BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

BIOMOLECULAR INTERACTIONS

There are three major types of biological macromolecules in living systems.

- Carbohydrates
- Nucleic acids
- Proteins

Though treated separately in different segments, the principles governing the organization of three-dimensional structure are common to all of them.

- Each of these macromolecules are made up of monomer units:
- monosaccharide -- for carbohydrate
- nucleotide -- for nucleic acids
- amino acid -- for proteins

Macromolecular interactions involving proteins:

- Quaternary structure refers to proteins formed by association of polypeptide subunits. Individual globular polypeptide subunits may associate to form biologically active oligomers.
- The subunits may be identical or they may be different.
- Subunit interaction is entirely **Noncovalent** between complementary regions on the subunit surface and involve following interaction:
  - Hydrophobic
  - Hydrogen bonding
  - Electrostatic (ionic)

- If **covalent links** exist (such as disulfide bridges) then the structure is not considered quaternary. In proteins with quaternary structure the disaggregated subunits alone are generally biologically inactive.
- Examples of quaternary structure, Hemoglobin.
- Hemoglobin is composed of four subunits of two types, alpha and beta, represented as $alpha_2beta_2$.
- Triose phosphate isomerase is a dimer of identical subunits.

Quaternary structure in proteins is the most intricate degree of organization considered to be a single molecule.

Incorporation of nonprotein components into proteins

- The resulting species are called conjugated proteins.

Classification of proteins by composition divide proteins into two categories.

- Simple proteins consist of only polypeptide.
- Conjugated proteins also contain a nonprotein moiety which frequently plays a role in biological function.
  - It may participate in function directly.
  - It may influence the shape of the protein.

Many different kinds of compound /macromolecules are found in conjugation/interaction with proteins. A few examples are:

- Protein-Carbohydrate
- Protein-lipid
- Protein DNA
- Protein-protein
- Protein-Heme

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

Protein-carbohydrate interaction:

One or more chains of monosaccharide units, 1 - 30 units long.It may be straight or branched, and it is usually covalently linked to the apoprotein in one of three major ways:

1. **N-linked (Type I):** N-acetylglucosamine (a sugar with an acetylated amino group in place of a hydroxyl group) at the reducing end of a carbohydrate chain is linked to the amide nitrogen of asparagine residue. The asparagine residue must be in the sequence, Asn X Thr (or Ser), where X is any amino acid residue.

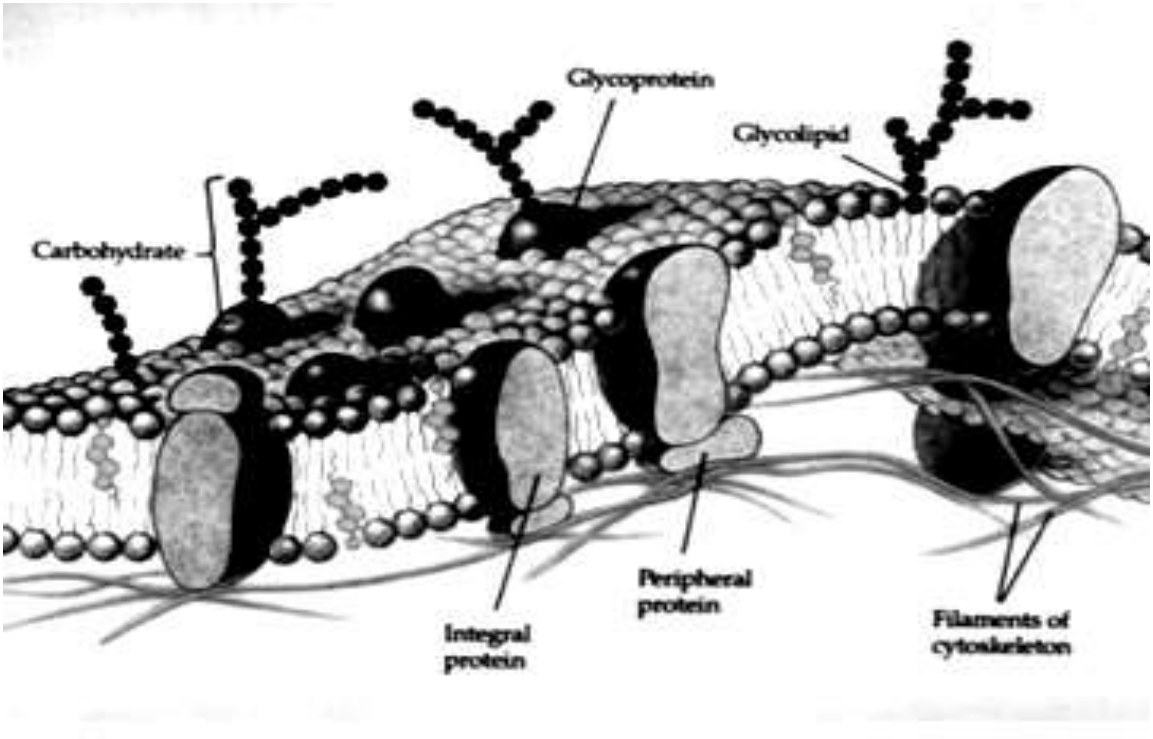


Fig1: Structure of membrane showing various components such as phospholipids bilayer, protein. Glycoproteins are important part of membrane wherein carbohydrate conjugates with proteins.

2.O-linked (Type II): Here the reducing end of a carbohydrate chain (usually N-acetylglucosamine residue) is linked to the hydroxyl of a seryl or threonyl residue.

3.O-linked (Type III): In this case The reducing end of a carbohydrate chain (usually N-acetylgalactosamine) is linked to the hydroxyl of a hydroxylysine residue in collagen. (Hydroxylysine is made from lysine in collagen after the collagen has been synthesized.)

Sl. No.	Glycoprotein	Function
1	Collagens	Structure
2	Mucins	Lubricant and protective agent
3	Transferrin	Transport
4	Immunoglobins	Immunologic
5	Hormone	Human chorionic gonadotropin (HCG), thyroid-stimulating hormone (TSH)
6	Enzyme	Various, e.g., alkaline phosphatase
8	Antifreeze protein	Certain plasma proteins of coldwater fish
9	Lectins, selectins	Interact with specific carbohydrates
10	Receptor	Various proteins involved in hormone and drug action
11	Calnexin	Affect folding of certain proteins

Table1: Examples of glycoproteins and their respective functions

Protein-lipid interaction: Protein associates with lipid through hydrophobic interactions involving the protein's hydrophobic R-groups. Lipoproteins are pseudomicellar structures. Micelles are orderly arrays of molecules having polar heads and hydrophobic tails. In water, the polar heads orient outward, and the non- polar tails cluster in the center of the micelle.

- Lipoproteins resemble micelles in some respects. The structure of lipoproteins typically includes the following features. Their outer surface is coated with polar lipids, with protein intermingled. Their interior is a region of randomly oriented neutral lipid.

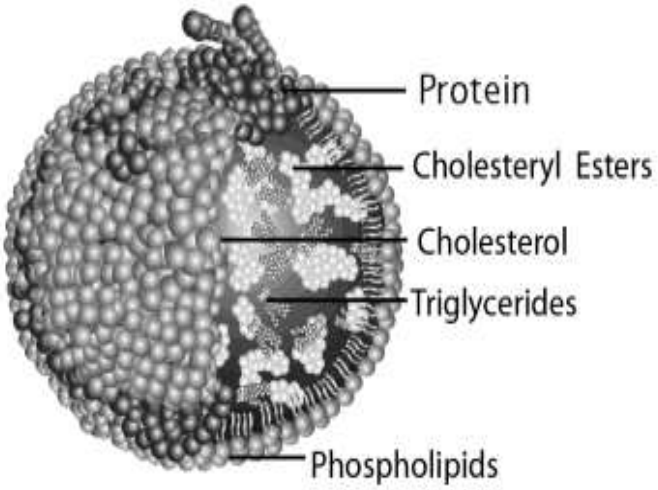


Fig2: Structure of typical micelle revealing lipid conjugation with protein

Lipid Conjugation/Protein-lipid interaction:

- The role of the polar lipid and protein on the surface is to solubilize the neutral lipid interior.
- Protein interacts with the lipid of lipoproteins through amphipathic helices. Alpha-helical regions of apolipoproteins have polar amino acids on one surface, and nonpolar ones on the opposite surface.
- The helix lies on the surface of the structure, with the polar groups oriented outward toward the water, and the nonpolar groups buried in the lipid.
- Consequence of charged surface is, it allows molecules to remain together.

Function of lipoprotein:

- Lipoprotein concentrations go up in infection. This has a protective effect.
  - They bind bacterial endotoxins.
  - They bind and neutralize a wide variety of viruses.
- Copper, transferrin and other proteins bind to HDL, making it more effective in preventing oxidation of LDL, thereby protecting against atherosclerosis.
- Membrane proteins are lipoprotein-like in that they have nonpolar amino acids in strategic locations to permit interaction with the membrane lipid. Proteins of the

membrane surface may be structured like the apoproteins of lipoproteins, with amphipathic helices.

- Some membrane proteins transverse the membrane. The region of the protein that is completely immersed in membrane should consist entirely of hydrophobic amino acids.
- A common structural motif to accomplish this is an alpha-helix consisting of at least 22 hydrophobic amino acyl groups. This makes an alpha-helix long enough to span a membrane.
- In arrays of membrane-spanning helices, helices in the interior of the array could be shorter.
- **Nucleic acid Conjugation/ Protein- Nucleic acid interaction:**

Functions that DNA-protein interactions :

- DNA replication, DNA repair, DNA recombination, transcription etc
- **Structure of DNA is regular:** a list of the positions of the atoms in the double helix, the stability of DNA and the Chargaff rules, and provided a model for how DNA stores genetic information.

- **Proteins are much less regular:**The first structures for NA binding proteins (NABP) were of the stable and abundant nucleases binding to single-stranded nucleotides.
- Work on more complicated protein-NA complexes (*e.g.*, repressors, polymerases, tRNA synthetases) required two advances techniques:
    - for overexpressing normally scarce proteins, and
    - for synthesizing large amounts of oligonucleotides (ON).

- Only in the late 1970s did it become reasonable to try to determine the structure of an ON, the protein it interacts with, and the complex between the two.
- **There are four major forces** that occur when proteins and NA interact:

➢ Electrostatic forces: salt bridges
➢ Dipolar forces: hydrogen bonds
➢ Entropic forces: the hydrophobic effect
➢ Dispersion forces: base stacking

- **Electrostatic forces:** (salt bridges) Electrostatic forces are long range, not very structure-specific, and contribute substantially to the overall free energy of association.
- Salt bridges are electrostatic interactions between groups of opposite charge. They typically provide ~40 kJ/mol of stabilization per salt bridge.
- In protein-NA complexes, they occur between the ionized phosphates of the NA and either the e-ammonium group of lysine, the guanidinium group of arginine, or the protonated imidazole of histidine**.**
- Salt bridges are influenced by the concentration of salt in the solution: as it increases, the strength of the salt bridges decreases.
- Salt bridges are much stronger in the absence of water molecules between the ionized groups (because water has high dielectric constant, e).
- Salt bridges alone, therefore, can not distinguish one B-DNA sequence from another.
- Patterns of salt bridges, however, could clearly be used to distinguish ss- from ds-NA, and B- from Z-DNA.

Dipolar forces: (hydrogen bonds)

- Hydrogen bonds are dipolar, short-range interactions that contribute little to the stability of the complex but much to its specificity.
- Hydrogen bonds occur between the amino acid side chains, the backbone amides and carbonyls of the protein, and the bases and backbone sugar-phosphate oxygens of the NA.
- When protein-NA molecules are not complexed, all their exposed hydrogen bond donors (X) and acceptors (Y) form linear hydrogen bonds to water. Hydrogen bonds are a result of dipole-dipole interactions:

$X^{d-}\_\_H^{d+}$ ---- $Y^{d-}\_\_R^{d+}$ (X, Y = nitrogen or oxygen)

When the complex forms, there is little change in the free energy due to hydrogen bond formation if the linear hydrogen bonds to water are replaced by a similar ones between the macromolecules.

By contrast, forming bent hydrogen bonds carries a free energy penalty of up to ~4 kJ/mol per hydrogen bond.

Thus hydrogen bonds are very important in making sequence-specific protein-NA interact.

Entropic forces: (the hydrophobic effect)

Hydrophobic forces are short range, sensitive to structure, proportional to the size of the macromolecular interface, and contribute to the free energy of association.

The hydrophobic effect is due to the behaviour of water at an interface.

When molecules aggregate, the ordered water molecules at the interface are released and become part of the disordered bulk water, thus stabilizing the aggregate by releasing the entropy of the system.

Dispersion forces: (base stacking)

Dispersion forces have the shortest range but are very important in base stacking in double-stranded NA and in the interaction of protein with ssNA.

Base stacking is caused by two kinds of interaction: the hydrophobic effect and dispersion forces**.**

Molecules with no net dipole moment can attract each other by a transient dipole-induced dipole effect. It is very sensitive to the thermal motion of the molecules.

For dsNA, dispersion forces are clearly important in maintaining the structure by base stacking.

For ssNA, they also help it to bind proteins because aromatic side chains can intercalate between the bases of a ssNA

Geometric constraints imposed by the nucleic acid:

All NA have repeating polyanionic backbones and so all proteins that bind them have strategically placed arginines and lysines that create an electrostatic field to neutralize the negative charge.

Possibly because the structure of RNA varies more than that of DNA, proteins seem to recognize RNAs in more ways than they recognize DNAs

All sequence-specific DNA binding proteins may bind DNA in two ways: one for tight, sequence-specific binding and the other for looser, non-sequence specific binding.

Non-specific interactions

Single-stranded nucleic acid binding proteins , Non-sequence-specific nucleases ,Polynucleotide polymerases

The need for packaging:

The fundamental building block of chromatin in eukaryotes is the nucleosome, a protein-DNA complex. The nucleosome core particle consists of 146 bp of DNA and eight small, highly basic histone proteins. The DNA wraps around the histone octomer to form a negative supercoil.

Bacteria also use small basic proteins to package DNA, such as the dimeric HU protein from *E. coli*, whose long b-strand arms presumably wrap around a double-stranded DNA molecule.

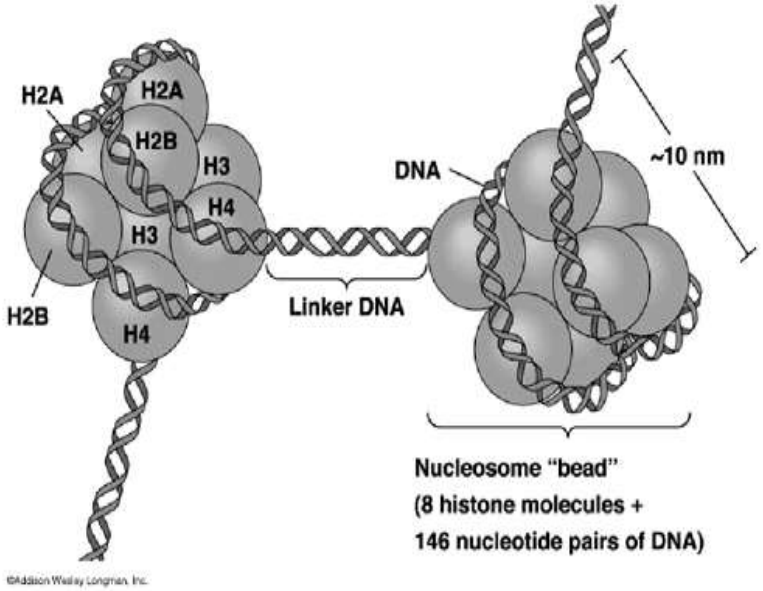


Fig3:Role of proteins in packaging of chromosomes in Eukaryotes

Single-stranded nucleic acid binding proteins:

ssDNA is formed during replication and most organisms produce proteins to bind it. These proteins form an important but diverse group but, with the exception of gene 5 protein from bacteriophage fd, there is little structural information on how they interact with NA.

A model has been suggested in which lysines and arginines neutralize the DNA phosphate backbone and the bases stack against aromatic amino acid side chains.

Non-sequence-specific nucleases:

All organisms must degrade NA during their life cycle. There is no one enzyme designed for this purpose, but rather a large number of enzymes with different specificities. These include exo- and endonucleases and enzymes specific for ss- and ds-NA and for base sequences. *e.g.*, RNase and DNase

RNase and DNase have different reaction mechanisms because RNase uses the ribose 2'-hydroxyl group, not present in DNA, to attack the 5'-phosphate ester linkage.

Ribonuclease A: RNase A is not sequence specific because it only interacts with the base at the active site; all other contacts are electrostatic ones to the sugar-phosphate backbone.

Deoxyribonuclease I: DNase I cleaves different sequences with different rates because of sequence-dependent steric hindrance at the active site. G-C tracts accommodate the catalytic loop better because they have wider minor grooves than A-T tracts.

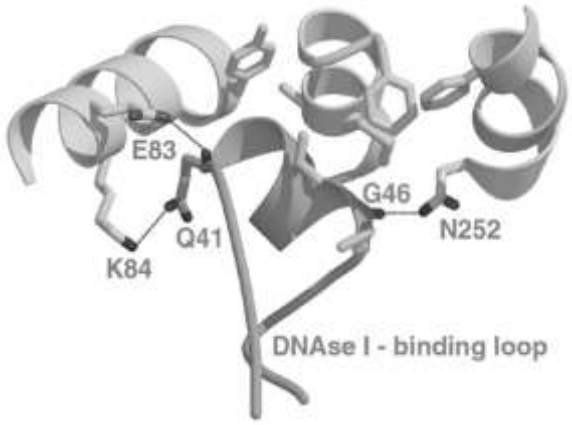


Fig4:Interaction between DNA and DNASE 1 Protein

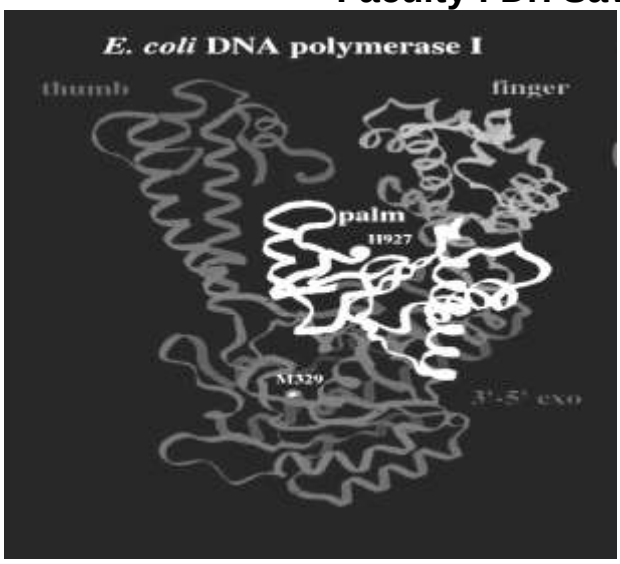
- Polynucleotide polymerases:There are four classes of template-directed polynucleotide polymerases: DNA- or RNA-dependent and DNA- or RNA-polymerizing.

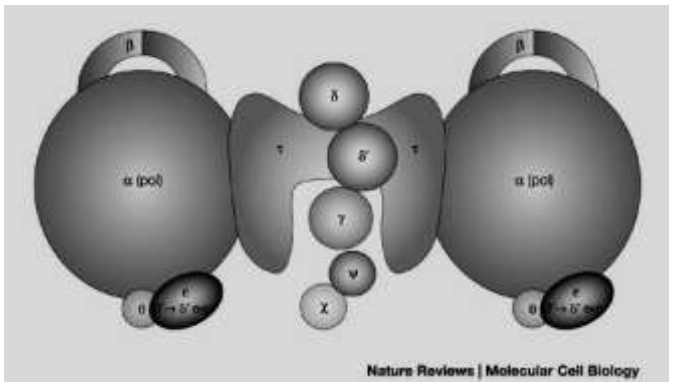
- All add nucleotides to the 3'-end of a growing polynucleotide chain but they differ widely in how accurately they replicate the NA (their fidelity) and how many nucleotides they add before dissociating (their processivity).

DNA-dependent DNA polymerases: E. coli DNA polymerase I (Pol I) and III

- All cellular DNA-dependent DNA polymerases have a 3'-5' proof-reading exonuclease, require a primer to begin synthesis, and replicate their own NA the most faithfully.
- The Klenow fragment of Pol I contains two widely-separated domains, one carrying the polymerase activity, and the other the 3'-5' proofreading exonuclease activity.

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat





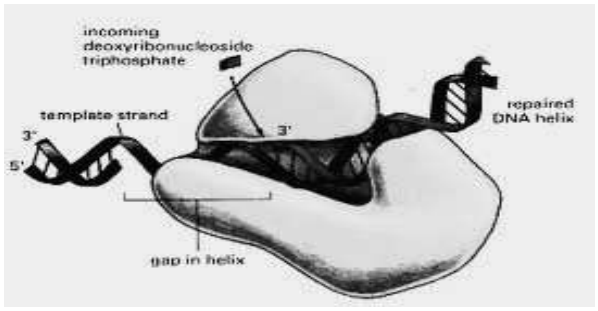


Fig 5: Structure of different DNA polymerases, DNA POLI(left) and DNA PolIII(right) and pol enzymes holding the DNA during replication (below)

Specific interactions

The placement of an a-helix in the major groove appears to be the most common way of recognizing a specific DNA sequence.

Other parts of the protein, which form hydrogen bonds and salt bridges to the DNA backbone, position the element on the DNA so that it can achieve recognition

Restriction endonucleases:

EcoRI and EcoRV

EcoRI and EcoRV have very different structures and interact with DNA differently: the former only in the major groove; the latter in both grooves.

However, both employ the same enzyme mechanism and catalytic residues and both achieve their high degree of sequence specificity similarly.

Fig6: Structure of restriction endonuclease , shown here the interaction between DNA and protein

The need for specificity:

For a cell to function at all, proteins must distinguish one NA from another very accurately.

Proteins that bind specific NA sequences also bind non-specific ones.

In some cases, like the transcriptional regulators, this binding is intrinsic to function; in others, like the tRNAs, the binding is merely unproductive.

Transcriptional regulators: the conserved motif

- **The zinc finger motif**
- **The leucine zipper**
- **The helix-turn-helix motif**

The zinc finger motif

- A small Zn-stabilized structural domain found in proteins that interact with nucleic acids. The zinc finger is a loop of about 25 amino acyl residues stabilized by a Zn atom.
- Zn complexed to His and/or Cys maintains the structure of the domain.
- Unlike a -S-S- bridge, the Zn complex will not be broken by reducing conditions within the cell.
- Unlike Cu or Fe, Zn does not participate in oxidation-reduction reactions that could generate free radicals which might damage nucleic acids.
- Other amino acyl residues in the loop are involved in binding to specific nucleotides of the nucleic acid or helping to maintain the folded structure of the domain.

- Zinc fingers occur in proteins occur in tandem arrays. They are joined to nearby zinc fingers by short linking regions of peptide. They are spaced to fit into the major groove of DNA, with the bases of the alpha-helices down in the grooves, and the beta-loops touching the double helix.

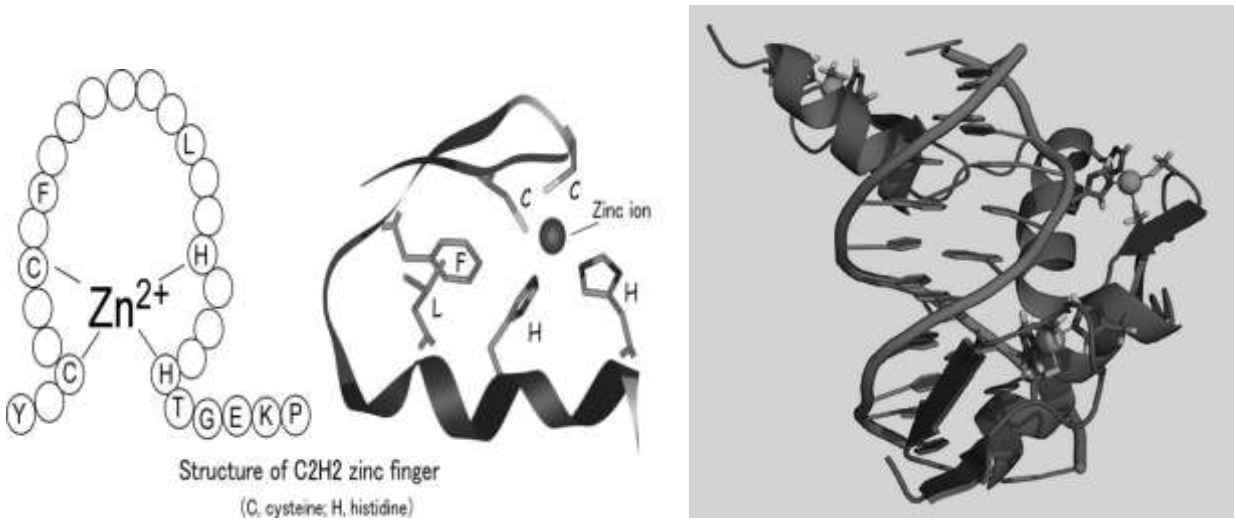


Fig7:Structure of CH2 type zinc finger motif and interactyion of DNA with the domain bearing zinc finger motif.

The leucine zipper

- A pair of amphipathic alpha-helices joining two subunits of a dimeric protein that binds to DNA. Some sites in DNA important to biological control have twofold symmetry: the base sequence is the same in both directions.
- Example: 5' ...TGACTCA... 3'
  3' ...ACTGAGT... 5'

A protein designed to bind at such a site might also be symmetric; this could be accomplished if the protein were a head-to-head dimer

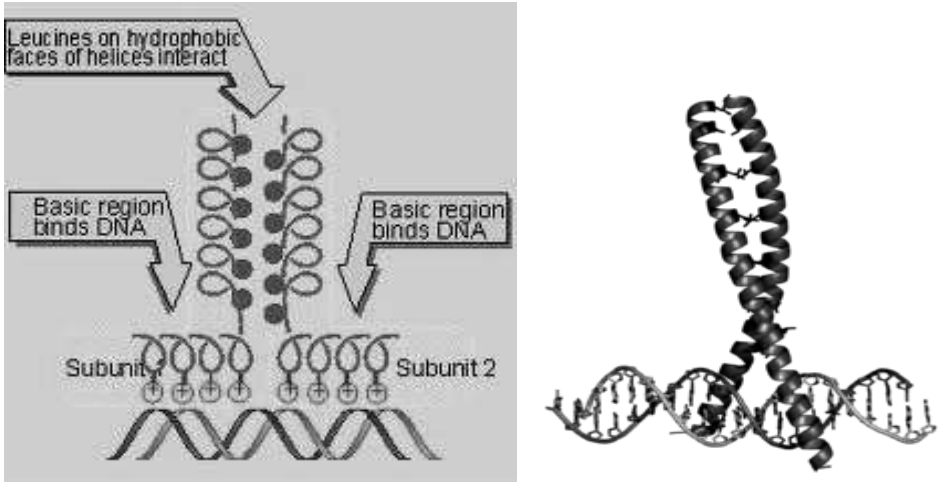


Fig8: Structure of leucine zipper and DNA interacting with zipper

- **The leucine zipper**
- A class of DNA binding proteins appears to form such dimers through alpha-helices having regularly spaced leucyl residues along one edge. Interaction between the protein monomer units is thought to be through leucyl residues along the edges of the amphipathic helices, sort of like the 4-helix bundle, but with just two helices. Originally it was thought that the leucyl residues interdigitated (hence the name, "leucine zipper"), but it is now believed that they face each other (reality in the form of x-ray crystallography strikes again). In any case, the symmetric dimer binds to the symmetric region of the DNA through special binding domains.

The helix-turn-helix motif :

- Two short adjacent alpha-helices that cross one another. One alpha-helix fits into the major groove of DNA, and interacts with specific bases; this is called the recognition helix. A short segment of protein links the recognition helix to a second helix; this is the turn, and is so named because it contains a so-called beta-turn, a well recognized structural element of proteins. The second helix lies across the major groove of DNA, and binds nonspecifically.
- A dimeric protein can have a helix-turn-helix motif in each subunit, and if the monomer units are identical it can thereby recognize and bind to symmetric DNA structures

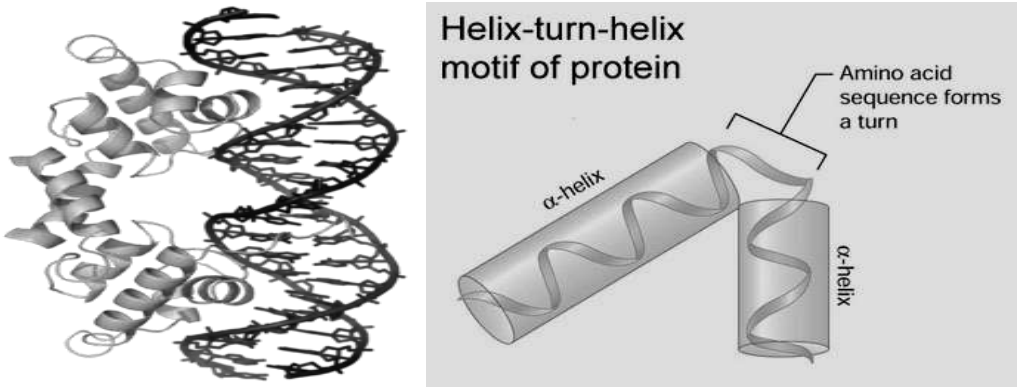


Fig9: Structure of helix turn helix motif and interaction of protein bearing HTH with major groove of DNA.

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

Protein Conjugation/Protein-protein interaction

insight review articles

Proteomics to study genes and genomes

Akhilesh Pandey* & Matthias Mann†

** Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, and Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA*
†Protein Interaction Laboratory (PIL), University of Southern Denmark, Campusvej 55, DK-5230 Odense M, and Denmark and Protana A/S, Staermosegaardsvej 16, DK-5230 Odense M, Denmark (e-mail: mann@cebi.sdu.dk)

Proteomics, the large-scale analysis of proteins, will contribute greatly to our understanding of gene function in the post-genomic era. Proteomics can be divided into three main areas: (1) protein micro-characterization for large-scale identification of proteins and their post-translational modifications; (2) 'differential display' proteomics for comparison of protein levels with potential application in a wide range of diseases; and (3) studies of protein–protein interactions using techniques such as mass spectrometry or the yeast two-hybrid system. Because it is often difficult to predict the function of a protein based on homology to other proteins or even their three-dimensional structure, determination of components of a protein complex or of a cellular structure is central in functional analysis. This aspect of proteomic studies is perhaps the area of greatest promise. After the revolution in molecular biology exemplified by the ease of cloning by DNA methods, proteomics will add to our understanding of the biochemistry of proteins, processes and pathways for years to come.

- Protein characterization for large scale identification of protein and PTM
- Differential display of proteomics for comparison of protein levels with application to wide range of diseases.
- Study of protein –protein interaction via Y2H and MS.
- Intrinsic to every cellular process
- Form the basis of phenomena
  -DNA replication and transcription
  -metabolism
  -signal transduction
  -cell cycle control

The Study of Protein-protein Interaction by Mass Spectrometry

A key question about a protein,in addition to when and where it is expressed, is with which other proteins does it interact. Interaction partners are an immediate lead into biological function and can potentially be exploited for therapeutic purposes. Creation of a protein–protein interaction map of the cell would be of immense value to understanding the biology of the cell.

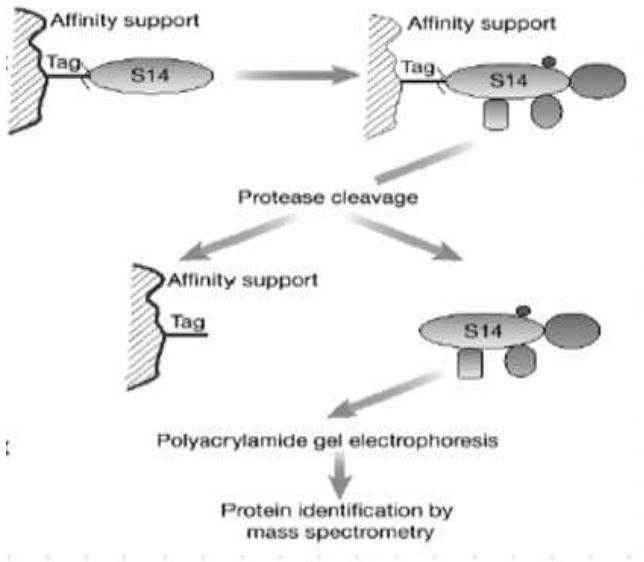


Fig 10:Steps involved in studying Protein interaction with other small molecules

The study of Protein-protein Interaction by Mass Spectrometry

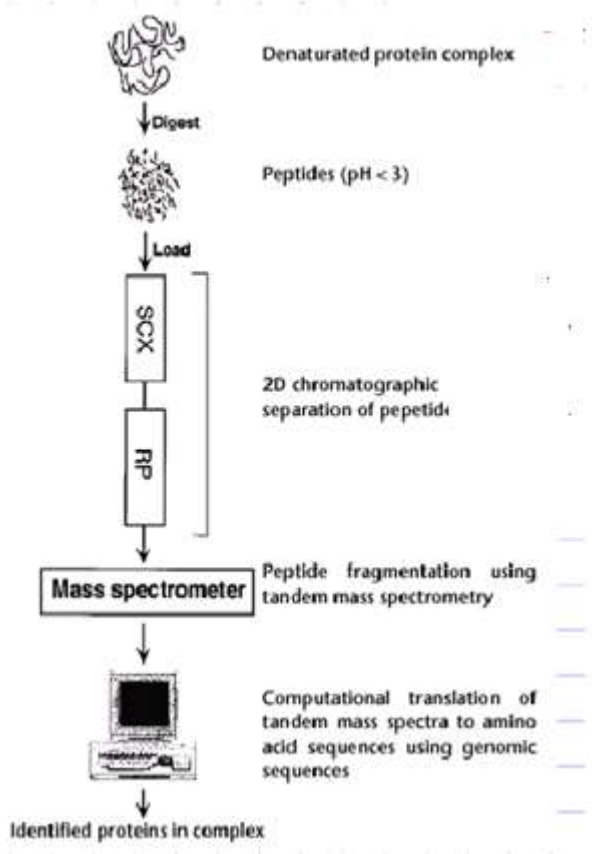
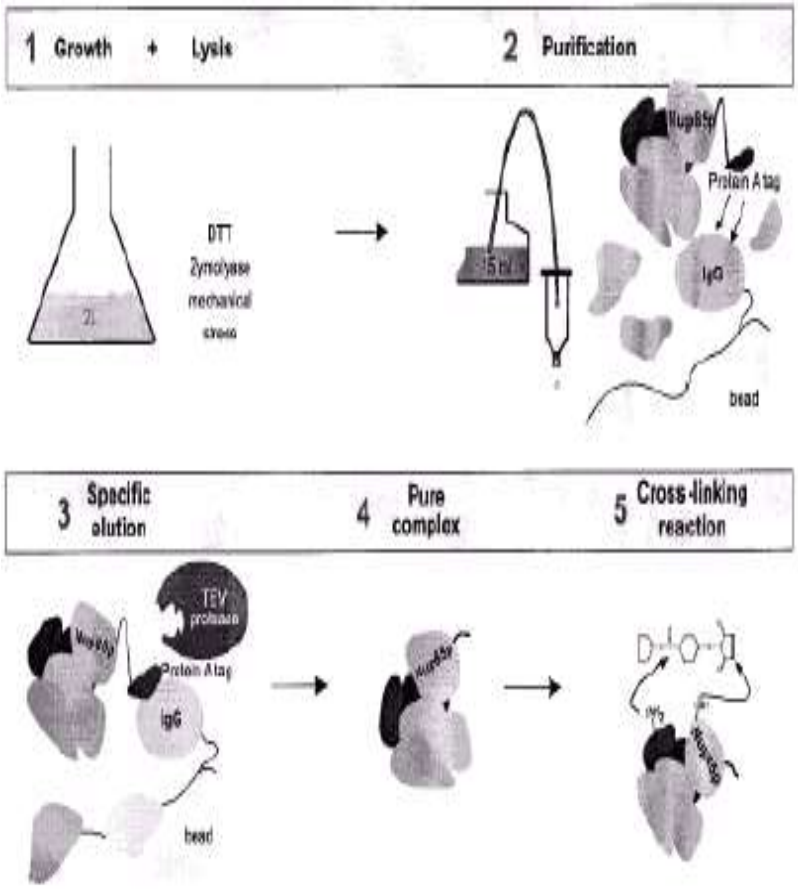


Fig11:Diagramatic representation of Mass spectroscopy protocol.

Generic Strategy to Analyze the Spatial Organization of Multi-protein Complexes by Cross-linking and Mass Spectrometry

- Growth
- Lysis
- Purification of Protein
- Specific elution
- Formation of pure complex
- Cross linking with other molecules
- SDS PAGE –separation
- Sample preparation
- MS and analysis



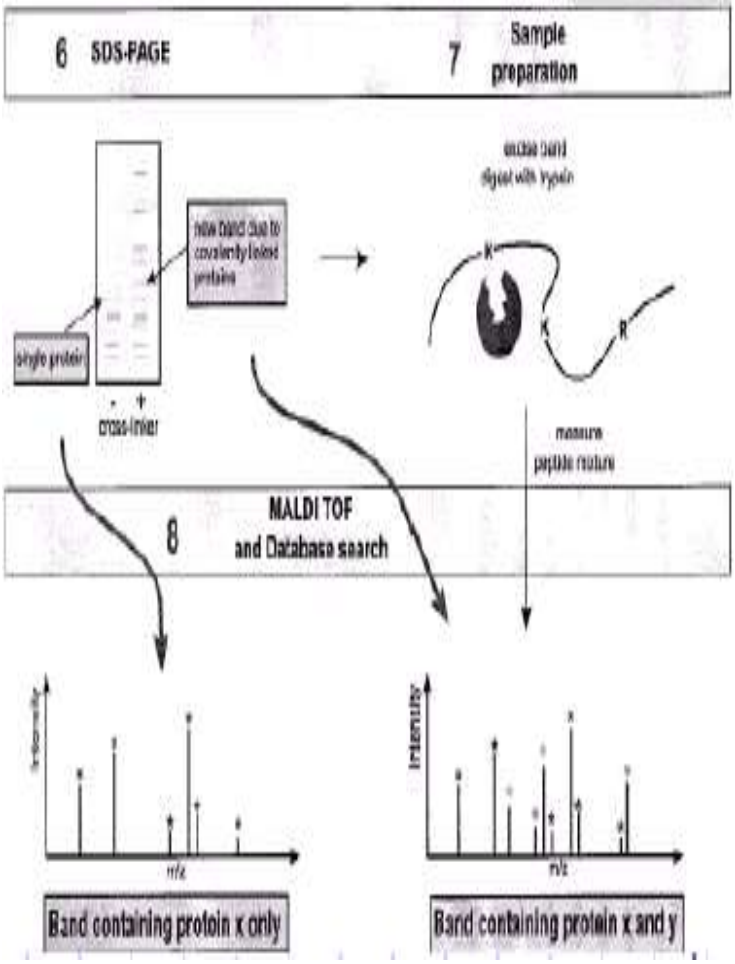


Fig12:Diagramatic representation of steps involved in complete characterization of protein and its interacting partner.

Yeast Two-hybrid

Researchers insert a gene in yeast for a "bait" protein alongside DNA for half of an "activator" protein. The other half of the activator DNA is then inserted alongside DNA for random "prey" proteins. The yeast cells are then grown up and the proteins are allowed to interact. If bait and prey proteins bind, the two halves of the activator protein be close enough to work together to turn on another yeast gene that turns the cell blue, signaling a match.

- Useful in the study of various interactions
- The technology was originally developed during the late 1980's in the laboratory Dr. Stanley Fields (Fields and Song, 1989, *Nature*).

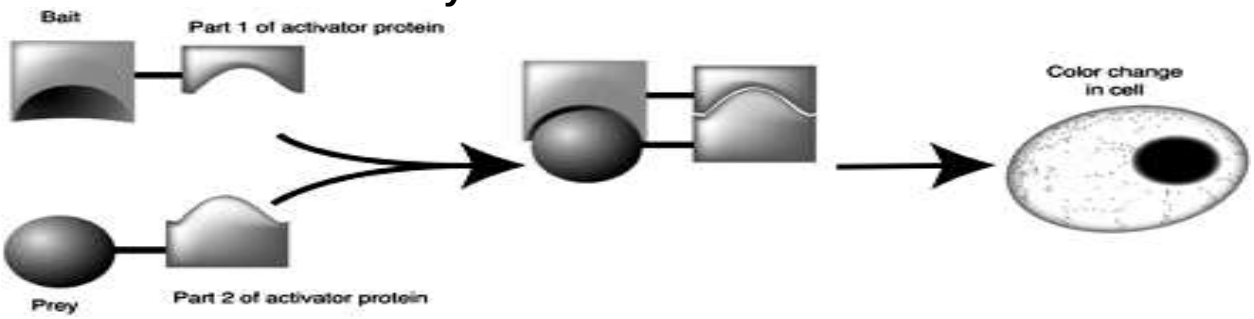


Fig 13: Y2 H system and expression of protein once bait and prey interact due to part1 & 2 of activator domain fit

Phage Display

- A method where bacteriophage particles are made to express either a peptide or protein of interest fused to a capsid or coat protein.
- It can be used for peptide epitopes, peptide ligands, enzyme substrates or single-chain antibody fragments.

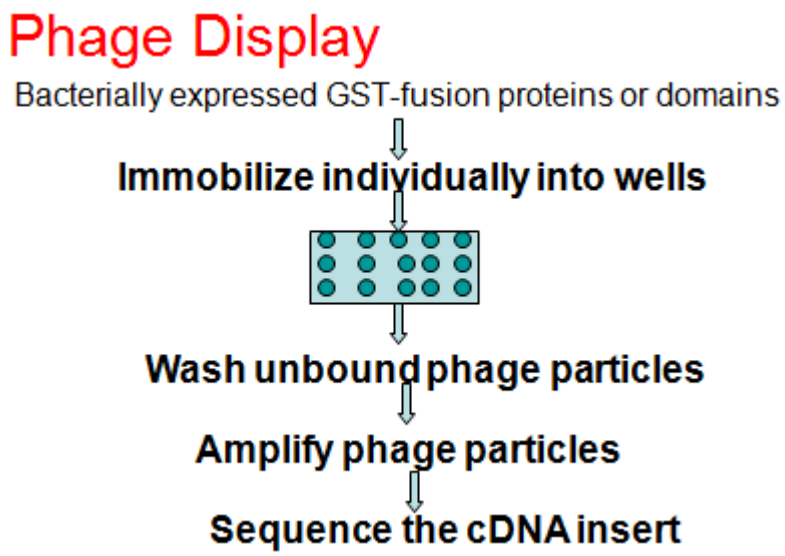


Fig 14: Flow chart displaying the method of studying protein protein interaction via phage display method.

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

Structural Biology

10BT43

UNIT 8:
MOLECULAR DYNAMICS

MOLECULAR MECHANICS AND DYNAMICS- Importance and relevance

- Modeling modeling and computational chemistry are important parts of modern biochemistry.
- Modeling is important to display in a meaningful and instructive fashion the large amounts of data produced when x-ray crystallography and NMR are used to determine the structure of large biological molecules and complexes.
- However, the primary x-ray crystal data which are in the form of electron density maps, must be interpreted like any other type of data.
- Structures need to be refined and energy minimized to produce more realistic structures for eg., without van der Waals overlap or missing atoms.
- In addition, atoms within any molecule are not static, but move as bonds vibrate, angles bend, etc. This implies that large biomolecules could adopt many possible conformations of different energies.
- For proteins, some of these conformations might center around a average conformations situated at a local or global energy minimum separated from each other by activation energy barriers.
- In contrast to small molecules whose structure can be minimized using *ab initio* or semi-empirical quantum mechanics (using programs such as Spartan), large molecular structures like DNA, RNA, proteins and their complexes must be minimized using molecular mechanics, based on Newton's laws.
- Atoms are treated as masses, and bonds as springs with appropriate force constants. A force field, containing all the relevant parameters for given atom (for example sp^3, sp^2, sp^2 aromatic, and sp C) and bond types is used to solve energy equations which sum all energies, over all atoms and bonds in the molecule.
- These energies include interactions among bonded atoms (stretching, bending, torsion, wagging) and those among nonbonded atoms (electrostatic and van der Waals).

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

Minimizations calculations :

- For minimizations calculations, the positions of the atoms within a molecule must be systematically or randomly moved and the energy recalculated with the goal of finding a lower energy and hence more stable molecule.
- Minimization calculations can not probe all conformational space and can not easily move a structure from a local minimum to a global minimum if two are separated by a large energy barrier. Energy minimizations are usually done in the absence of solvent.

Software tools for force fields:

- A common force fields used for macromolecules is CHARMM, AMBER, and GROMOS. Parameters for specific atom type in a given bond include atomic mass, van der Waals radius, partial charge for atoms (from quantum mechanics) and bond length (from electron diffraction data), angles, and force constants for bonds (modeled as springs, obtained from IR).
- These parameters are derived from experiments and theoretical (usually quantum mechanical) calculations on small organic molecules. A potential energy equation comprised of terms from bond stretching, angle bending, and torsion angle changes (bonded interactions) as well as electrostatic and van der Waals interactions (nonbonded) is then solved .
- The goal of molecular dynamics is to simulate the actual changes in a molecule as a function of time after an energy input (heat application at a higher temperature) is added to a molecule at equilibrium.
- To make the simulation realistic, the structure is placed in a "bath" of thousands of water molecules. As is described below, if the energies of atoms in a large molecule are known, the forces acting on those atoms can be deduced.
- From Newton's Second Law (F=ma), the velocity or change of position of an atom in the structure with time can be determined.
- If the dynamic simulation can be run for a long enough period of time, alternate conformations (perhaps those centered around a global minimum as well as those nearby in energy space - a local minimum) may be sampled.
- By determining what fraction of the simulated conformations resemble the two alternative conformations, the DG for the interconversion of the two states can be calculated.

- These calculations require large amounts of computer time. They give very important information, however, since protein conformational changes are often, if not always associated with binding of a biological molecule to a binding partner, *In silico* experiments offer important clues and support to results obtained using other methods of study.

  Relationship between Energy (E), Force (F) and Motion

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

To make the energy equations for the individual components more understandable, it is useful to consider the relationship between force and energy. There are two general force equations in chemistry and introductory physics. One is Coulomb's Law, which describes the electrostatic force of attraction, F_C, between two charges, q_1 and q_2, separated by a distance r.

$$F_c = k\frac{q_1 q_2}{r^2}$$

- The other is Hooke's Law, which describes the restorative force on a mass connected to a spring on stretching or compression of the spring.

$$F = -kx$$

- where x is the displacement of the spring from an equilibrium (at rest) position. Equations leading to determination of PE:
- Its important to understand how these equations might lead to equations which describe the potential energy of a two charge system or of a compressed or stretched spring.
- This can be best understood by studying the simple example of a ball placed at various locations on a hill. If placed on a flat surface at the top and bottom of the hill, there is no net force on the ball ($F_{net} = 0$), so it will not move.
- If placed at various locations on the downslope, it will experience a net downward force, shown in a qualitative fashion in the figure below.

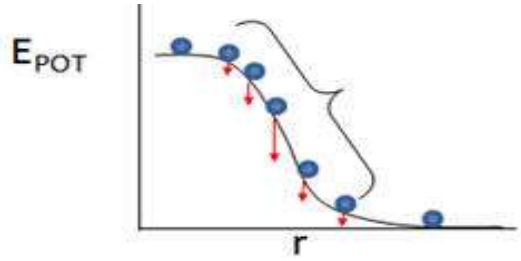


Fig1: plot: Energy vs distance.

- The magnitude of the force vector is proportional to the slope.From this simplistic approach, following equation is derived for relating F to E:

$$F = -\frac{dE}{dr}$$

- The minus sign is required since the force is downward but the energy increases upward.

- This simplified approach can be extended into three dimensions, to give the following equation where F is the negative gradient of the potential energy:

$$F = -\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}\right)E = -\nabla E$$

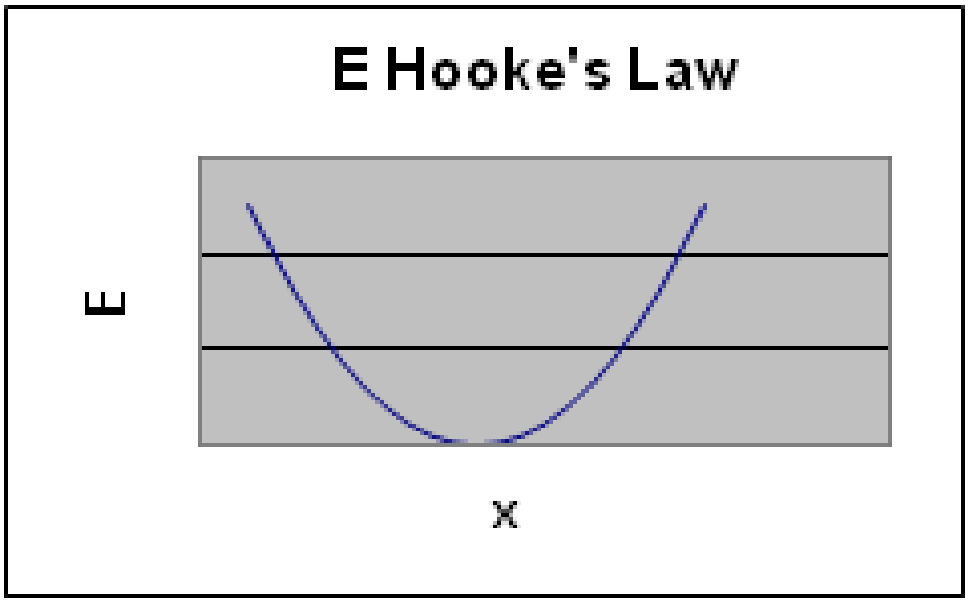
- Applying the 1D equation to Hooke's Law gives

$$dE = -Fdr = -kxdx$$

$$\int dE = -k\int xdx$$

$$E = \frac{kx^2}{2}$$

This gives a parabolic graph of E vs displacement



The same approach can be applied to Coulombs Law. Notice that the result equation for E results in increasingly negative values as r get smaller only when q_1 and q_2 have opposite charges

$$dE = -F dr = -k\frac{q_1 q_2}{r^2} dr$$

$$\int dE = -kq_1 q_2 \int r^{-2} dr$$

$$E = \frac{kq_1 q_2}{r}$$

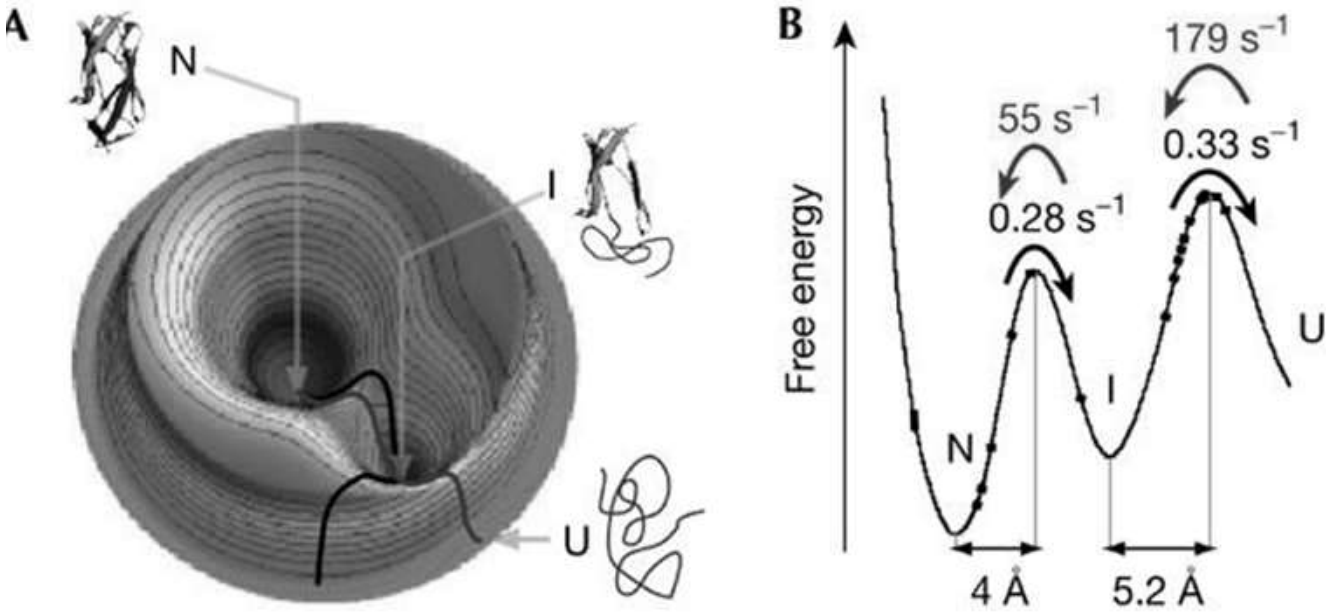


Fig2:A 3D energy landscape for the folding of a complicated protein molecule is shown .

MOLECULAR MECHANICS:

The mechanical molecular model was developed out of a need to describe molecular structures and properties in as practical a manner as possible. The range of applicability of molecular mechanics (MM) includes:

- Molecules containing thousands of atoms
- Organics, oligonucleotides, peptides, and saccharides
- Vacuum, implicit, or explicit solvent environments.
- Ground state only
- Thermodynamic and kinetic properties.

The great computational speed of MM allows for its use in procedures such as molecular dynamics (MD), conformational energy searching, and docking, that require large numbers of energy evaluations.

- MM methods are based on the following principles:
- Nuclei and electrons are lumped into atom-like particles
- Atom-like particles are spherical (radii obtained from measurements or theory) and have a net charge (obtained from theory)
- Interactions are based on springs and classical potentials
- Interactions must be preassigned to specific sets of atoms
- Interactions determine the spatial distribution of atom-like particles and their energies.

These principles differ from those of quantum mechanics.

The Anatomy of a MM Force Field:

- The mechanical molecular model considers atoms as spheres and bonds as springs. The mathematics of spring deformation can be used to describe the ability of bonds to stretch, bend and twist:

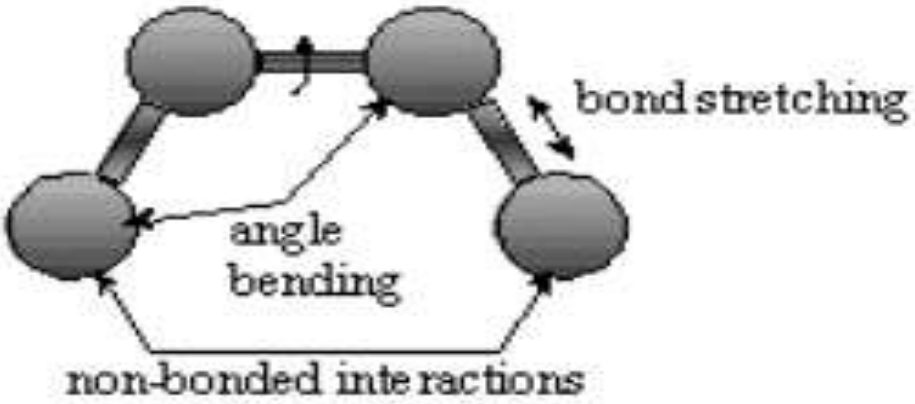
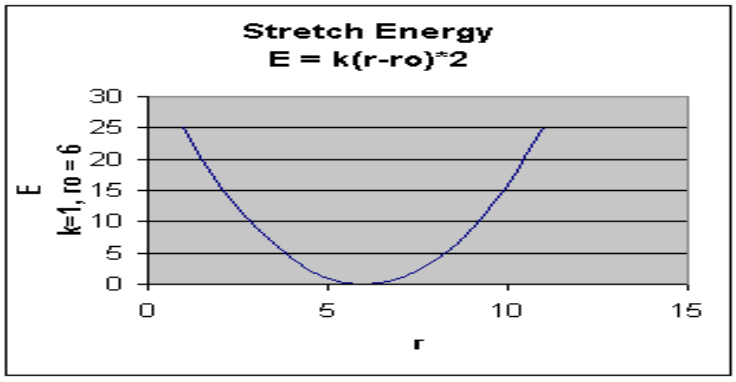


Fig 3:types of interaction between 2 atoms.

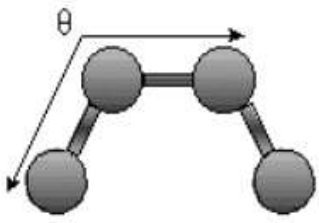
- Non-bonded atoms (greater than two bonds apart) interact through van der Waals attraction, steric repulsion and electrostatic attraction/repulsion. These properties are easiest to describe mathematically when atoms are considered as spheres of characteristic radii.
- The object of MM is to predict the energy associated with a given conformation of a molecule. However, MM energies have no meaning as absolute quantities. Only difference in energy between two or more conformations have meaning. A simple MM energy equation is given by:

Energy (E) = $E_{Stretch} + E_{Bending} + E_{Torsion} + E_{Non\text{-}bonded\ Interactions}$

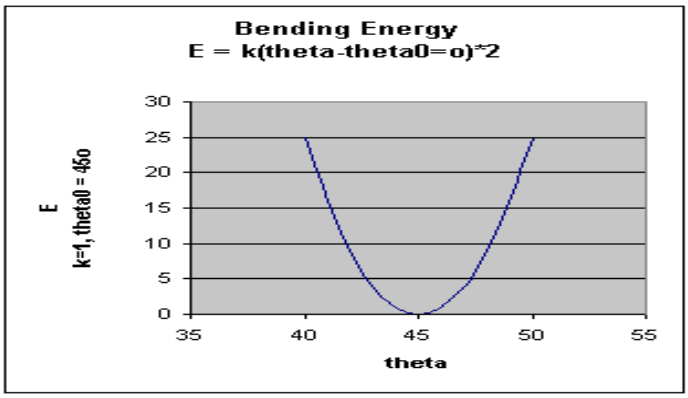
- These equations together with the data (parameters) required to describe the behavior of different kinds of atoms and bonds, is called a force-field. Many different kinds of force-fields have been developed over the years. Some include additional energy terms that describe other kinds of deformations. Some force-fields account for coupling between bending and stretching in adjacent bonds in order to improve the accuracy of the mechanical model.
- The mathematical form of the energy terms varies from force-field to force-field.
- Stretching Energy: $E_{stretch} = S_{bonds} k_b (r - r_o)^2$
- The stretching energy equation is based on Hook's law. The k_b parameter controls the stiffness of the bond spring, while r_o defines its equilibrium length. Unique k_b and r_o parameters are assigned to each pair of bonded atoms based on their types (e.g. C-C, C-H, etc.). This equation estimates the energy associated with vibration about the equilibrium bond length. This is the equation of a parabola, as can be seen in the following plot:



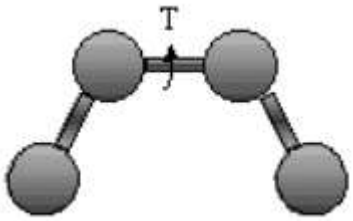
- Note that the model tends to break down as a bond is stretched toward the point of dissociation.



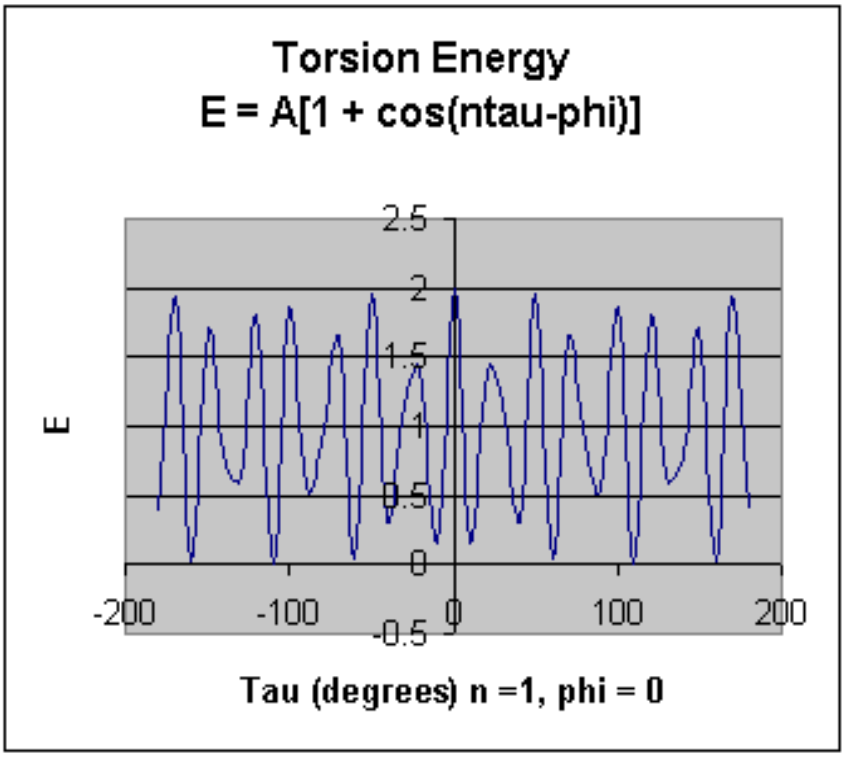
- Bending Energy: $E_{bending} = S_{angles} k_Q (Q - Q_o)^2$
- The bending energy equation is also based on Hook's law. The k_Q parameter controls the stiffness of the angle spring, while the Q defines it equilibrium angle. This equation estimates the energy associated with vibration about the equilibrium bond angle .



- Unique parameters for angle bending are assigned to each bonded triplet of atoms based on their types. (e.g. C-C-C, C-O-C, etc.).
- The effect of the kb and k_Q parameters is to broaden or steepen the slope of the parabola. The larger the value of k, the more energy is required to deform an angle (or bond) from its equilibrium value. Shallow potentials are achieved for k values between 0.0 and 1.0.
- Torsion Energy $E_{torsion} = S_{torsions}$ A [1 + cos (ntau - phi)]
- The A parameter controls the amplitude of the curve, the n parameter controls its periodicity, and phi shifts the entire curve along the rotation angle axis (tau).



- The torsion energy is modeled by a s simple periodic function, as can be seen in the following plot:



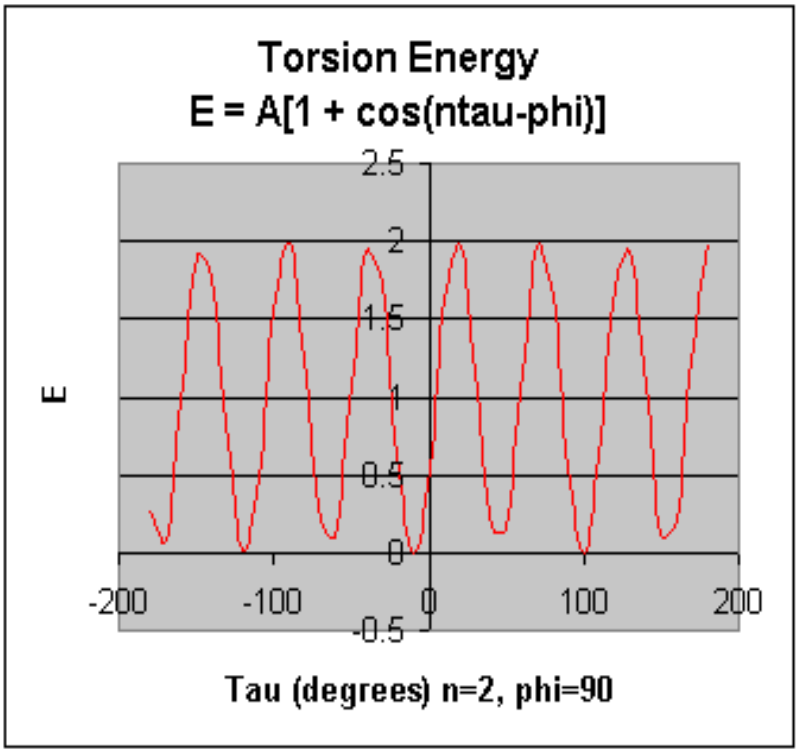


Fig 4: Modeled torsion energy by a s simple periodic function.

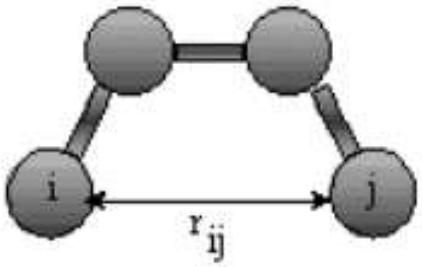
The torsion energy in MM is primarily used to correct the remaining energy terms rather than to represent a physical process. The torsional energy represents the amount of energy that must be added to r_o subtracted from the Stretching Energy + Bending Energy

+ Non-Bonded Interaction Energy term to make the total energy agree with experiment or rigorous quantum mechanical calculation for a model dihedral angle (ethane, for example, might be used as a model for an H-C-C-H bond).

- The A parameter controls the amplitude of the curve, the n parameter controls its periodicity, and phi shifts the entire curve along the rotation angle axis (tau).
- The parameters are determined from curve fitting. Unique parameters for torsional rotation are assigned to each bonded quartet of atoms based on their types (e.g. C-C-C-C, C-O-C)N, etc.).
- Notice that n reflects the type symmetry in the dihedral angle. A CH3-CH3 bond, for example, ought to repeat its energy every 120 degrees. The cis conformation of a dihedral angle is assumed to the zero torsional angle by convention. The parameter phi can be used to synchronize the torsional potential to the initial rotameric state of the molecule whose energy is being computed.

Non-Bonded Energy

- The non-bonded energy represents the pair-wise sum of all the energies of all possible interacting non-bonded atoms i and j:
- $E_{nonbonding} = S_i S_j [-A_{ij}/r_{ij}^6 + B_{ij}/r_{ij}^{12}] + S_i S_j (q_i q_j) / r_{ij}$
- The first bracketed term above represents van der Waals interactions among the atoms, while the second bracketed term represents Coloumbic electrostatic interactions.(A-repulsive, B attractive)



- The non-bonded energy accounts for repulsion, van der Waals attraction and electrostatic interactions. van der Waals attraction occurs at short range, and rapidly dies off as the interacting atoms move apart by a few angstroms.
- Repulsion occurs when the distances between interaction atoms becomes even slightly less than the sum of their contact radii.
- Repulsion is modeled by an equation that is designed to rapidly blow up at close ranges ($1/r^{12}$ dependency).

- The energy term that describes attraction/repulsion provides for a smooth transition between the two regimes.

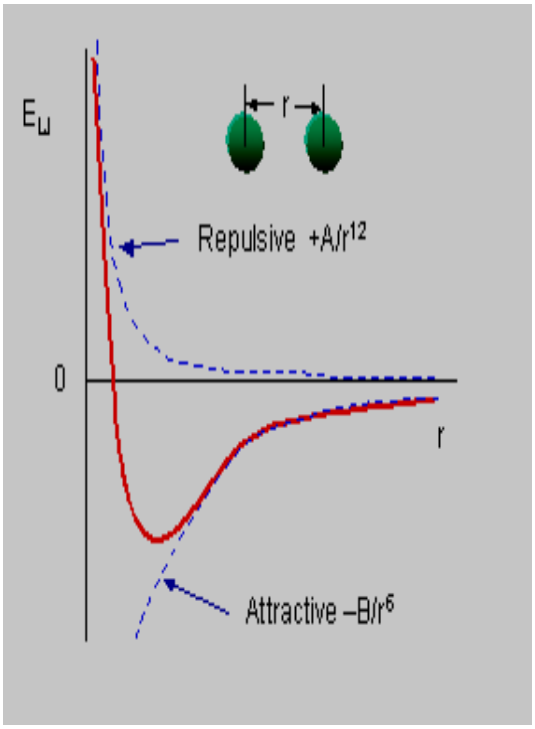


Fig5: The energy term describing attraction and repulsion providing smooth transition between the two regimes.

- The A and B parameters control the depth and position (interatomic distance) of the potential energy well for a given pair of non-bonded interacting atoms (e.g. C:C, O:C, etc.). In effect, A determined the degree of stickiness of the van der Waals attraction, and B determines the degree of hardness of the atoms (e.g. marshmallow-like, billiard ball-like, etc.).
- The A parameter can be obtained from atomic polarizability measurements, or it can be calculated quantum mechanically. The B parameter is typically derived from crystallographic data so as to reproduce observed average contact distances between different kinds of atoms in crystals of various molecules. The equation for the Lennard-Jones potential is:
- $E_{nonbonding\ (LJ)} = S_i S_j [-A_{ij}/r_{ij}^6 + B_{ij}/r_{ij}^{12}]$

Lennard-Jones Potential:

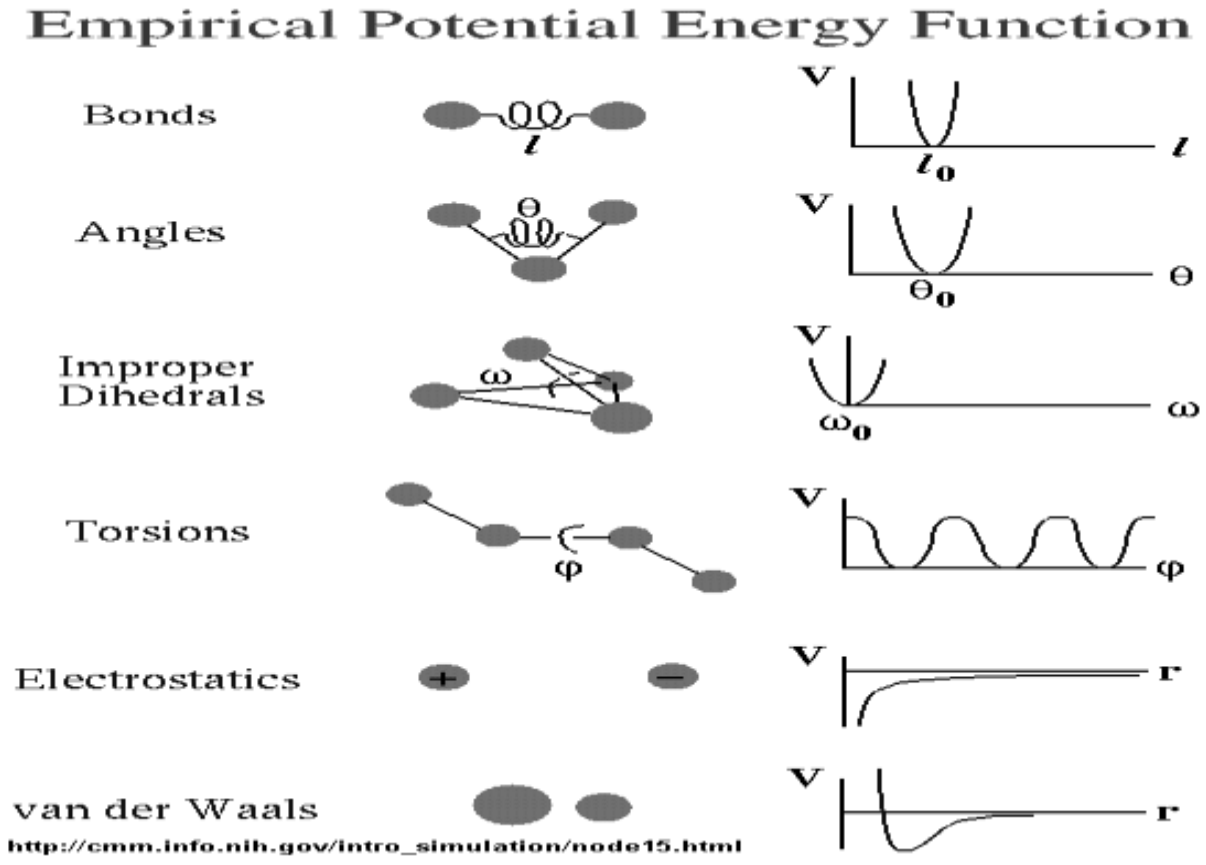
The electrostatic contribution in modeling using a Columbic potential:

- The electrostatic energy is a function of the charge on the non-bonded atoms, their interatomic distance, and a molecular dielectric expression that accounts for the

attenuation of electrostatic interaction by the environment (e.g. solvent or the molecule itself).

- Often the molecular dielectric is set to a constant value between 1.0 and 5.0. A linearly varying distance-dependent dielectric (i.e.1.r) is sometimes used to account for the increase in environmental bulk as the separation distance between interacting atoms increases.
- Partial atomic changes can be calculated for small molecules using an ab initio or semiempirical quantum technique (usually MOPAC or AMPAC). The equation for the electrostatic potential is:

- $E_{nonbonding\ (electrostatic)} = S_i S_j (q_i q_j) / r_{ij}$
- Some programs assign charges using rules or templates, especially for macromolecules. In some force-fields, the torsional potential is calibrated to a particular charge calculation method (rarely made known to the user).
- Use of a different method can invalidate the force-field consistency. Sometimes, an additional bonded interaction term, improper dihedrals, are added as illustrated below. The potential for that is given by the following equation:
- $E_{improper} = S_{angles}\ k_w (w - w_o)^2$



Fi6: The potential energy functions are illustrated in the graph above, where V is the potential energy.

MOLECULAR DYNAMICS:

- In the broadest sense, MD is concerned with molecular motion . Motion is inherent to all chemical and biological processes. Simple vibrations, like bond stretching, and angle bending, give rise to IR spectra. Chemical reactions, hormone-receptor binding, and other complex processes are associated with many kinds of intra- and intermolecular motions.
- The driving force for these processes is described by thermodynamics. The mechanism by which these processes occurs is described by kinetics. Thermodynamics dictates the energetic relationships between different chemical states, whereas the sequence or rate of events that occur as molecules transform between their various possible states is described by kinetics.
- Conformational transitions and local vibrations are the usual subjects of molecular dynamics studies.
- MD alters the intermolecular degrees of freedom in a step-wise fashion, analogous to energy minimization.
- The individual steps in energy minimization are merely directed at establishing a down-hill direction to a minimum.
- The steps in MD, on the other hand, meaningfully represent the changes in atomic position, ri, over time (i.e. velocity).
- Newton's equation ($F_i = m_i a_i$) is used in the MD formalism to simulate atomic motion.
- The rate and direction of motion (velocity) are governed by the forces that the atoms of the system exert on each other as described by Newton's equation.
- In practice, the atoms are assigned initial velocities that conform to the total kinetic energy of the system, which in turn, is dictated by the desired simulation temperature.
- This is carried out by slowly heating the system (initially at absolute zero) and then allowing the energy to equilibrate among the constituent atoms.
- The basic ingredients of MD are the calculation of the force on each atom, and form that information, the position of each atom through a specified period of time (typically on the order of picoseconds = 10^{-12} seconds).
- The force on an atom can be calculated from the change in energy between its current position and its position a small distance away.

- This can be recognized as the derivative of the energy with respect to the change in the atom's position: $-dE/dr_i = F_i$
- Energies can be calculated using either MM or quantum mechanics methods.
- MM energies are limited to applications that do not involve drastic changes in electronic structure such as bond making/breaking.
- Quantum mechanical energies can be used to study dynamic processes involving chemical changes. The latter technique is extremely novel and of limited availability.
- Knowledge of the atomic forces and masses can then be used to solve the positions of each atom along a series of extremely small time steps (on the order of femtoseconds).
- The resulting series of snapshots of structural changes over time is called a trajectory.
- The use of this method to compute trajectories can be more easily seen when Newton's equation is expressed in the following form
  $-dE/dr_i = m_i d^2r_i/dt^2$
- In practice, trajectories are not directly obtained from Newton's equation due to lack of an analytical solution.
- First the atomic accelerations are computed from the forces and masses.
- The velocites are next calculated from the accelerations based on the following relationship:

$$a_i = dv_i/dt.$$

- Lastly, the positions are calculated from the velocities: $v_i = dr_i/dt$. A trajectory between two states can be subdivided into a series of sub-states separated by a small time step, delta t (e.g. 1 fs).
- The initial atomic positions at time t are used to predict the atomic positions at time t = delta t. The positions at t = delta t are used to predict the positions at t = 2delta t, and so on.
- The leapfrog method is a common numerical approach to calculating trajectories based on Newton's equation. The method derives its name from the fact that the velocity and position information successively alternate at ½ time step intervals.
- MD has no defined point of termination other than the amount of time that can be practically covered. Unfortunately, the current ps order of magnitude limit is often not long enough to follow many kinds of state to state transformations, such as large conformational transitions in proteins.

Molecular Dynamics and *in silico* Free Energy Calculations

- MD simlulations can be used to obtained theoretical values for DG and Keq values for conformational changes, binding of small ligands and changes in protonation states for side chains.
- This process is based on the idea that the conformations sampled in *in silico* MD simulations reflect those found *in vitro* (i.e. they are part of the thermodynamically expected and conformations available to the molecules during its normal conformational shifts). This is called the Ergodic Hypothesis.
- Given the short time spans for MD simulations (limited by computer power) this hypothesis can't apply to the dynamic results unless the sample conformations are close in energy without a large activation energy barrier between them. If it is, then the following equation could apply:

  $DG^0 = -RTlnK_{eq}$

  $DG^0 = -RTlnP_2/P_1 = -RTlnf_2/f_1$ where

  P_n is the probability of being in a given state and f_n is the fraction in a given state.

- A biomolecular system can be understood as a mechanical system (i.e: atoms interacting with each other accordingly to the classic laws of physics)
- A mechanical system can be found in different states
- A state is defined by a conformation and energy
- Different conformations can have the same energy (degenerated states) (ex: gauche- and gauche+)

  To fully characterize our mechanic system we just need to know the probabilities P(s) of occurrence of all possible state s

  In a Isothermal system:

  The probability P of finding a system in a state s, P(s) is given by:

$$P(s) = \frac{e^{-E_s/RT}}{\sum_s e^{-E_s/RT}}$$

  where Es is the energy of state s, R is the gas constant and T is the temperature.

  The denominator is impossible to calculate for all states

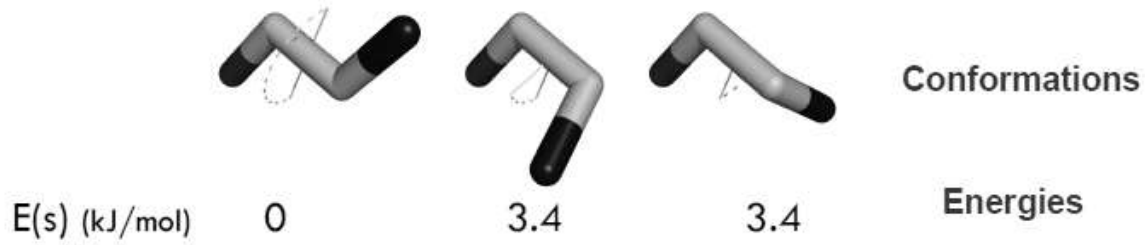
  It is impossible to calculate this probability distribution for large systems!

- System with 1 DOF, we can enumerate 360 states. One state for each torsion degree.
- States grow very fast with the number of DOF.

- A simple aminoacid has, at least, $50*10^6$ states!

Solution:

- If we cannot count all possible states of a mechanical system, the only solution is to get a representative sample of different states.
- We need to have methods to sample states:
- Generate conformations (ex: Molecular dynamics)
- Calculate the energy of each conformation (ex: Force field)



Molecular Mechanics

- Solely mechanical changes are involved (no electronically excited states or electronic rearrangements)
- Classical molecules do not exist, however …
- Statistical ensembles of a sufficient size are generated to allow for the convergence of the observables of interest and calculate thermodynamic properties
- We need an interaction function that incorporates the mechanical properties of the molecular systems
- Interaction function is generally not derived from first principles, but rather tailored and calibrated in an empirical way
- The molecular interaction function is split into a sum of functionally simple and physically meaningful energetic terms, chosen on the basis of chemical intuition
- This molecular function is known as, the Force Field

The Force Field provides a energy for a given molecular configuration:

Force field incorporates contributions from(common terms):

- Bonds
- Angles
- Dihedrals
- van der Walls
- Electrostatic interactions

We have an equation that gives the potential energy of a given molecular state.To fully characterize our molecular system, we need to generate an ensemble of different molecular configurations (states)

¨ Two common ways:

- Monte Carlo
- Molecular dynamics

- Monte Carlo Simulations
- The system jumps from state to state by means of random trial moves, that may be accepted or not
- The acceptance of a trial move depends on the resulting energy change ΔE
- ΔE ≤ 0 à accept the move
  ΔE > 0 à accept the move with probability

$$e^{-\Delta E/RT}$$

Integration of the equations of motion

Integration is broken in many small fixed time δt steps

The algorithms must:

- Conserve energy and momentum
- Be time-reversible
- Permit a long time step δt

Algorithms used:

- Verlet algorithm
- Leap-frog algorithm

Verlet algorithm

- Uses the positions and accelerations at time t, and the positions from the previous step, r(t-δt), to calculate the new positions at t+δt, r(t+δt).

$$\boldsymbol{r}(t+\delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\delta t) + \delta t^2 \boldsymbol{a}(t)$$

- Velocities are obtained by:

$$\boldsymbol{v}(t) = \left[\boldsymbol{r}(t+\delta t) - \boldsymbol{r}(t-\delta t)\right] / 2\delta t$$

- Initialization
- How to get position at "previous time step" when starting out?
- At t=0:

$$\boldsymbol{r}(t_0+\delta t) = 2\boldsymbol{r}(t_0) - \boldsymbol{r}(t_0-\delta t) + \delta t^2 \boldsymbol{a}(t_0)$$

$$\mathbf{r}(t_0-\delta t) = \mathbf{r}(t_0) - \mathbf{v}(t_0)\delta t$$

Verlet algorithm

Advantages

- Very simple to implement
- Intuitive, straight forward application of Newton's equation of motion.

Disadvantages

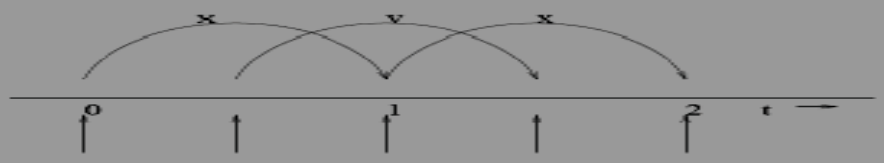
- Addition of large and small numbers leading to numerical imprecision

Leapfrog algorithm

- Positions and velocities are not synchronized

$$r(t+\delta t) = r(t) + \delta t v\left(t + \frac{1}{2}\delta t\right)$$

$$v\left(1+\frac{1}{2}\delta t\right) = v\left(1-\frac{1}{2}\delta t\right) + \delta t a(t)$$



- Initialization
- How to get velocity at "previous time step" when starting out?
- At t=0:

$$v\left(1+\frac{1}{2}\delta t\right) = v\left(1-\frac{1}{2}\delta t\right) + \delta t a(t_0)$$

$$\mathbf{v}(t_0 - \delta t) = \mathbf{v}(t_0) - \frac{1}{m}\mathbf{F}(t_0)\frac{1}{2}\delta t$$

Leapfrog algorithm

Advantages

- Eliminates the problem of adding small and large numbers
- High numerical precision and stability

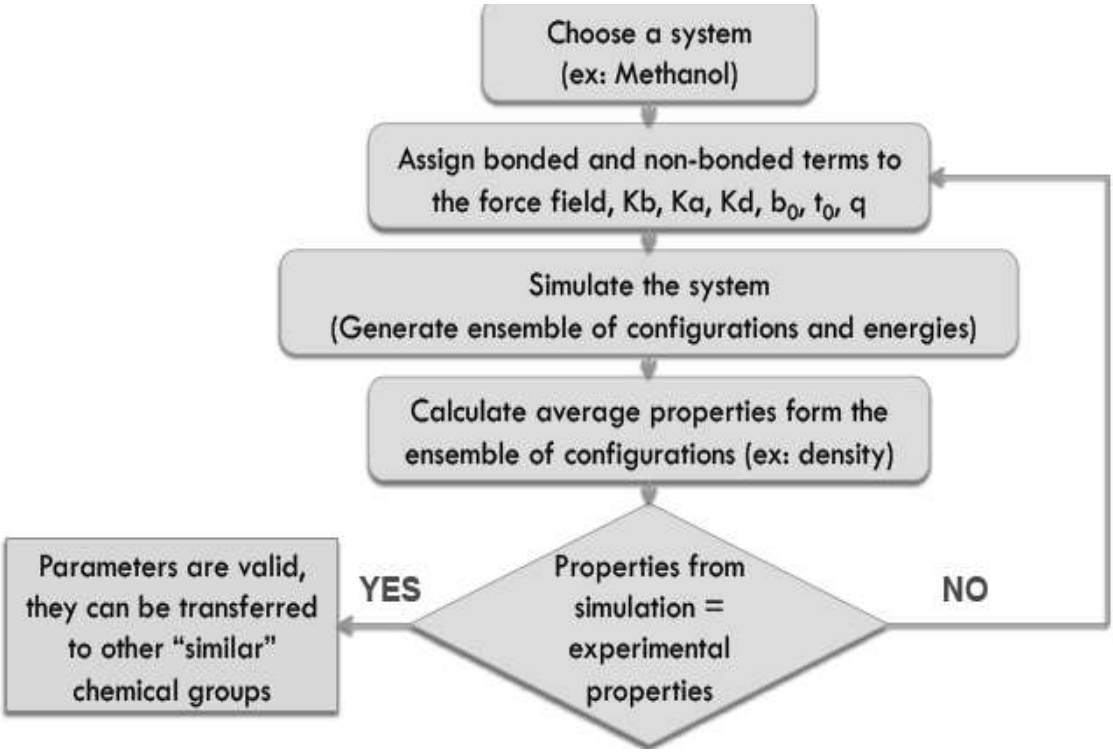
Disadvantages

- Velocities are not synchronized with positions. Extra work to synchronize. Initial velocities are randomly assigned
- Velocities are given by a Maxwell-Boltzmann distribution

$$prob(v_x) \propto \exp\left(-\frac{1}{2}mv_x^2 / kT\right)$$

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

Force Field parameter calibration & simulation check



Molecular Dynamics Simulations

- GROMOS force field development (GROningen MOlecular Simulation)
- Reproduce the free enthalpies of hydration and apolar solvation for a range of compounds.
- Thermodynamic properties of pure liquids of a range of small polar molecules
- The solvation free enthalpies of amino acid analogs in cyclohexane (53A5),Hydration free enthalpies in water (53A6)
- Relative free enthalpy of solvation between polar and apolar environments is a key property in many biomolecular processes of interest, such as protein folding, biomolecular association, membrane formation, and transport across the membranes.

Empirical force fields for biological macromolecules

- CHARMM: Chemistry at HARvard Macromolecular Mechanics
- AMBER: Assisted Model Building with Energy Refinement
- OPLS:Optimized Potentials for Liquid Simulations

BMS COLLEGE OF ENGINEERING, BANGALORE-19
DEPARTMENT OF BIOTECHNOLOGY
Faculty : Dr. Savithri Bhat

Molecular Dynamics Simulations

- **Molecular dynamics software**
  - DO NOT CONFUSE SIMULATION SOFTWARE WITH FORCE FIELDS!
  - Most force field developers also have their simulation software

Force Field	Software name
CHARMM	CHARMM
AMBER	AMBER
GROMOS	BIOMOS
OPLS	BOSS (Biochemical and Organic Simulation System)

References for Unit4, 7, & 8:

- Biophysical Chemistry by Cantor R. and Schimmel P.R, W. H. Freeman.
- Physical Biochemistry by David Freifelder, W H Freeman and Company.
- Proteins Structure – A Practical Approach by Creighton, Oxford University Press
- Molecular Biology of the Cell by Alberts et al., Garland Publishing
- Biochemistry by Voet & Voet, Wiley New York
- Principles of Biochemistry by Albert Lehninger, CBS publishers