaches for molecular modeling.

**(Refer Slide Time: 16:40)**




One is called the force field based approach, the other one is called the quantum mechanics approach. Force field based approach is the easiest one, it does not require too much computer resources, complexities, but we can get only some set of parameters using force field based approach we can get the structural properties that means 3 dimensional structure. We can get the shape of the molecule.

We can look at diffusion coefficient of the molecule, all these are called the force field based approach, so, here the tools are molecular mechanics tools, okay, the other approach is based on the quantum mechanics based approach, here in the quantum mechanics based approach we can look at energy required to break a bond, form a bond, thermochemical calculations, electronic calculations, optical calculations.

So we can do all these using the quantum mechanics approach, okay, whereas in the force filed based approach or molecular mechanics based approach we cannot do these type of calculations because these calculations involve electronic energy, here we can do only structural features, we can look at the shape of the molecule, we can look at the diffusion coefficient.

We can look at minimum energies, okay, in the quantum mechanics again we have different types of quantum mechanics, semi-empirical quantum mechanics, this is slightly easier version of quantum mechanics, other one is ab initio quantum mechanics, which is detailed quantum mechanics. Again in ab initio we have 2 types, the density function approach, the Hartree-Fock approach and then as you go down it becomes correlated post Hartree-Fock approach.

So the quantum mechanics involves looking at electrons, nucleus, interaction energies and in quantum mechanics again you have the simpler version that is called semi-empirical and the more detailed version that is called ab initio methods, okay, so, we can get lot of information about the molecule as we go down in the complexity in the computation whereas if you want to look at a ligand binding to a protein.

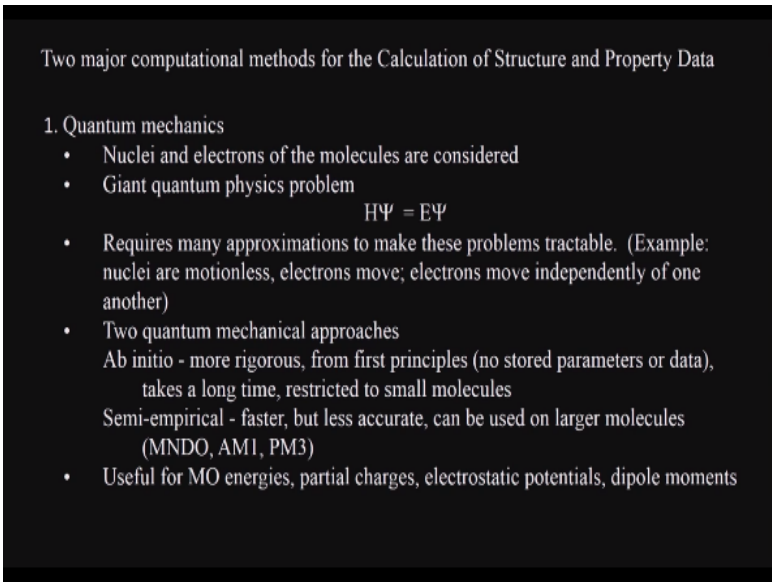
We do not go into quantum mechanics we generally use force field approach because force field approach can handle 1000s of atoms whereas quantum mechanics can handle only 100s of atom where as if you go to ab initio type of methods you can handle only 10s of atoms, so depending upon the amount of data you want depending upon the size of the molecule we decide on what type of approach we need to do.

But if you are interested in bond formation, bond breaking, electronic energy we cannot do it by force field approach we have to do by quantum mechanics approach, so if you are interested in docking of ligands, if you are interested in looking at the diffusion coefficient, if you are looking at the molecular dynamic type of studies, if you are look at confirmations the molecules can take.

It is good enough to do force field approach or molecular mechanics type of approach because it is simpler, it is faster, okay, so the level of complexity, computational resource requirement keeps going up as we go down this line, okay, so we will spend more time on

this force field and molecular mechanics approach and less time on the other quantum mechanics approach, okay.

**(Refer Slide Time: 20:08)**




2 major computational methods as I said, one is called the quantum mechanics, the other is called the molecular mechanics or force field, okay, so, what is quantum mechanics, so nuclei and electrons of the molecules are considered, so it differentiates electrons nucleus, giant quantum physics problem requires many approximations, because if I am going to do molecule which has got 1000s of atoms, I cannot go very deep into the quantum mechanics.

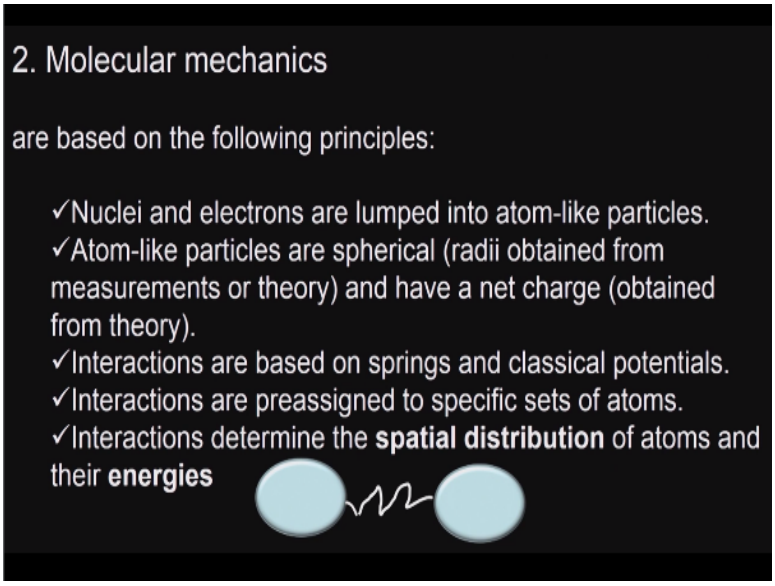
So I have to use many approximations this is very, very important to make these problems attractable, nuclear or motionless that means we assume nucleus, electrons move independently of one another okay, but we may consider only the electrons in the outermost orbit which form the bond, we will ignore electrons inside, like that you know, so in the quantum mechanics like I said we have the ab initio method which is very rigorous from first principles.

No stored parameters or data takes a long time restricted to small molecules, semi-empirical faster less accurate can be used on large molecules, these are the methods, MNDO, AM1, PM3. You will come across these names often. This is useful for molecular orbital energy, partial charges, electrostatic potentials, dipole moments, so we can use even semi-empirical methods to calculate all these things, okay.

These are very useful when you want to consider how a ligand goes and binds into the active site, so I want to know the electrostatic forces, I want to know the partial charges, okay, all these information can be obtained from this actually. There are no stored parameters in ab initio method whereas semi-empirical methods assume some parameter so it makes the calculations much simpler.

So this is the quantum mechanics method 2 divisions, the ab initio and the other one is the semi-empirical. The ab initio is much more detailed whereas we cannot do for too many atoms only for small, the other one is semi-empirical we can do lot of things with semi-empirical methods, okay it is much faster than ab initio methods and so on.

**(Refer Slide Time: 22:26)**




Okay, now the second approach in fact we are going to talk more about this because most of the ligand protein docking is done through molecular mechanics predominantly, okay, they are based on the following principles. It ignores nucleus and electrons and they are lumped into atom like particles okay, so it is like balls. They are spherical in shape, the radius obtained from measurements or theory and have a net charge obtained from theory.

So they are spherical in shape they have charge, interactions are based on springs, okay, so Hooke's law comes into picture, okay, the energy is proportional to delta x square, delta x is the change in the spring length, okay, so interactions are based on springs and classical potentials that is Hooke's law, interactions are pre-assigned to specific set of atoms okay, so we have carbon-carbon interaction.

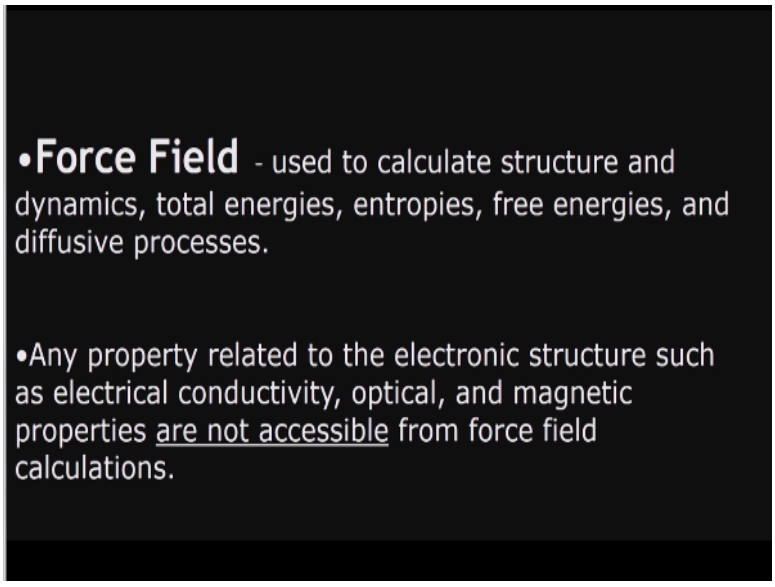
We will get different constant values, carbon-nitrogen will have different constant value and so on. Interactions determine the spacial distribution of atoms and their energy that means if some bonds get pulled or some bonds get pushed, some bonds get bent, okay, so we will have different confirmations the molecules takes that is what it is on the spacial distribution, so 2 types of techniques, the molecular mechanics, and the other one is the quantum mechanics.

Quantum mechanics differentiates electrons and nucleus, in quantum mechanics we have 2 types ab initio and semi empirical. Semi-empirical is the easier version of quantum mechanics okay, which can also determine lot of important information but it assumes certain parameters, stored parameters, it neglects many aspects that are considered in ab initio. Ab initio is good for only 10s of atoms.

So in our molecular modeling we hardly use ab initio. We use generally the force field or molecular mechanics method and little bit of semi-empirical method. The molecular mechanics methods assume the atom as a ball, it does not look at nucleus and electrons and the bonds as springs, so springs get stretched, springs get bend and there could be interactions because of the stretching and bending aspects okay.

So 2 types the molecular mechanics approach or force field approach, other one is called the quantum mechanics, in quantum mechanics we have the ab initio and the semi-empirical method, okay, force field approach. As I said we are going to spend lot of time on force field approach.

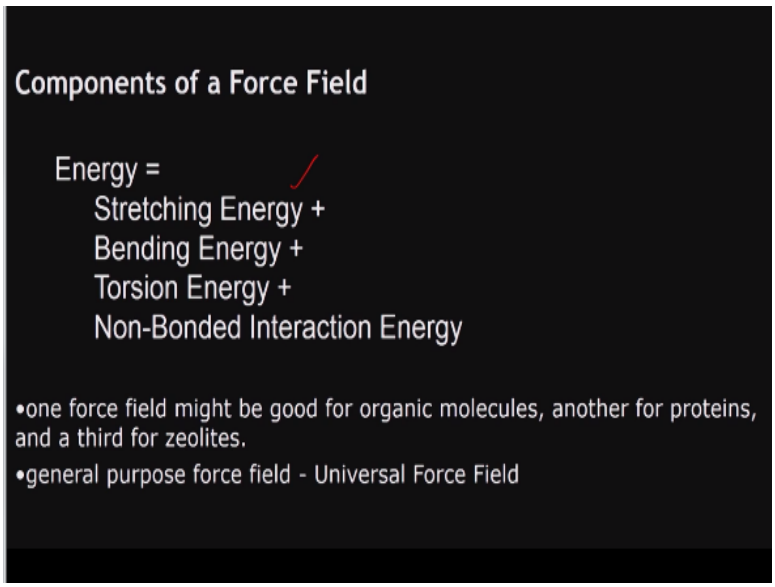
**(Refer Slide Time: 25:12)**



It is used to calculate structure, we can get the 3 dimensional structure of the molecule, conformations, dynamics, we can do molecular dynamics, we can calculate diffusion coefficient. We can calculate total energy, what is the energy of minimum energy we can calculate entropies, free energies, diffusion. Any property related to electronic structures cannot be done with molecular mechanic's approach.

Electrical conductivity, optical conductivity, magnetic properties are not accessible from force field calculations. Because it does not differentiate the electron and the nucleus, it assumes an atom as a ball, okay. So we can do free energy calculations, shape, diffusion, conformations, but anything beyond that cannot be done, okay, in the force field approach energy there are 4 terms.

**(Refer Slide Time: 26:15)**

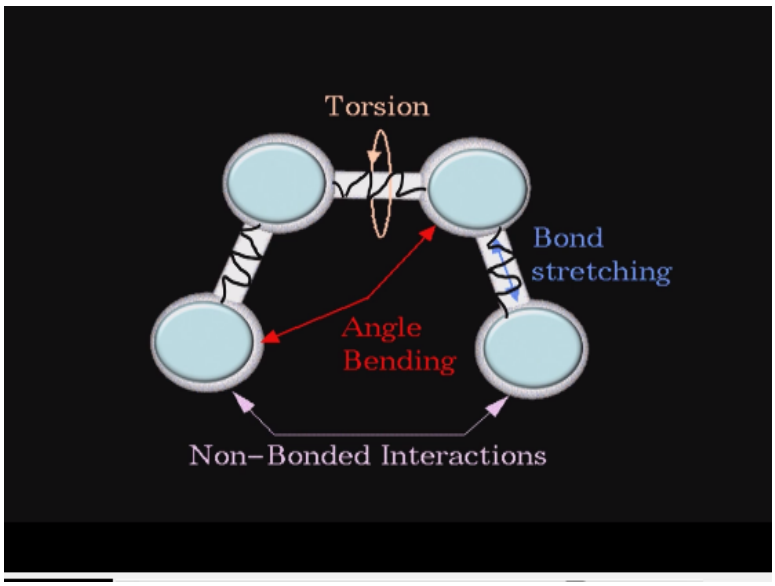



Stretching energy, that means, if you have the bond, bond can get stretched or it can be compressed. Bending energy, if there are 2 bonds, 3 atoms 2 bonds, there could be bending okay, third one is torsion energy. Torsion energy is like in the third dimension and the fourth one is the non-bonded interaction, there could be interaction between atoms which are not connected because of electrostatic forces, because of van der Waals forces.

Okay, so energy of a molecule is made up of 4 terms, stretching energy, the bond gets stretched or compressed. Bending energy, if there are 3 atoms with 2 bonds and there is a bending, torsion is the change in the third dimension, fourth is the non-bonded interaction energy. There are different types of force field, so there are force fields for organic molecules, for proteins, we look at it for say inorganic molecules like zeolites.

And of course, there are some universal force fields and so on, and so there are many different software, many different force fields depending upon what type of parameters they have, what type of approximations this each of these force fields have taken, so we have the energy of the molecule is made up of 4 terms, the stretching, bond stretching, bending, torsion, non-bonded interactions, okay. So we will look at each one of these term now in more detail.

**(Refer Slide Time: 27:57)**




The first one, look at this, assume there is a molecule with 4 atoms or 4 balls connected with 3 bonds, so this is the bond stretching, the bond gets pulled or bond gets pushed inside, so we can have both, delta x is the change in the bond length either plus or minus, but when you calculate energy like I said it is based on Hooke's law so it will become delta A square.

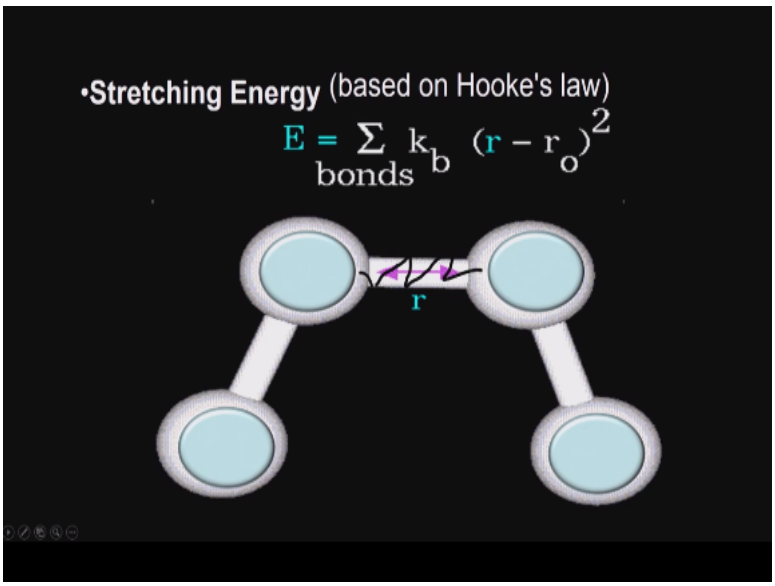
Angle bending, so if we have 3 balls or 3 atoms connected with 2 springs or 2 bonds so we can, angle can be bent here okay, so there is energy. Now we have torsion suppose you look at these 3 atoms turned, twisted in the third dimension that is called the torsion energy. The fourth one is called non-bonded interaction, look at this molecule, these 2 are not connected through any bond, but there could be some interaction because of electrostatic charges van der Waals forces that is called the non-bonded interactions.

So a molecule can have 4 types of energy terms, the bond stretching, okay, here the bond angle bending here, the torsion, the non-bonded interaction, so if you have terms mathematical relationship for each one of them and we add up all these that should give us

the total energy of a molecule, okay, there could be many bond stretching, there could be many angle bending, there could be many torsion, there could be many non-bonded interactions that are possible.

Okay, so we will look at each one of these mathematical terms. The first one is the bond stretching, it is based on Hooke's law like I said, so energy is equal to kb, kb is a constant, okay.

**(Refer Slide Time: 29:57)**




We have r-r0 square, r is the distance, r0 is the number which the distance will be or the bond length will be at equilibrium, so this shift in r0 to r squared that will contribute to the energy is the constant kb, this is summation. Okay, for example if I have say, 2 carbon ethane okay, carbon-carbon, but when I put in another carbon like propane, another carbon butane that length will not be constant.

It may get stretched, alright, that could be r whereas r0 could be the distance at equilibrium when it is not stretched so this square because Hooke's law like I said delta x square at kb, kb is a constant, so I need to know the kb, I need to know r0 to calculate the energy. Now this r0 can vary depending upon what each of these atoms are, for example, I may have carbon-carbon, carbon nitrogen, carbon oxygen, carbon sulfur, so the r0 can change.

And kb also can change and in addition carbon-carbon the distance is not going to be same you may have carbon-carbon sp3, so that will have some distance, we may have carbon-carbon double bond we call it sp2, that distance r0 may be different, we may have carbon-

carbon triple bond we call it sp, that distance is different, so I need a database which contains all these information.

So when I draw a structure the software should identify what type of carbon-carbon it is and then it will pick up the correct r0 and correct kb, okay and calculate the energy, so I need to have lot of parameters okay that is very, very important parameters, so molecular mechanics requires lot of parameters like I said, not only carbon oxygen, carbon nitrogen, carbon sulfur.

But within carbon-carbon I may have sp3 carbon, carbon-carbon in aliphatic, carbon-carbon in aromatic and carbon-carbon in 5 membered ring, so each one of them will have different r0 different kb similarly carbon oxygen can be an ether bond, can be a ketone bond, can be in a benzene ring, can be in a 5 membered ring, so you will have different values for r0 and kb for different situation.

So I need a big database of that and that is called parameterization, so I need to have parameter, so each software may have different sets of parameters taken from literature, some software may ignore some parameter, so when you run different force field, I will get different energy values, so if I am doing calculations for a set of molecules do not change the force field, use the same force field.

Otherwise you will get different answers, why because your parameter values maybe different depending upon from which literature they took and some software may ignore some type of situations or environment, so we will always get different energy values, okay, so we will continue more on these molecular mechanics or force field based energy. Thank you very much for your time.