



Computational Chemistry

1

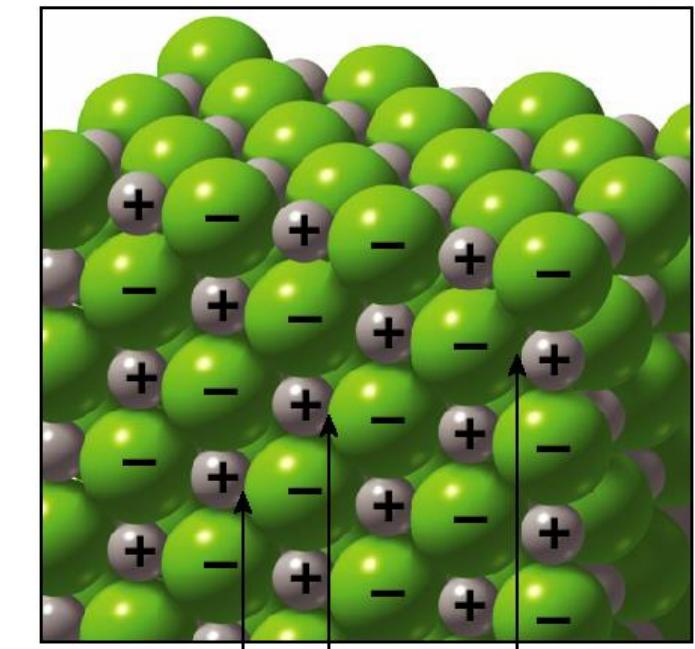
Dr. Swarna M Patra
Assistant Professor
Dept. of Chemistry
RV College of Engineering®
Bengaluru – 560059
Email: swarnamp@rvce.edu.in

UNIT-II**Computational Chemistry****08 Hrs**

Scope, cost and efficiency of computational modeling. Stabilizing interactions: Bonded and non-bonded interactions. Molecular topology, topological matrix representation, topological indices, QSAR/QSPC concept for insilico prediction of properties. 3D co-ordinate generation for small molecules, geometry optimization.

- Intermolecular forces are interactions that exist between molecules. Functional groups determine the type and strength of these interactions.
- There are several types of intermolecular interactions.
- Ionic compounds contain oppositely charged particles held together by extremely strong electrostatic interactions. These ionic interactions are much stronger than the intermolecular forces present between covalent molecules.
- Covalent compounds are composed of discrete molecules.
- The nature of the forces between molecules depends on the functional group present. There are three different types of interactions, shown below in order of increasing strength:
 - van der Waals forces
 - dipole-dipole interactions
 - hydrogen bonding

Ion-ion interactions



strong electrostatic interaction

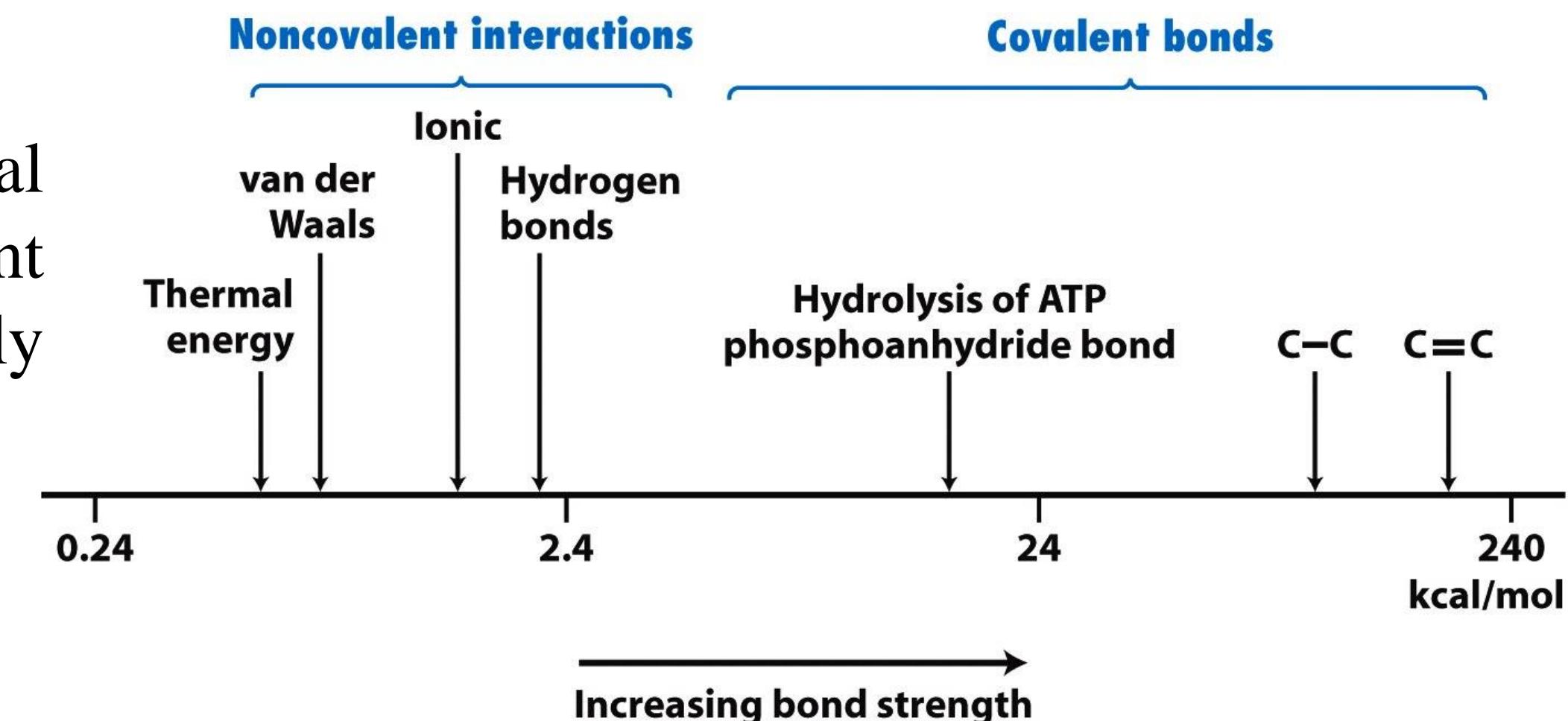
intramolecular IONIC
COVALENT

intermolecular Hydrogen
Dipole-Dipole
Induced Dipole

Strongest

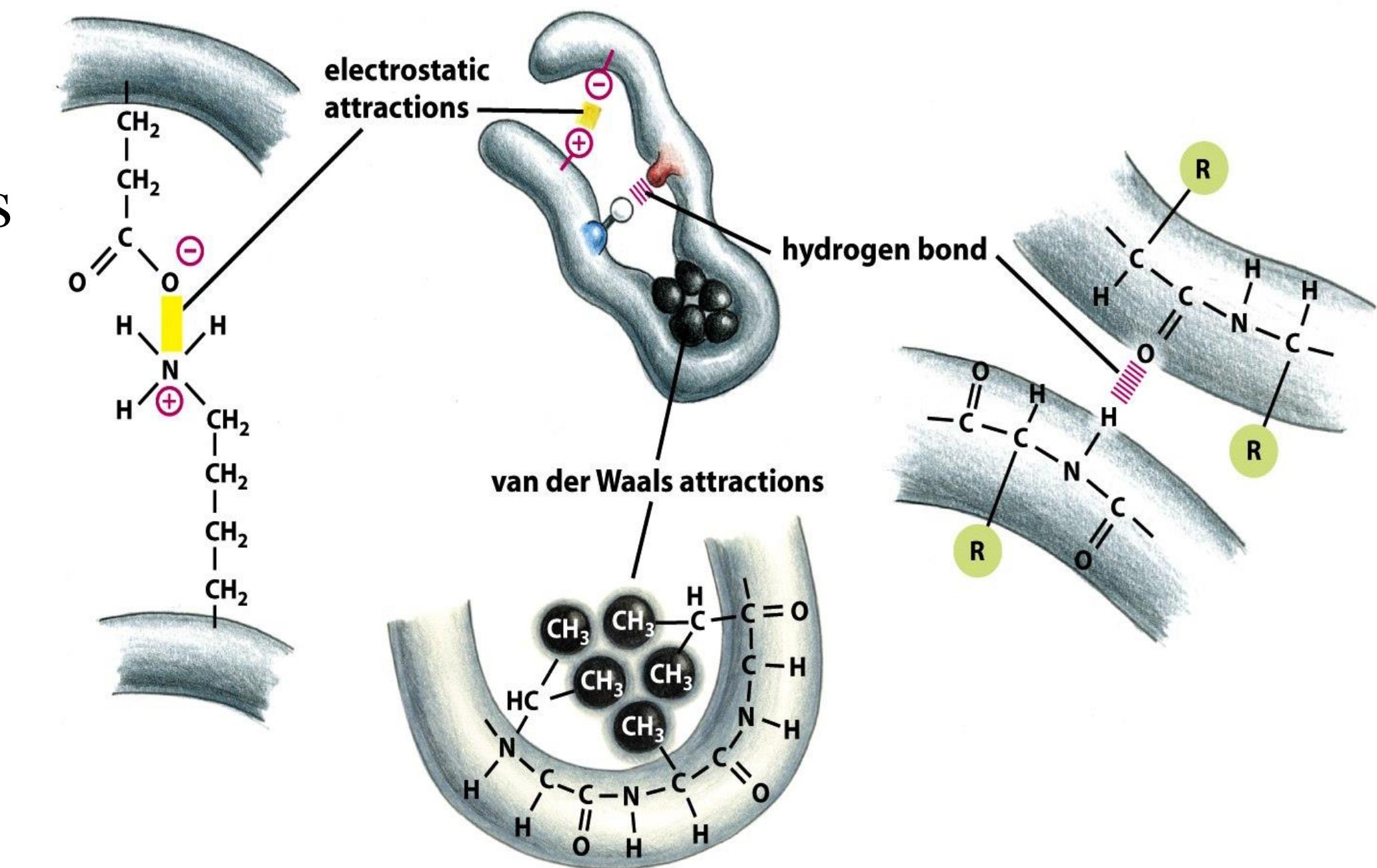
Weakest

Noncovalent interactions are weak electrical bonds between molecules. Noncovalent interactions (1-5 kcal/mol) are typically ~100-fold weaker than covalent bonds



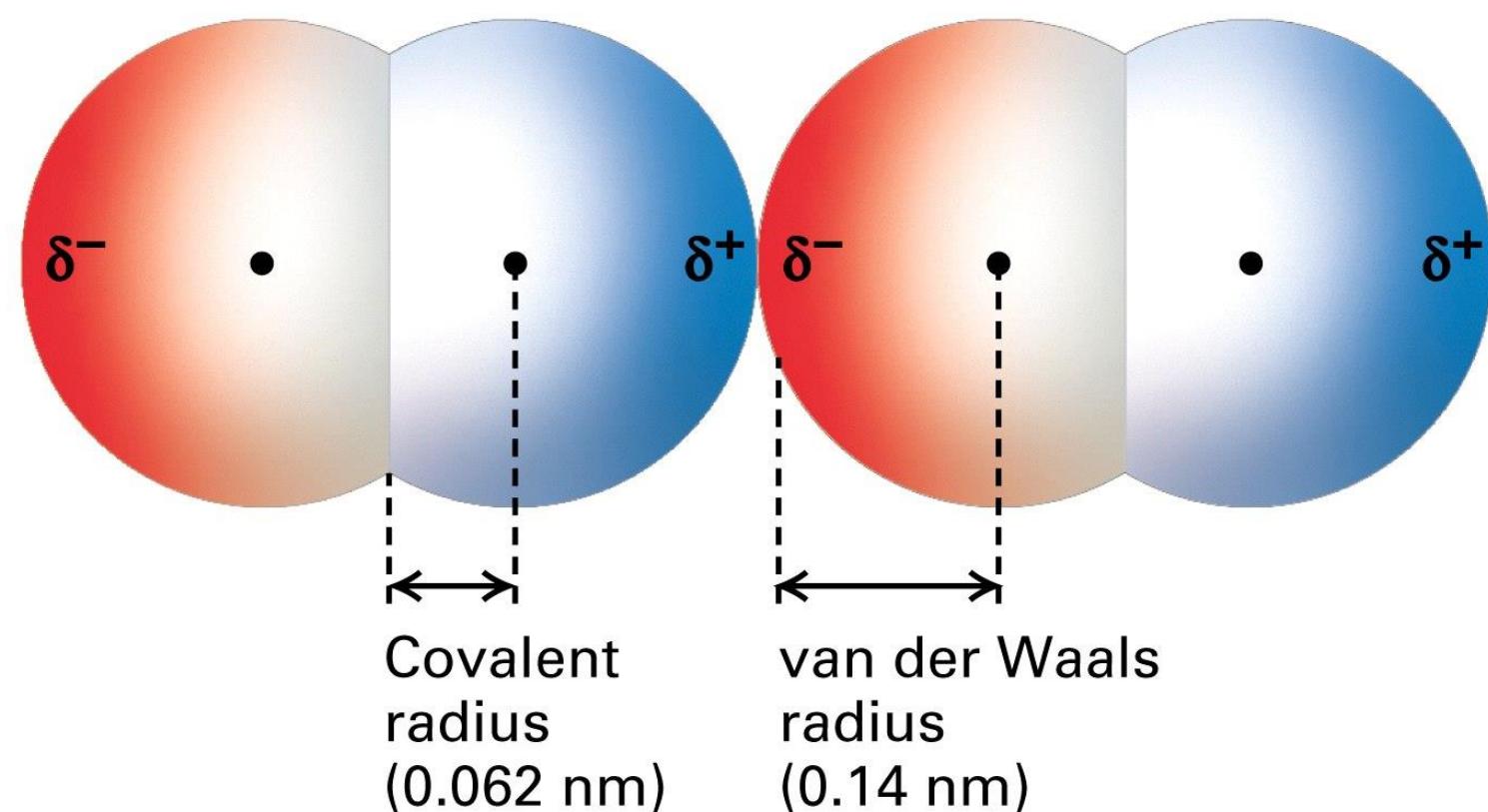
Noncovalent interactions determine protein structure

- Amino acids are connected by *covalent* bonds called peptide bonds.
- Four types of *noncovalent* interactions between amino acids affect protein structure:
 - van der Waals interactions
 - Electrostatic interactions (salt bridges)
 - Hydrogen bonds
 - Hydrophobic forces



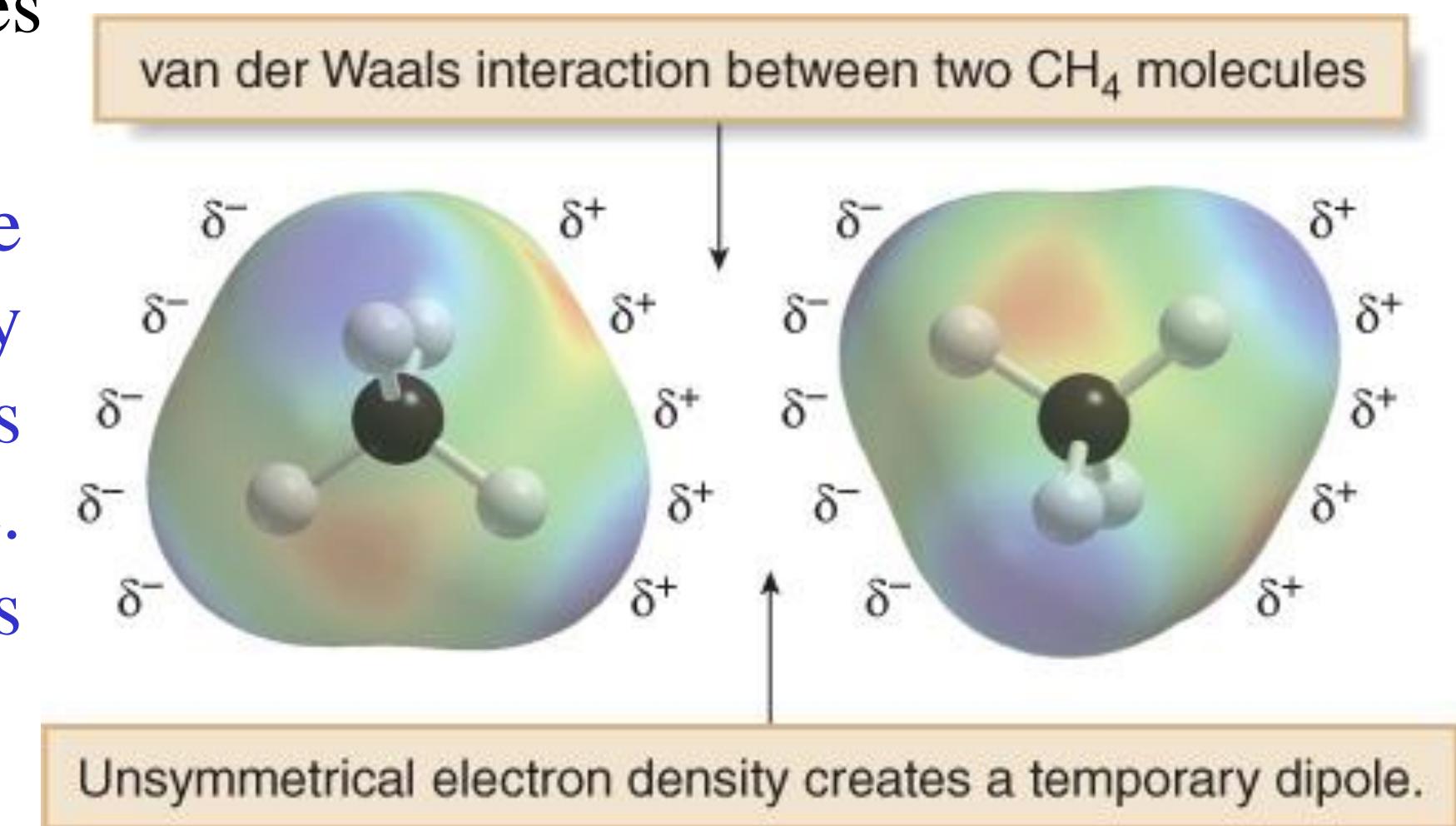
Force two atoms together and they will push back. When two atoms are close together, the occupied orbitals on the atom surfaces overlap, causing electrostatic repulsion between surface electrons. This repulsive force between atoms acts over a very short range, but is very large when distances are short.

van der Waals interactions are bonds between fluctuating, induced dipoles within the electron clouds of interacting molecules. These bonds can occur between nonpolar or polar molecules. van der Waals bonds are extremely dependent on the distance of separation between molecules, and are significant only when the electron clouds of the molecules are just touching. van der Waals interactions are demonstrated for two O₂ molecules and the covalent and van der Waals radii are shown.

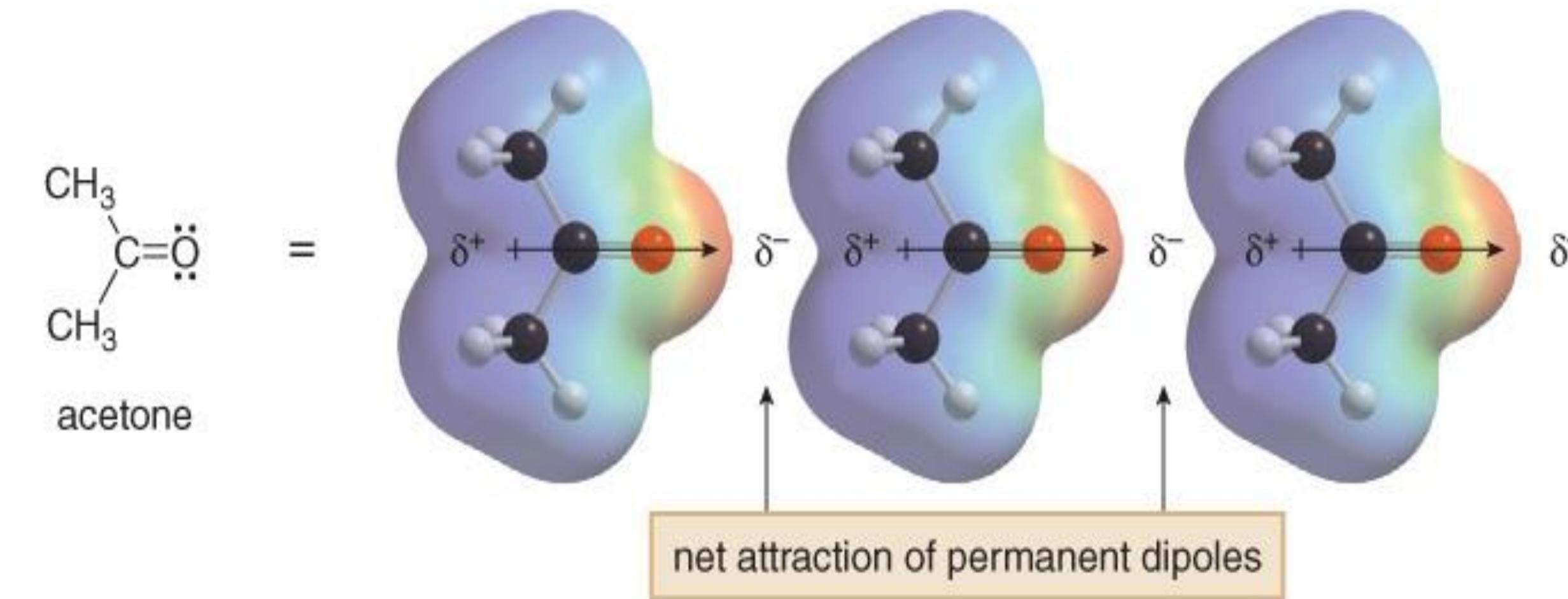


- Van der Waals forces are also known as London forces.
- They are weak interactions caused by momentary changes in electron density in a molecule.
- They are the only attractive forces present in nonpolar compounds.
- All compounds exhibit van der Waals forces.
- The surface area of a molecule determines the strength of the van der Waals interactions between molecules. The larger the surface area, the larger the attractive force between two molecules, and the stronger the intermolecular forces

Even though CH_4 has no net dipole, at any one instant its electron density may not be completely symmetrical, resulting in a temporary dipole. This can induce a temporary dipole in another molecule. The weak interaction of these temporary dipoles constitutes van der Waals forces.



Dipole—dipole interactions are the attractive forces between the permanent dipoles of two polar molecules.



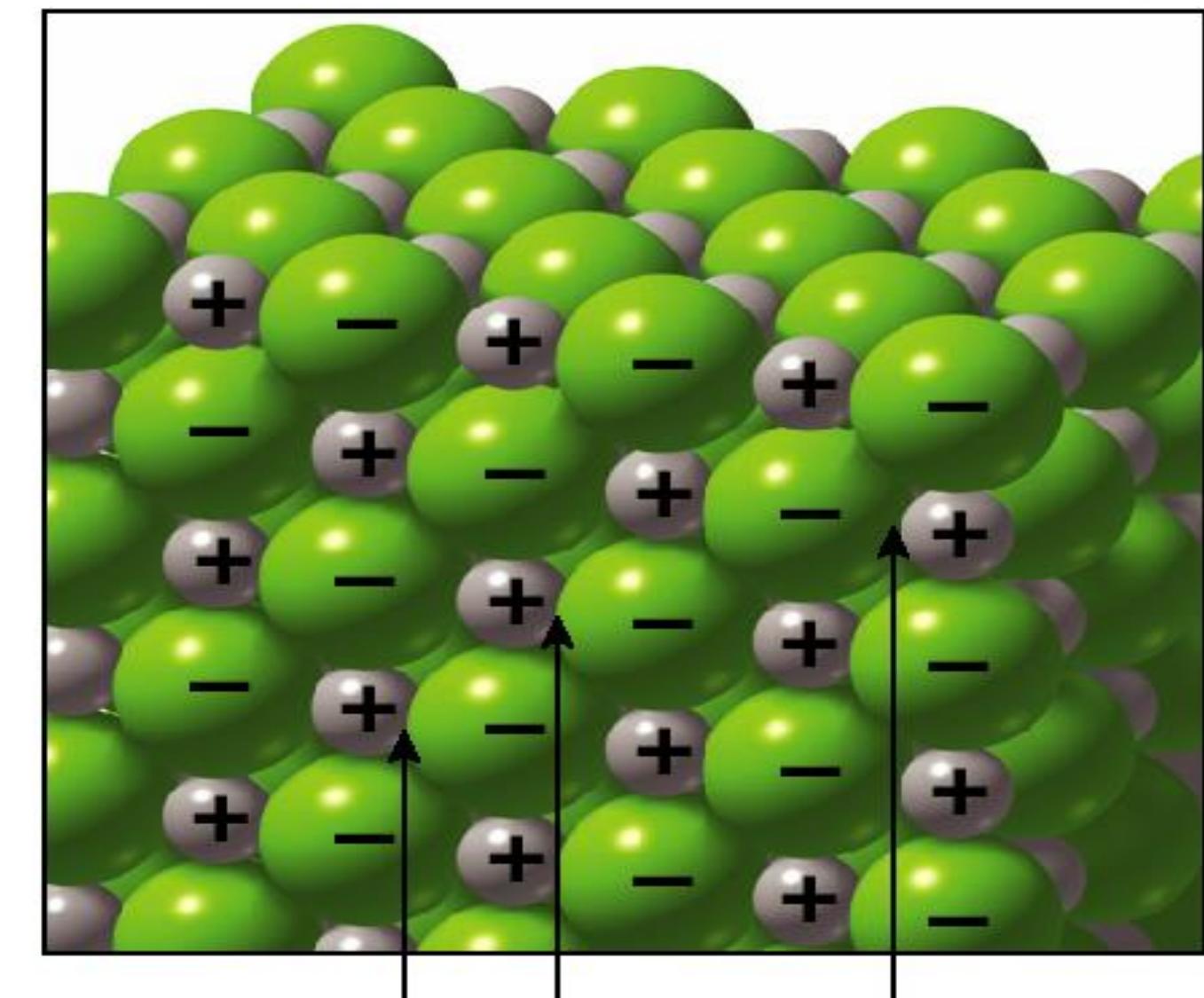
- Consider acetone. The dipoles in adjacent molecules align so that the partial positive and partial negative charges are in close proximity. These attractive forces caused by permanent dipoles are much stronger than weak van der Waals forces.

- Intermolecular forces are interactions that exist between molecules. Functional groups determine the type and strength of these interactions.
- There are several types of intermolecular interactions.
- Ionic interactions occur between cations and anions.
- These bonds are non-directional
- These ionic interactions are much stronger than the intermolecular forces present between covalent molecules.
- Strength depends on the distance of separation (r) according to $1/r^2$. Strength also depends on the medium (dielectric constant), and is less in polar than nonpolar solvents.

9

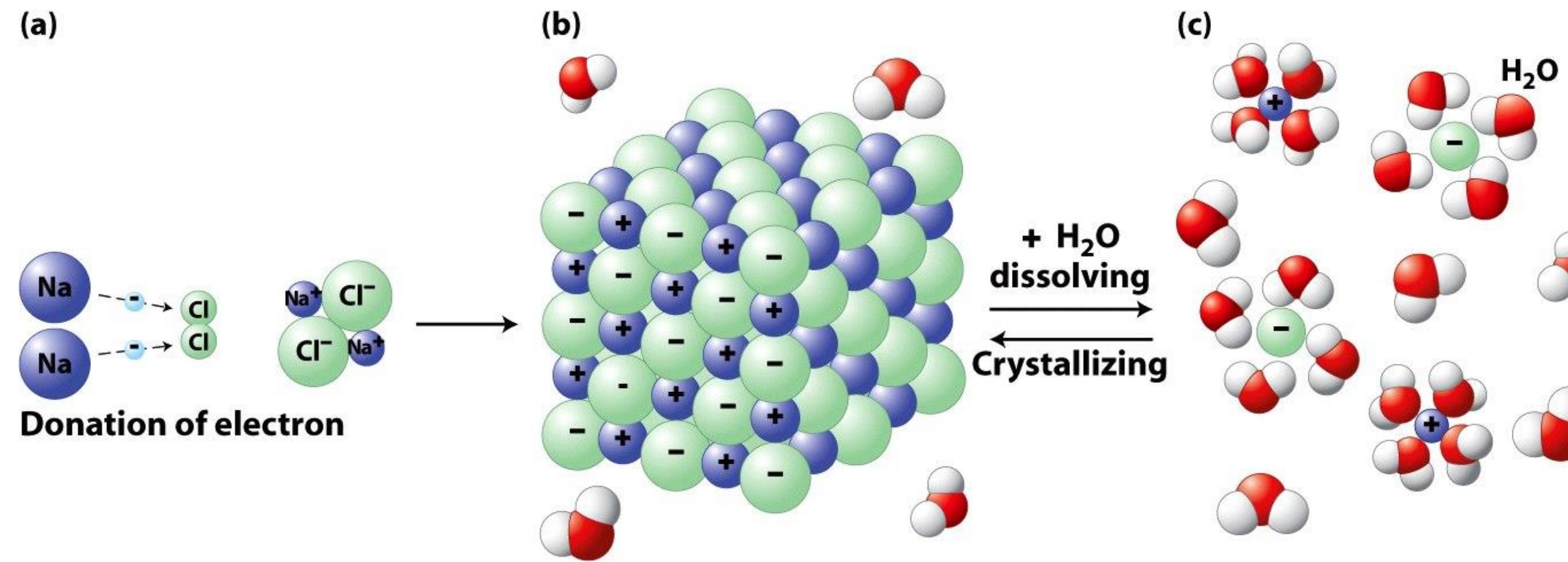
Ionic compounds such as NaCl are readily dissolved in water. Solvation spheres of water molecules surround ions in solutions. Water molecules orient so that the negative ends of their dipoles contact cations and the positive ends contact anions in solution.

Ion–ion interactions



strong electrostatic interaction

Ionic compounds such as NaCl are readily dissolved in water. Solvation spheres of water molecules surround ions in solutions. Water molecules orient so that the negative ends of their dipoles contact cations and the positive ends contact anions in solution.

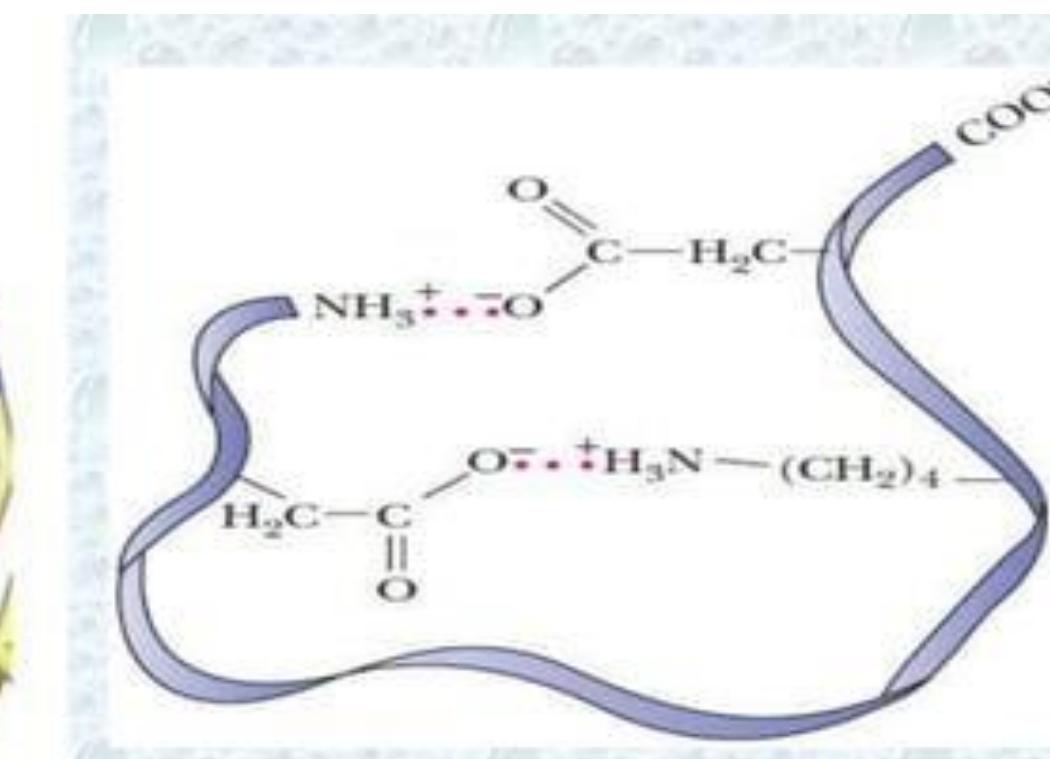
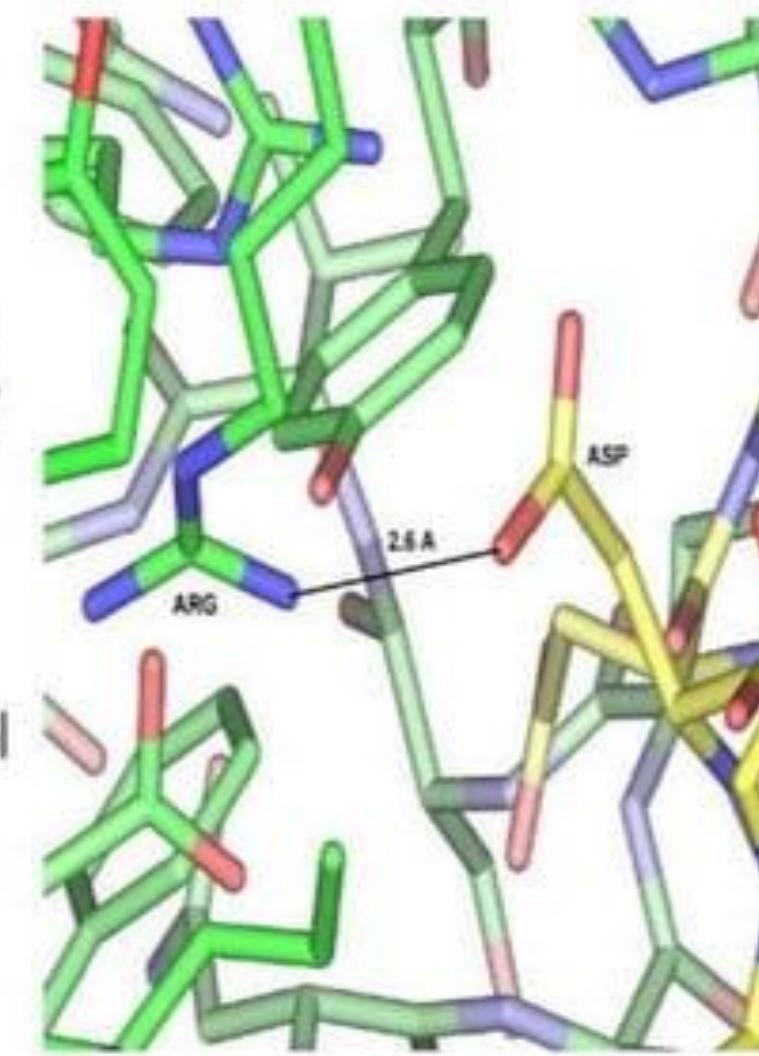


Favorable electrostatic interactions cause the vapor pressure of sodium chloride and other salts to be very low. If you leave crystals of table salt (NaCl; Na⁺=cation, Cl⁻=anion) on a hot pan, how long does it take before they vaporize and sublime away? A very very long time; electrostatic interactions are very very strong. The electrostatic interactions within a sodium chloride crystal are called ionic bonds. But when a single cation and a single anion are close together, within a protein, or within a folded RNA, those interactions are considered to be non-covalent electrostatic interactions

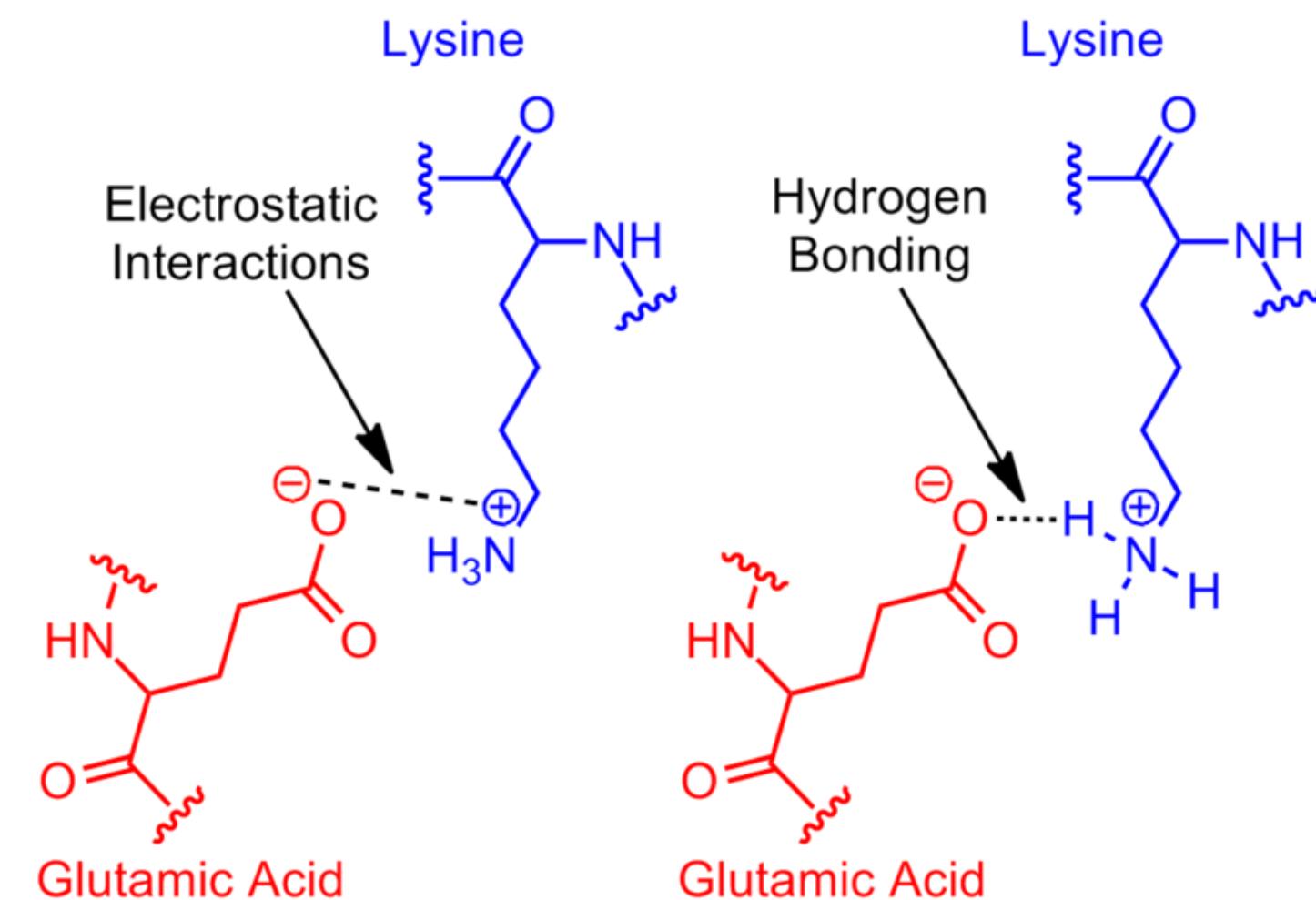
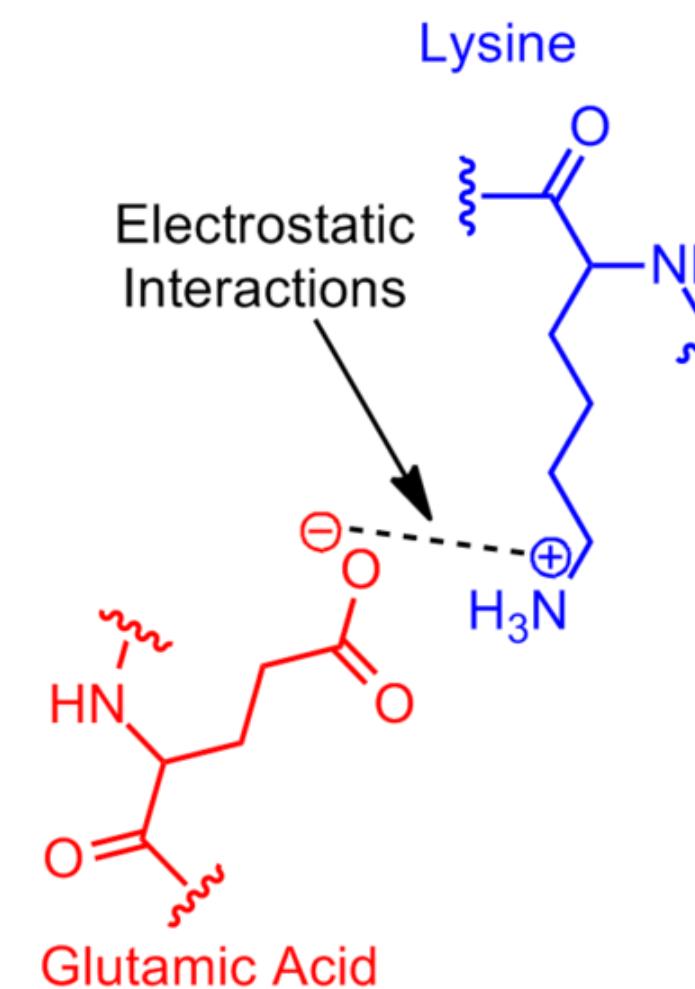
A salt bridge is a *non-covalent interaction between two ionized sites*. It has two components: a hydrogen bond and an electrostatic interaction. Salt bridges in proteins are bonds between oppositely charged residues that are sufficiently close to each other to experience electrostatic attraction.

Salt Bridges

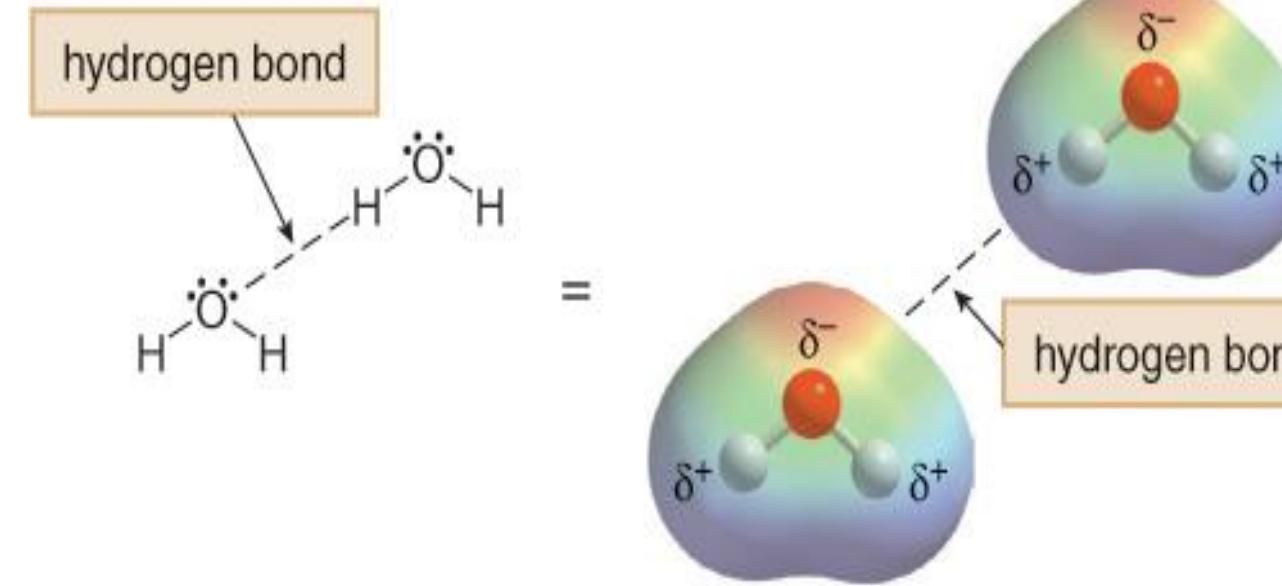
- Salt bridges are electrostatic bonds between oppositely charged groups
- The strength is usually 4-7 kcal/mol



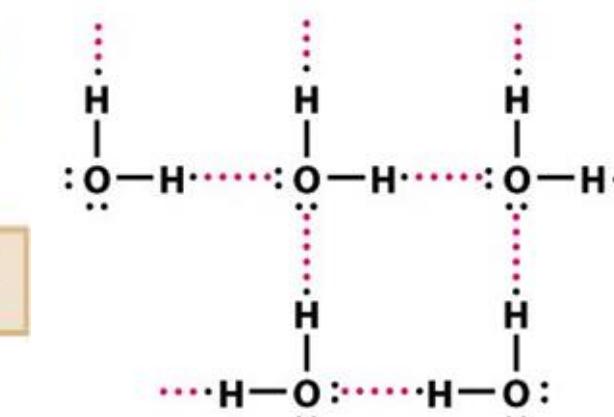
- Ions on R groups form salt bridges through ionic bonds.
- NH^3+ and COO^- areas of the protein attract and form ionic bonds.



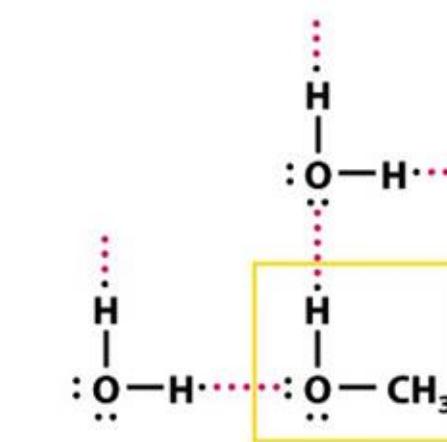
Hydrogen bonding typically occurs when a hydrogen atom bonded to O, N, or F, is electrostatically attracted to a lone pair of electrons on an O, N, or F atom in another molecule.



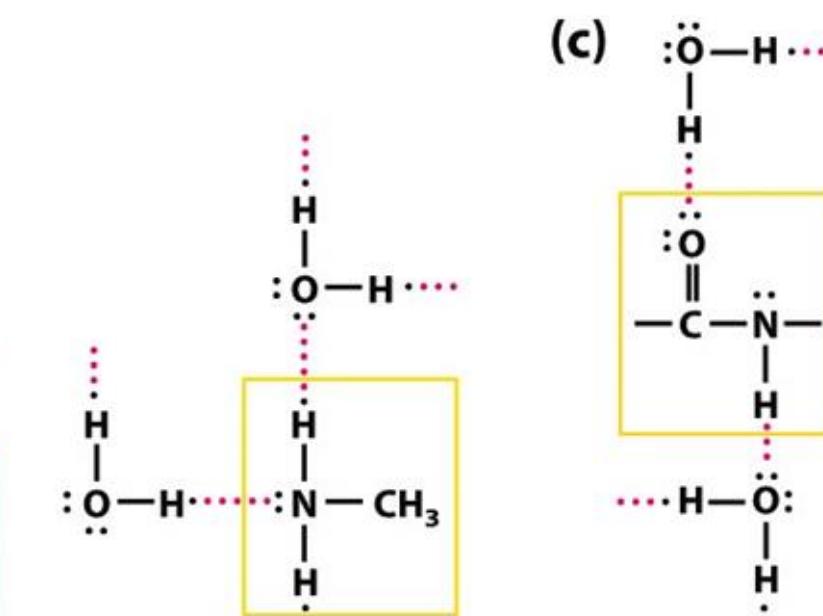
(a)



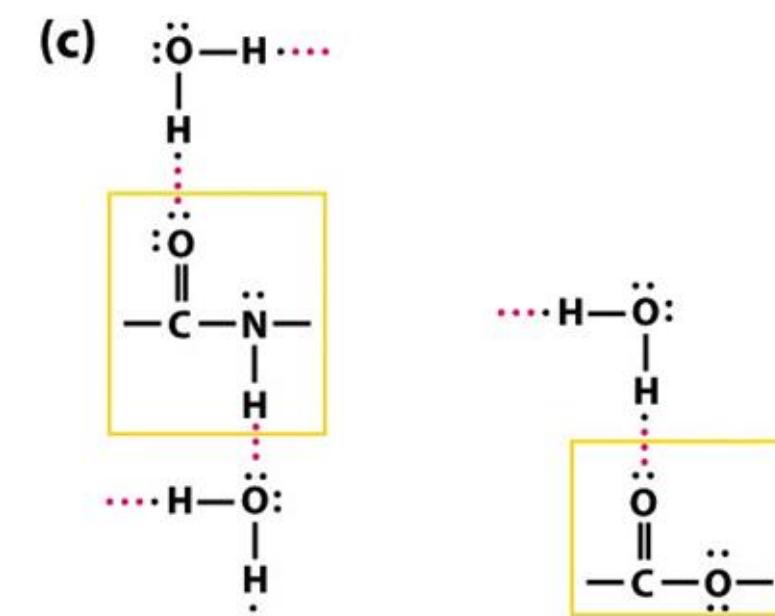
(b)



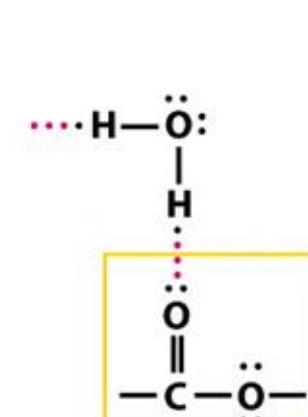
Water-water



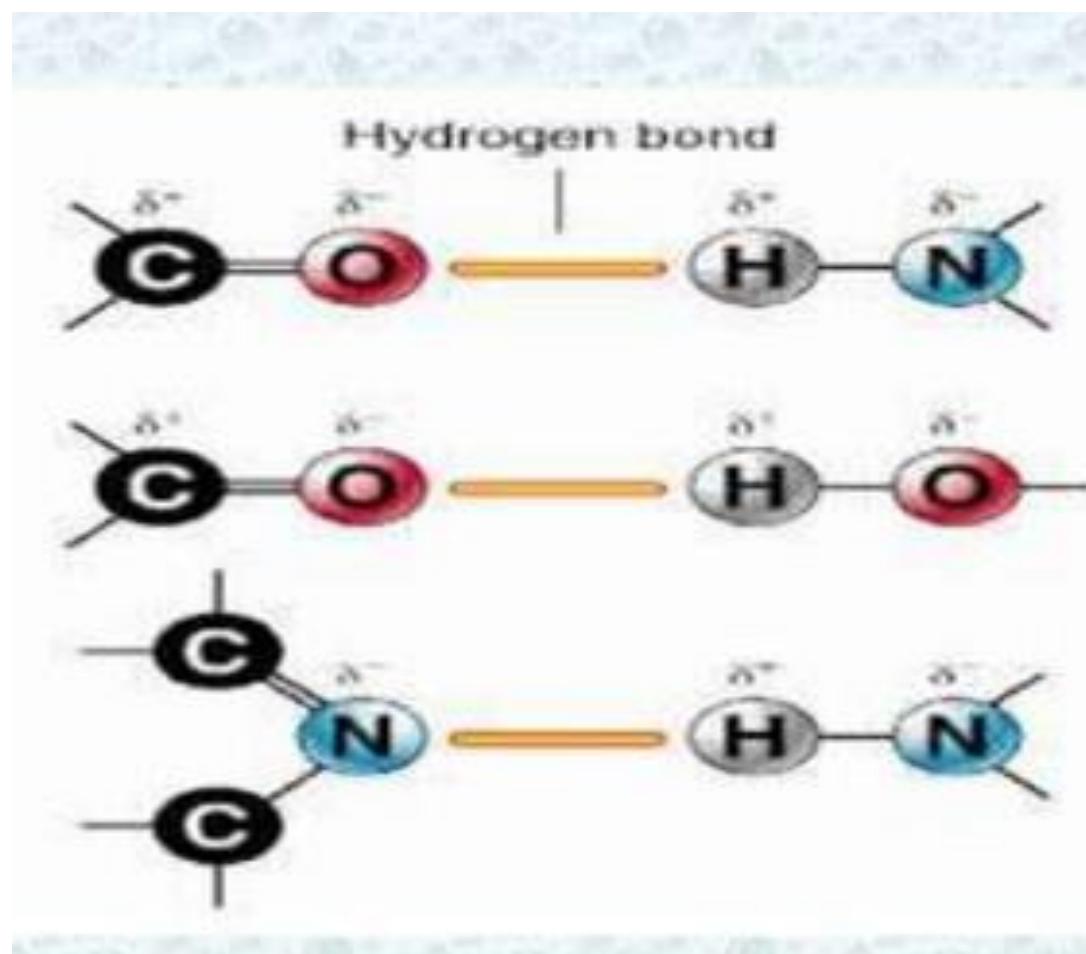
Amine-water



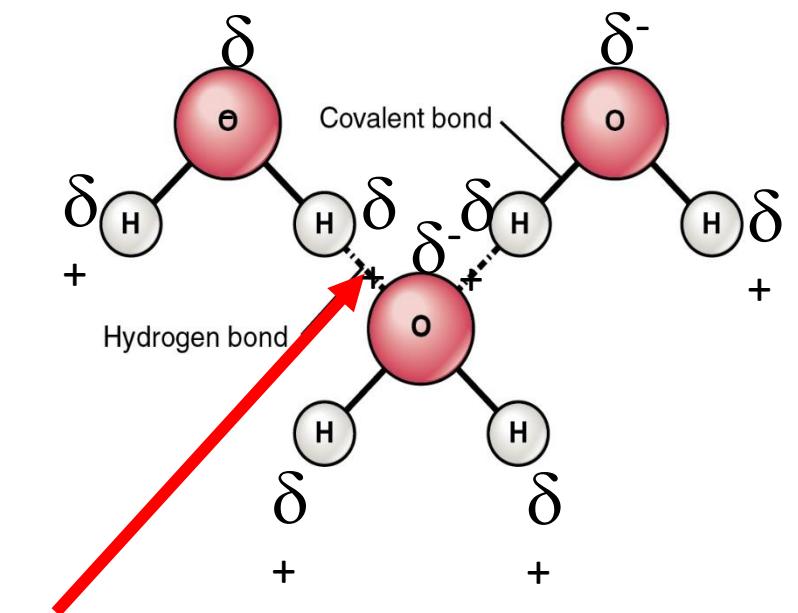
Peptide group-water



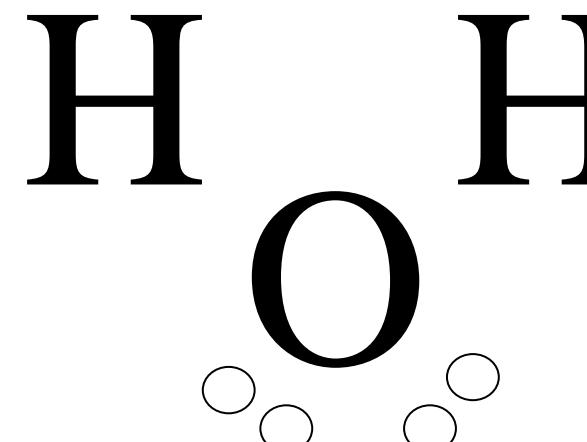
Ester group-water



- ❖ H bonds are weak which allows to be broken and reformed easily.
- ❖ Allows structural change and produces 'functional' molecules

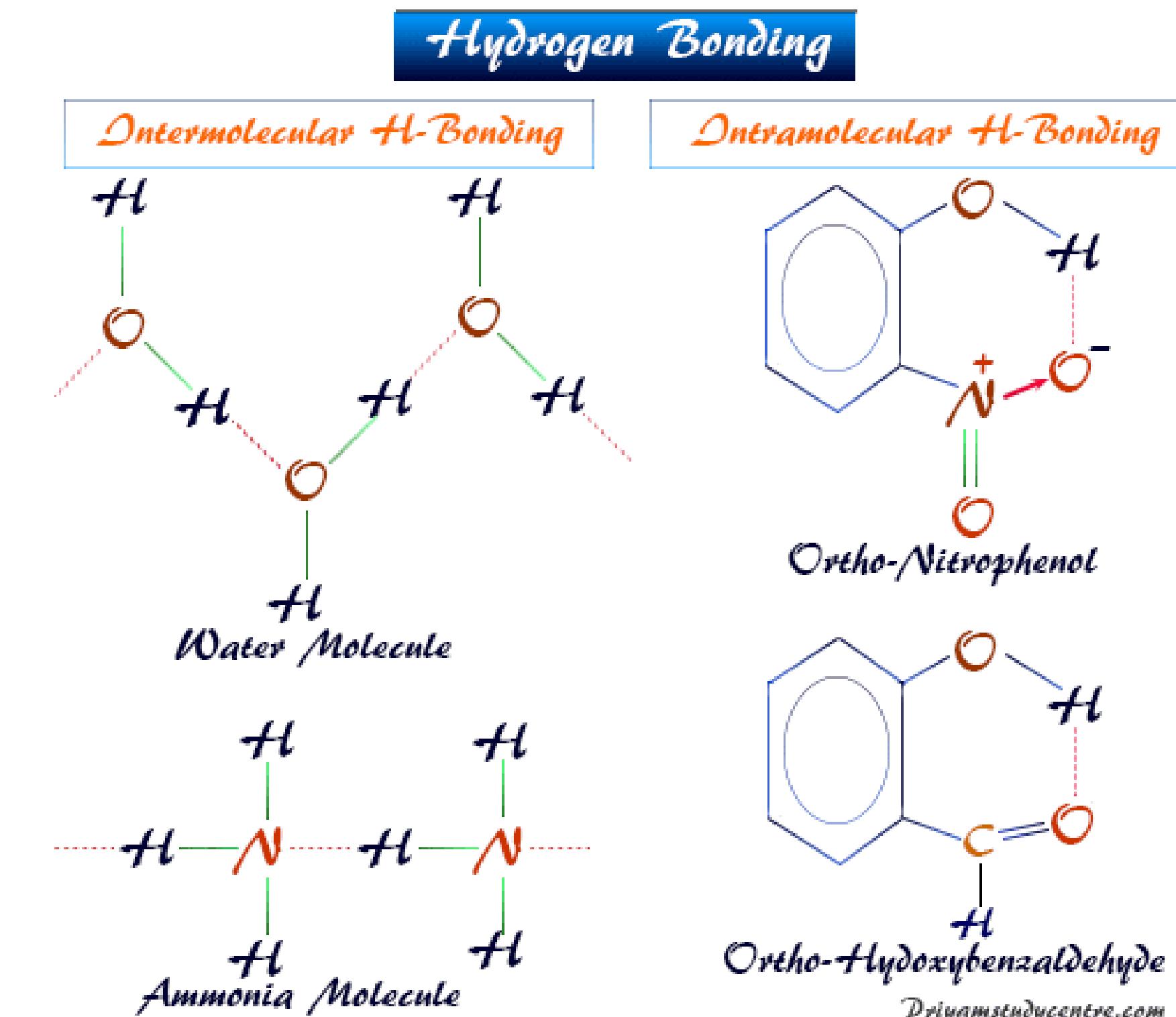
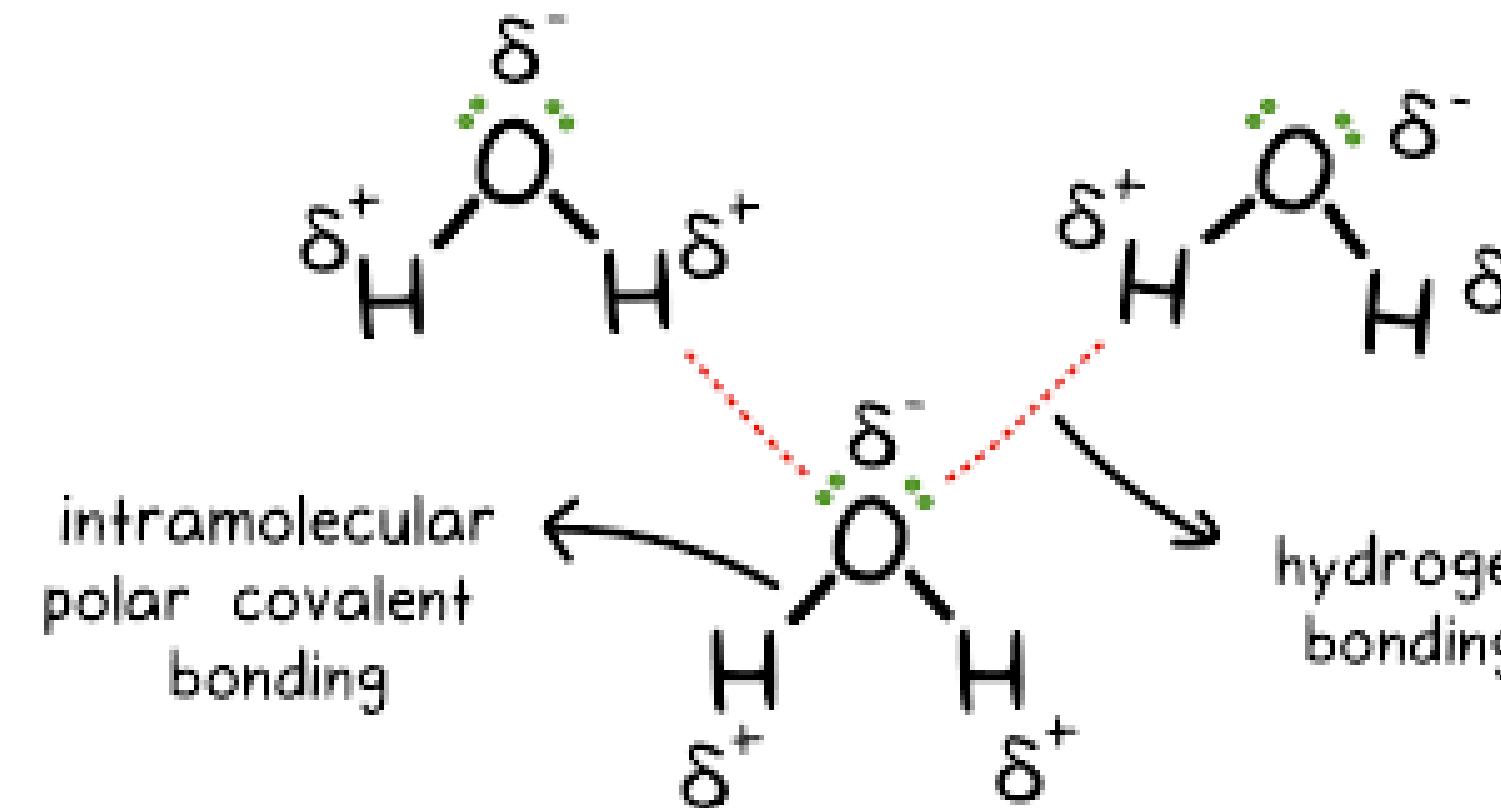


Hydrogen “bond”



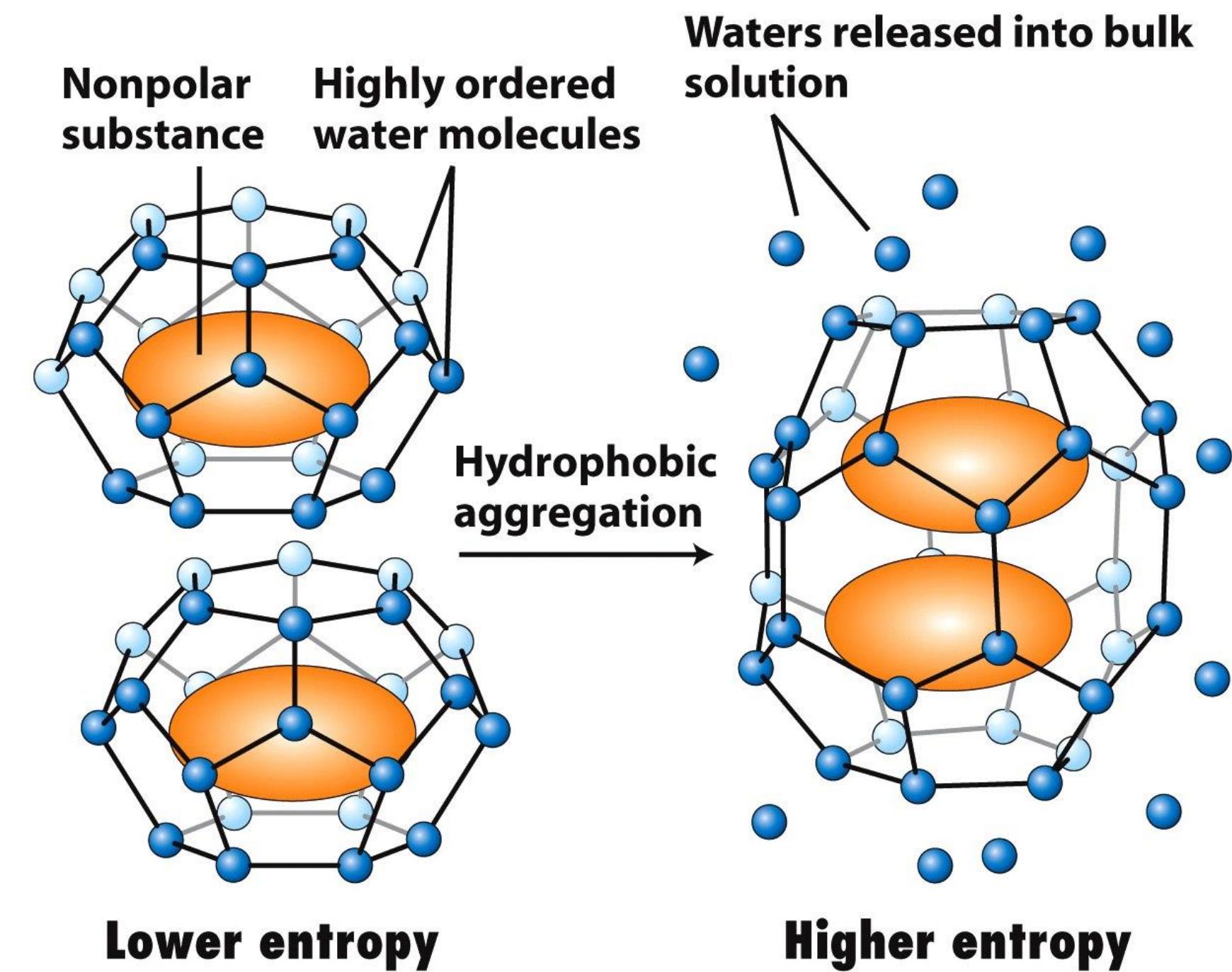
More on *intermolecular forces Hydrogen “Bonding”*

STRONG *intermolecular force* Like magnets
Occurs *ONLY* between H of one molecule and N, O, F of another molecule



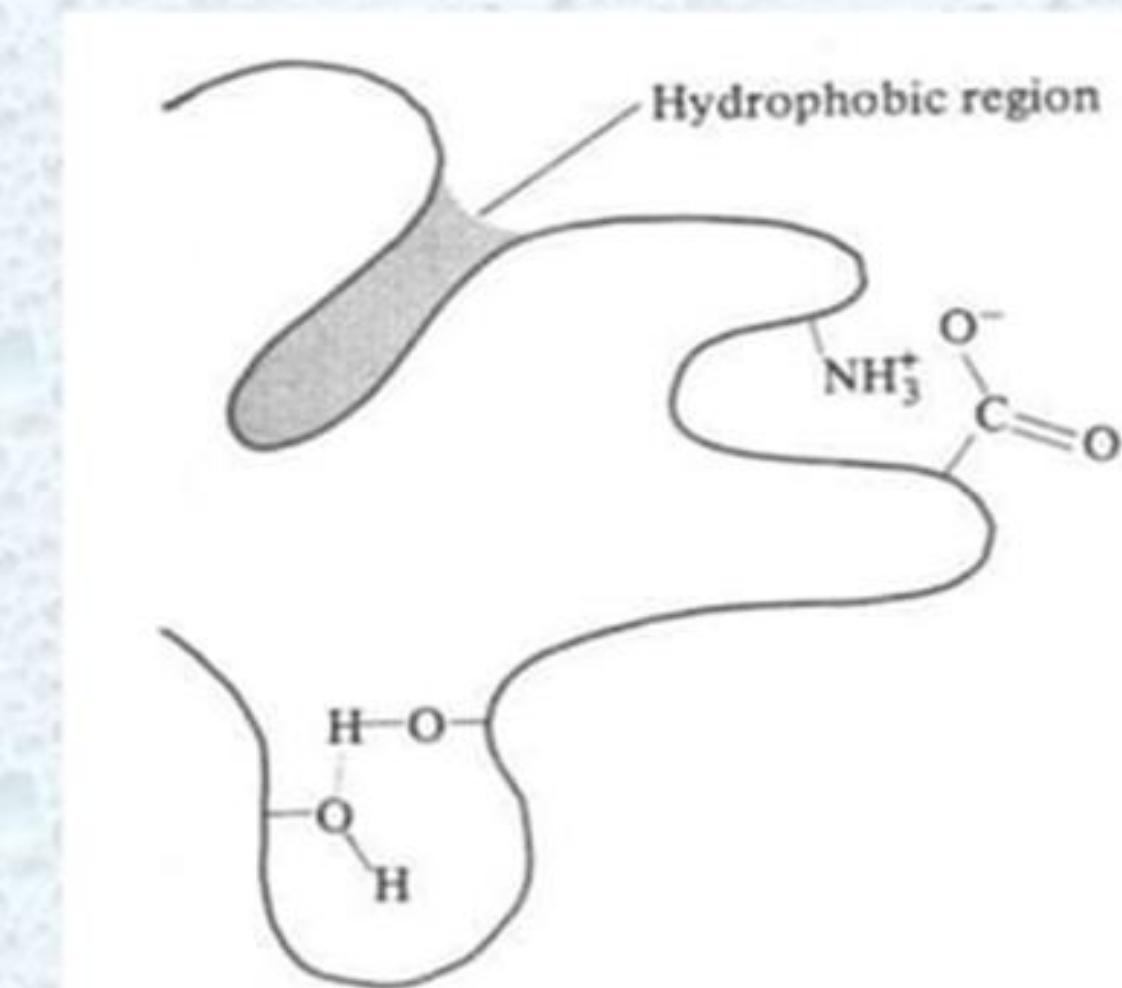
The hydrophobic effect refers to the entropy-driven aggregation of nonpolar molecules in aqueous solution that occurs to minimize the ordering of water molecules with which they are in contact. This is not an attractive force, but rather a thermodynamically driven process.

The cage-like structures formed by water molecules surrounding a nonpolar solute. The hydrophobic effect drives the formation of membranes and contributes to the folding of proteins and the formation of double helical DNA.



Tertiary structure - hydrophobic forces

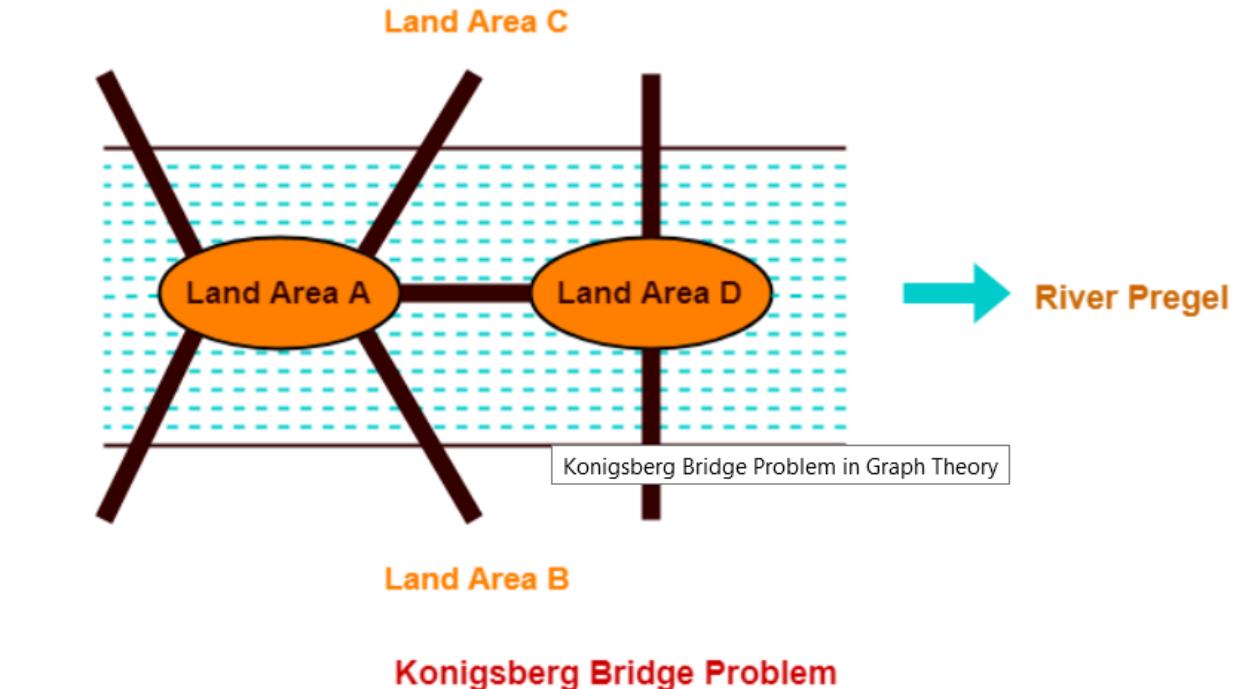
- Close attraction of non-polar R groups through dispersion forces.
- They are non attractive interactions, but results from the inability of water to form hydrogen bonds with certain side chains.
- Very weak but collective interactions over large area stabilize structure.
- Repel polar and charged molecules/particles.





Molecular Topology

Seven Bridges of Königsberg



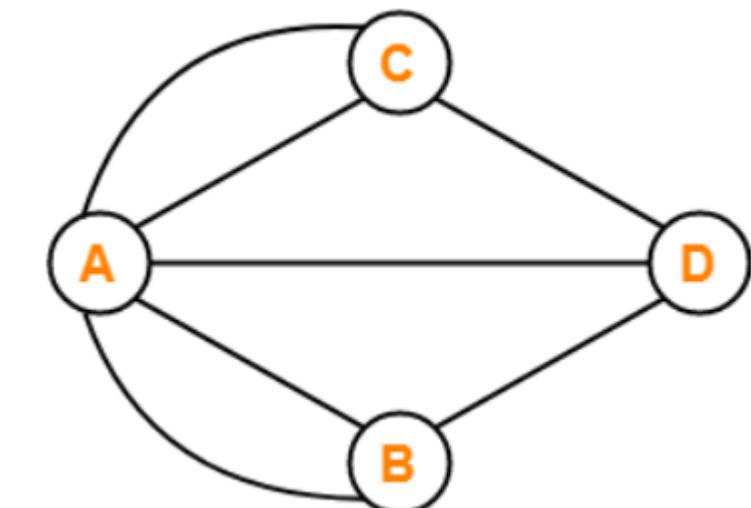
“Starting from any of the four land areas A, B, C, D, is it possible to cross each of the seven bridges exactly once and come back to the starting point without swimming across the river?”

In 1735, A Swiss Mathematician Leonhard Euler solved this problem.

- He provided a solution to the problem and finally concluded that such a walk is not possible.

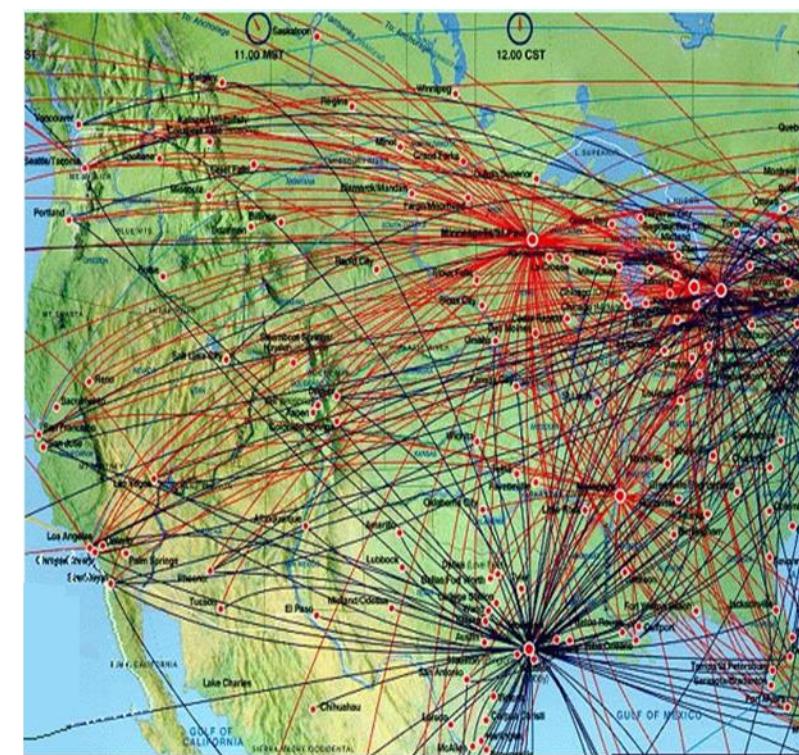
Euler found that only those networks are traversable that have either-

- No odd vertices (then any vertex may be the beginning and the same vertex will also be the ending point)
- Or exactly two odd vertices (then one odd vertex will be the starting point and other odd vertex will be the ending point)

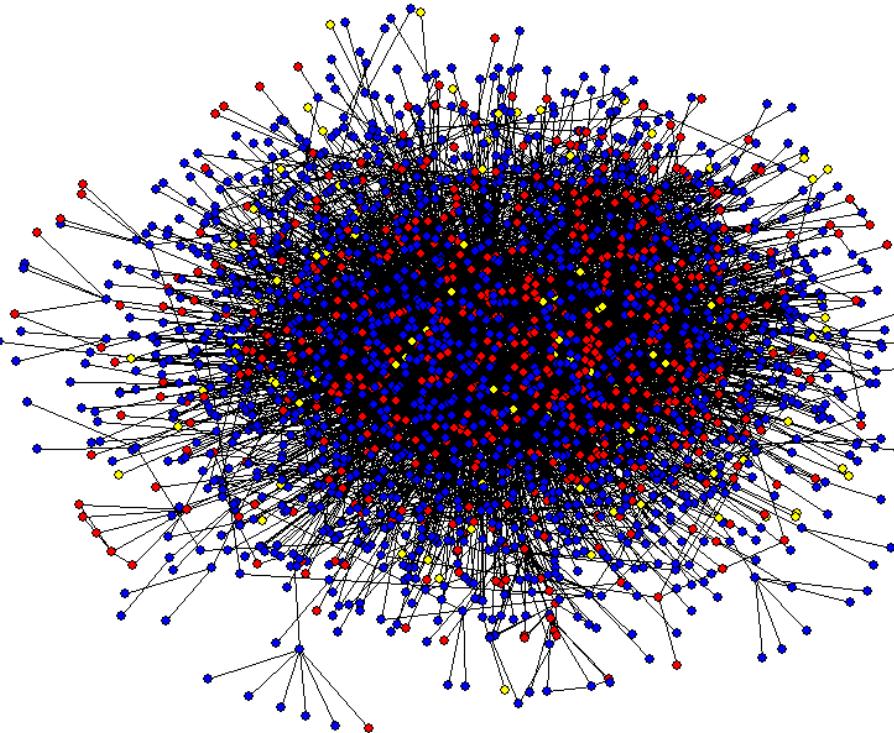


What is a Network?

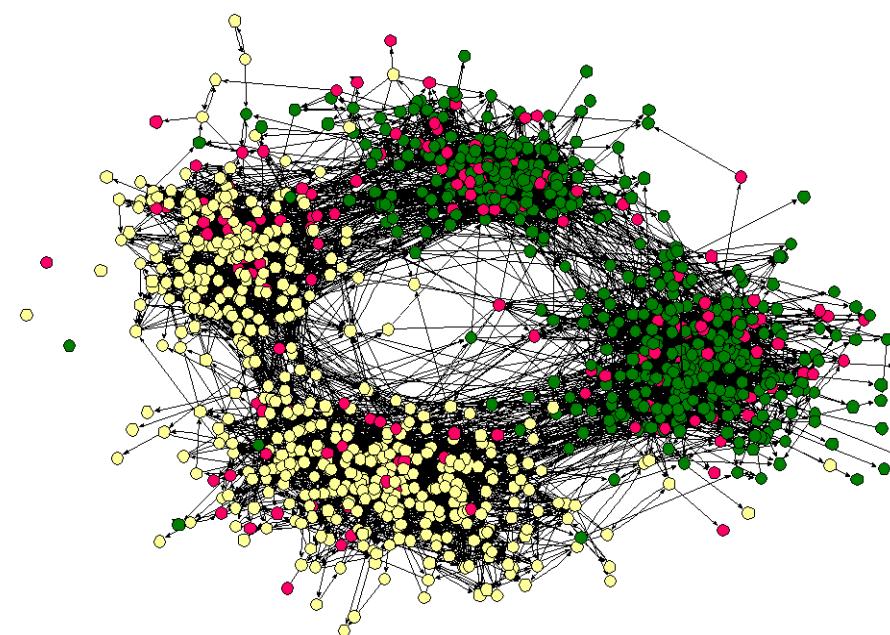
Transportation



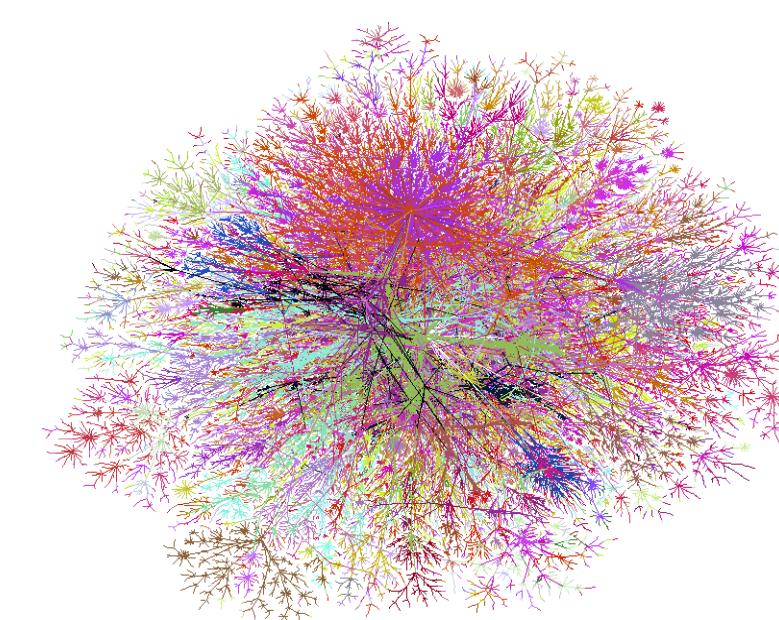
Protein-Protein Interaction



Friendship



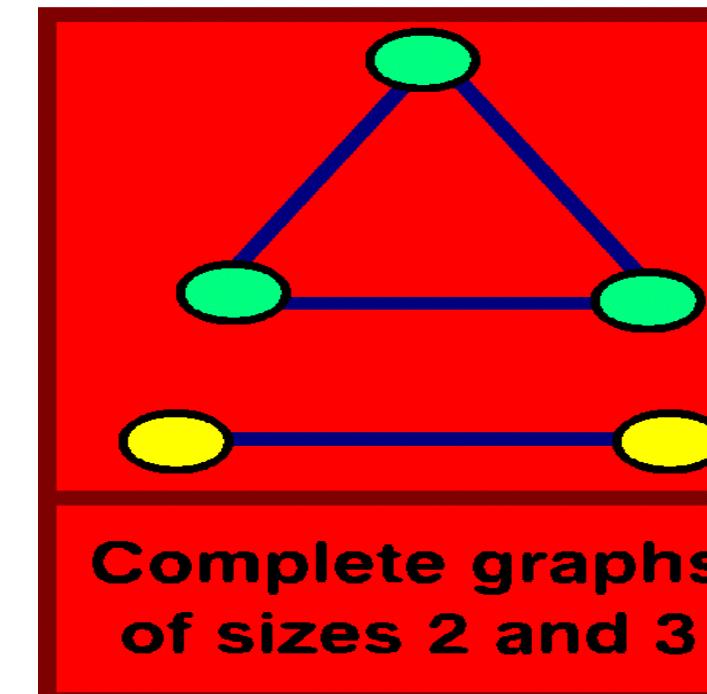
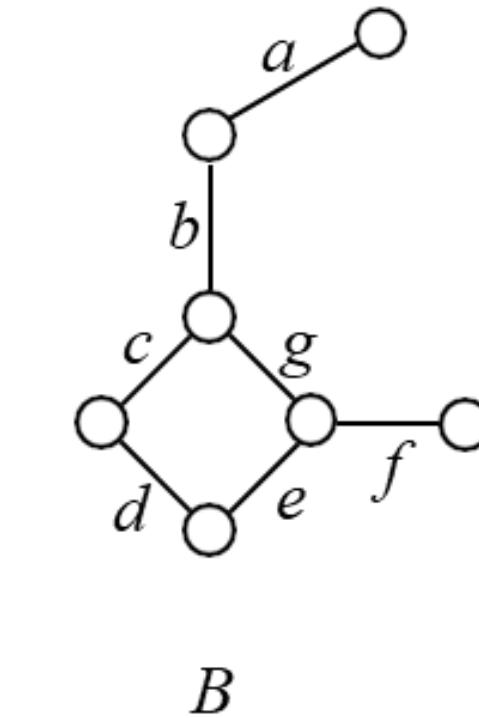
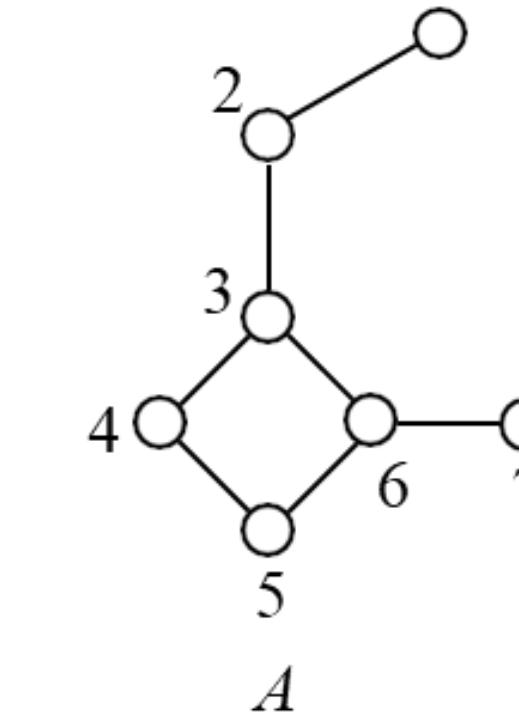
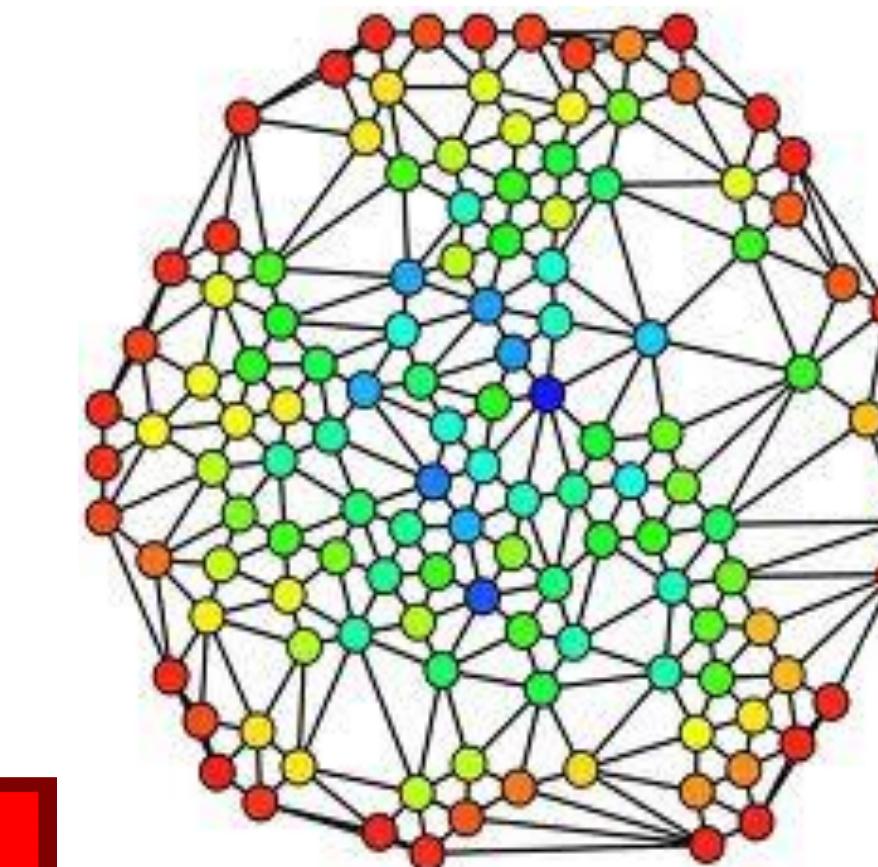
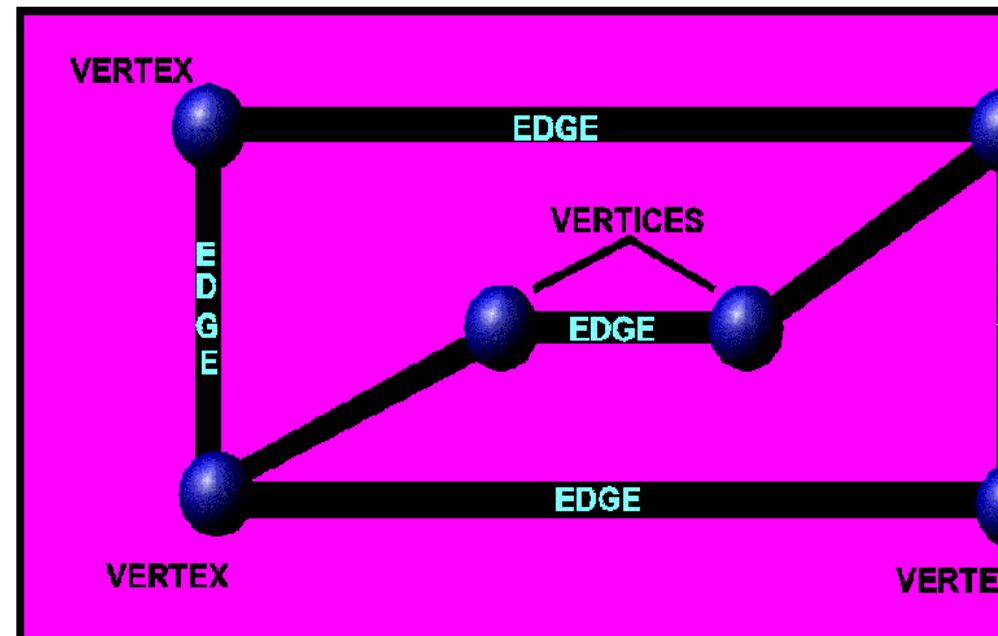
Internet



- Network = Graph
- Informally a *graph* is a set of nodes joined by a set of lines or arrows.

What is a graph?

Graph is a pair $G = (V, E)$,
where V is the Vertex set and E is the Edge set



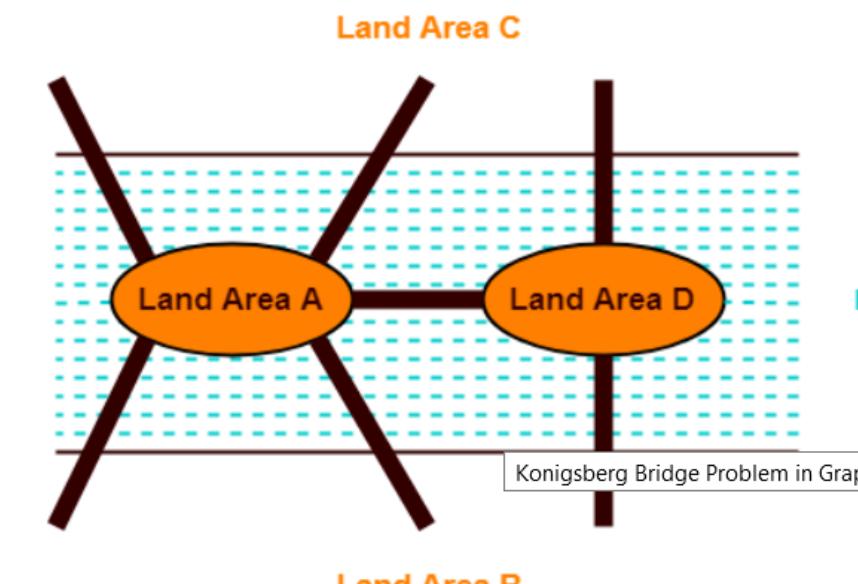
Vertex-labeled (A) and edge-labeled (B) graph G

Graph-Based Representations

- Representing a problem as a graph can provide a different point of view
- Representing a problem as a graph can make a problem much simpler
 - More accurately, it can provide the appropriate tools for solving the problem

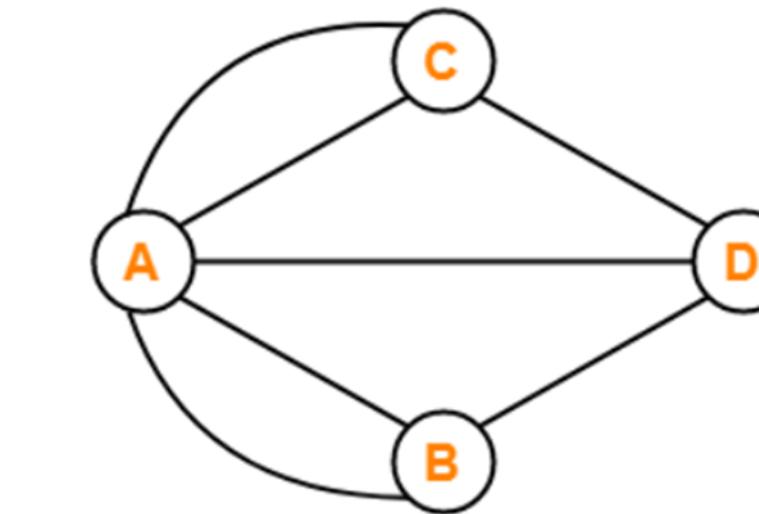
What makes a problem graph-like?

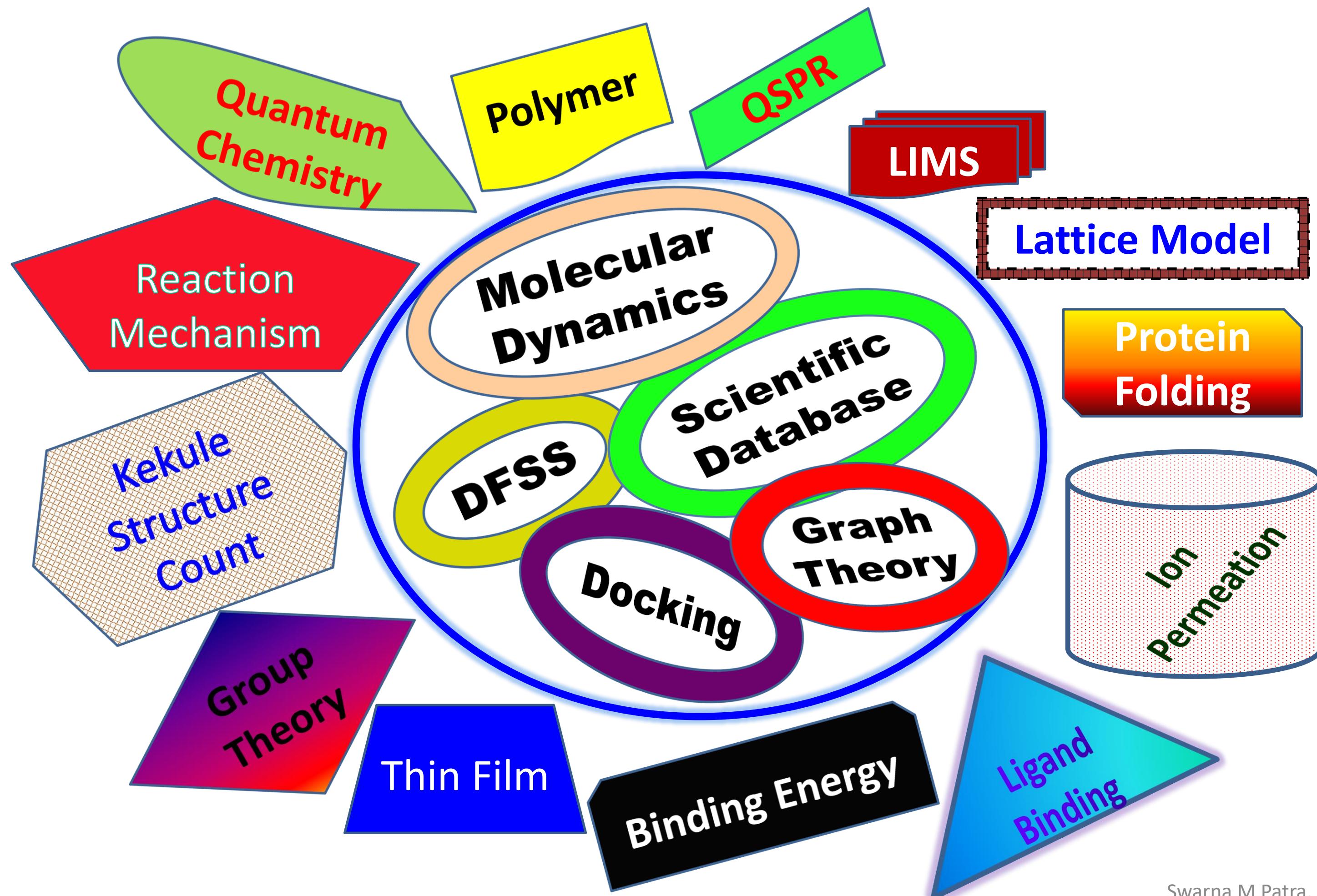
- There are two components to a graph
 - Nodes and Edges
- In graph-like problems, these components have natural correspondences to problem elements
 - Entities are nodes and interactions between entities are edges
- Most complex systems are graph-like

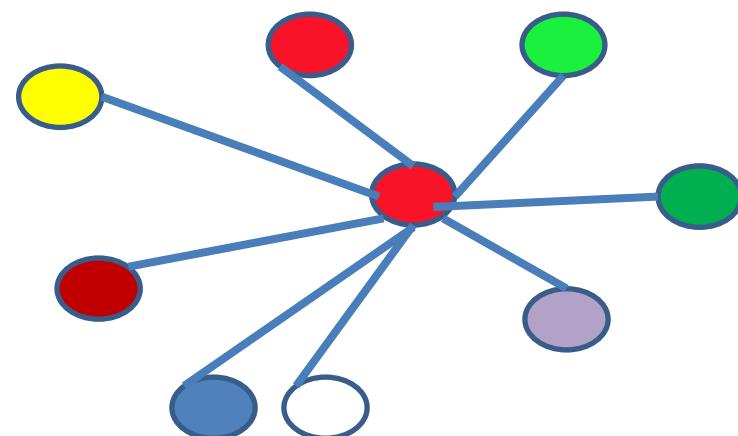
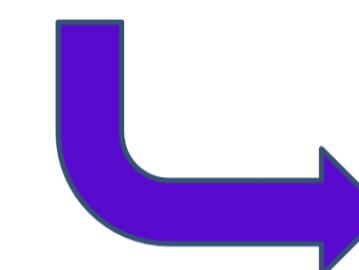
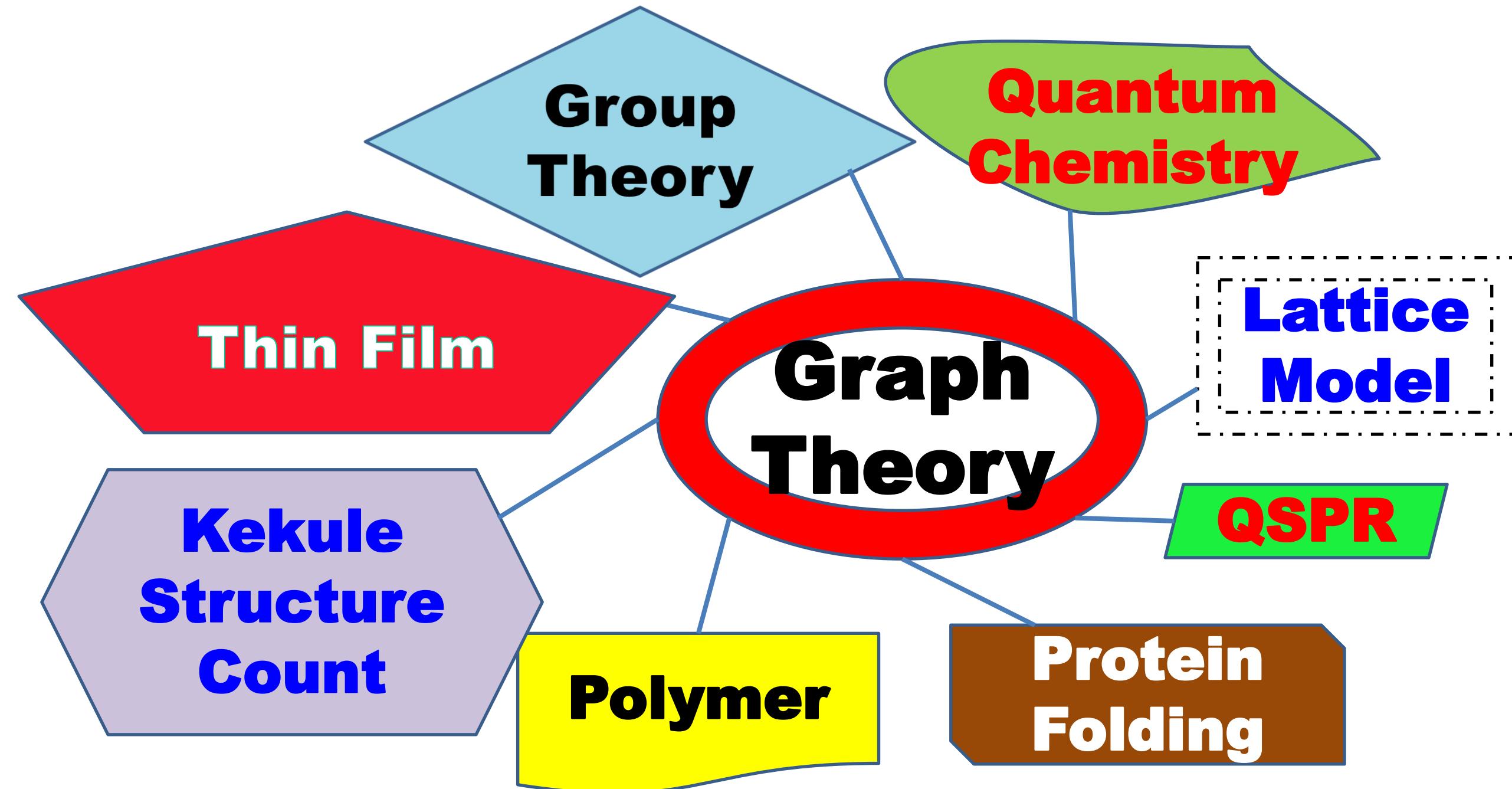


River Pregel

Konigsberg Bridge Problem

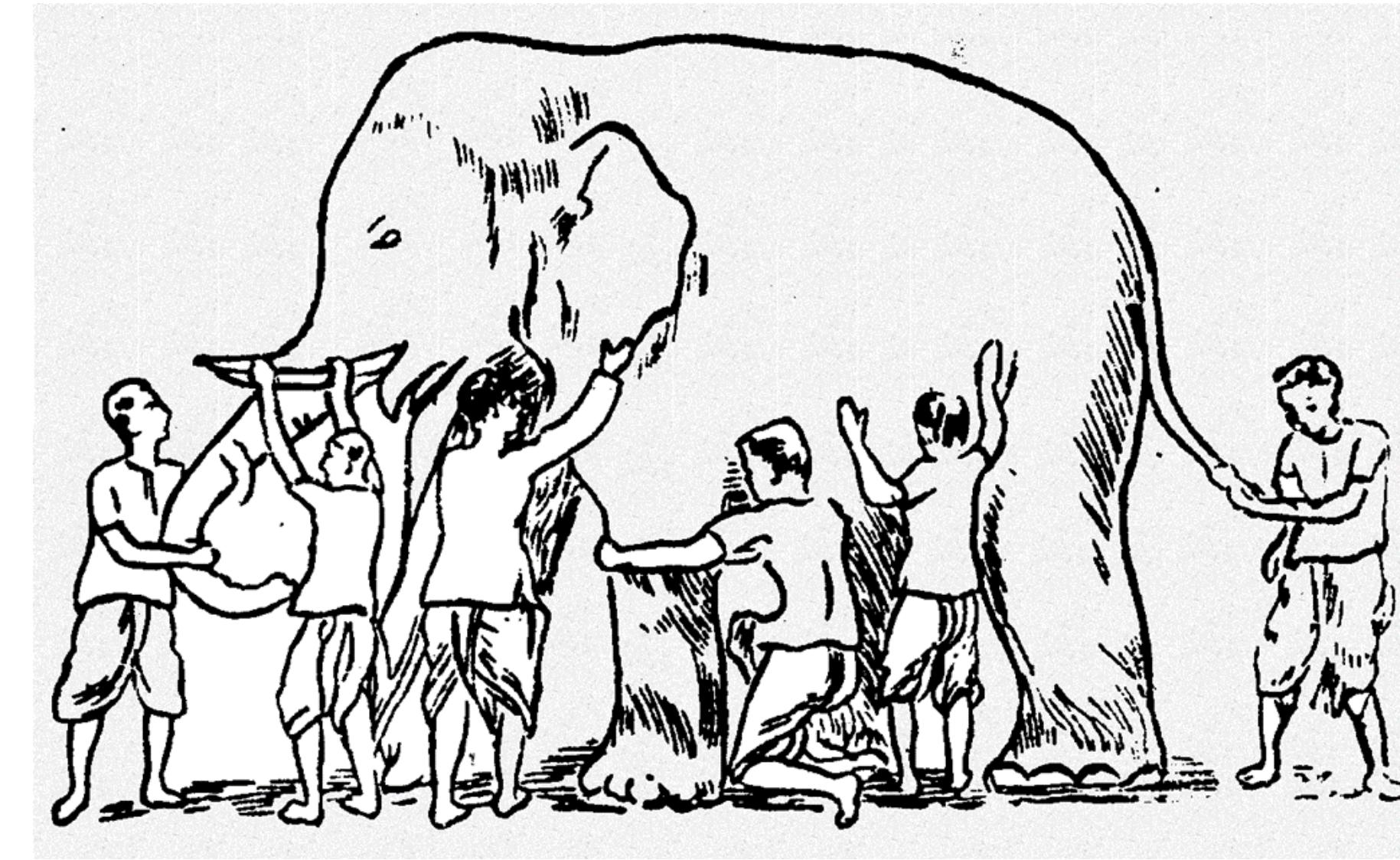






- **adjacency** – u, v adjacent if $\{u, v\}$ edge
- **vertex degrees** – number of adjacent vertices
- **paths** – sequences $u_1, u_2 \dots u_n$ such that each $\{u_i, u_{i+1}\}$ is an edge + no multiple edges
- **circuits** – closed paths
- **cycles** – circuits with all vertices appearing only once
- **simple graphs** – no loops and no multiple edges
- **connected graphs** – every two vertices connected by a path
- **trees** – connected graph without cycles

- “Similarity” can have quite different meanings in chemical approaches.
- Molecular Similarity does not just mean similarity of structural features.
- Similarity in a chemical context must include additional properties.



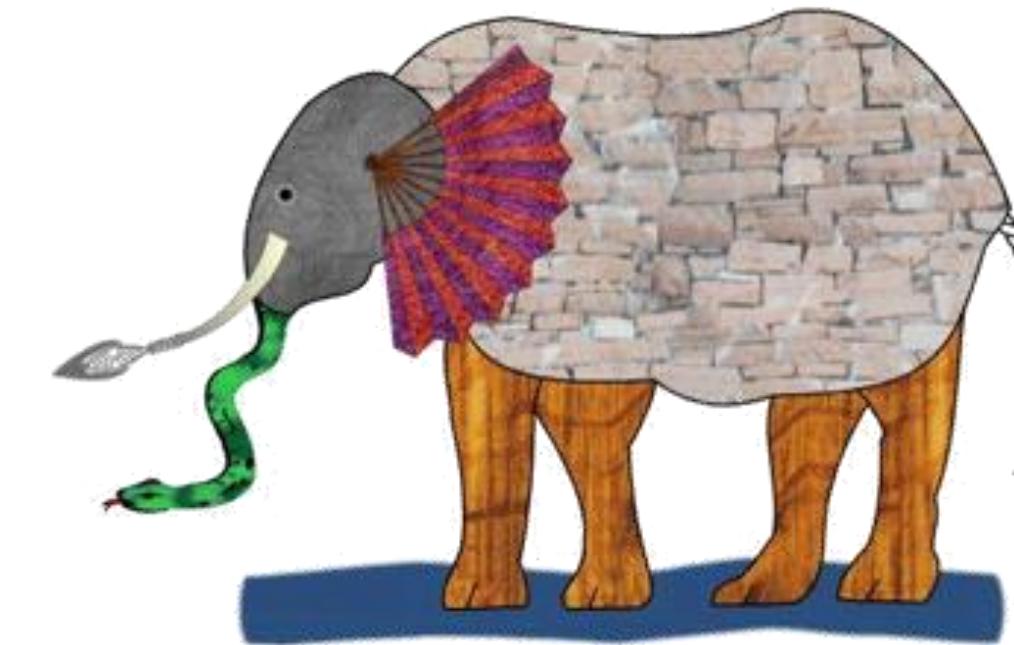
It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind

The First approached the Elephant,
And happening to fall
Against his **broad and sturdy side**,
At once began to bawl:
“God bless me! but the Elephant
Is very like a **wall!**”

The Second, feeling of the **tusk**,
Cried, “Ho! what have we here
So very round and smooth and sharp?
To me ’tis mighty clear
This wonder of an Elephant
Is very like a **spear!**”

The Third approached the animal,
And happening to take
The squirming **trunk** within his hands,
Thus boldly up and spake:
“I see,” quoth he, “the Elephant
Is very like a **snake!**”

It's an Elephant



The Fourth reached out an eager hand,
And felt about the **knee**.
“What most this wondrous beast is like
Is mighty plain,” quoth he;
“ ‘Tis clear enough the Elephant
Is very like a **tree!**”

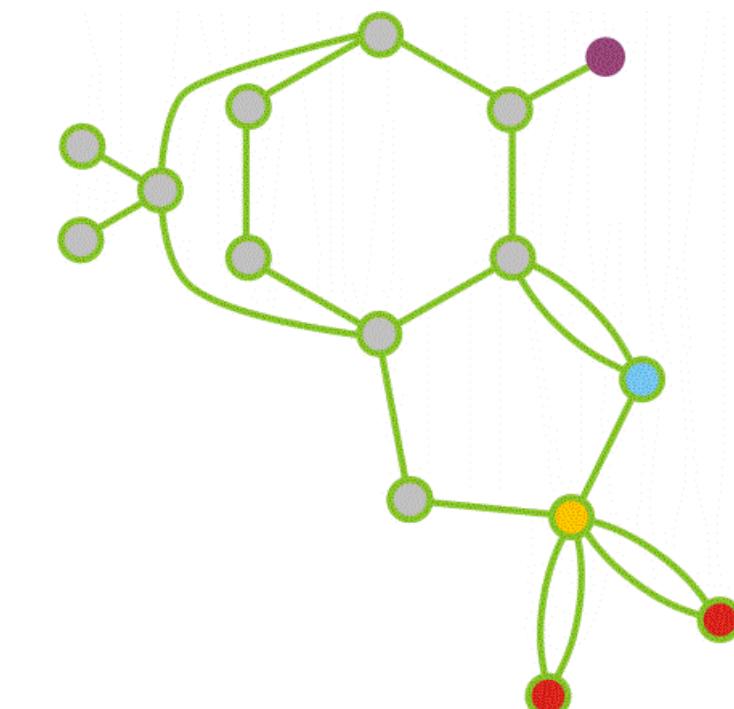
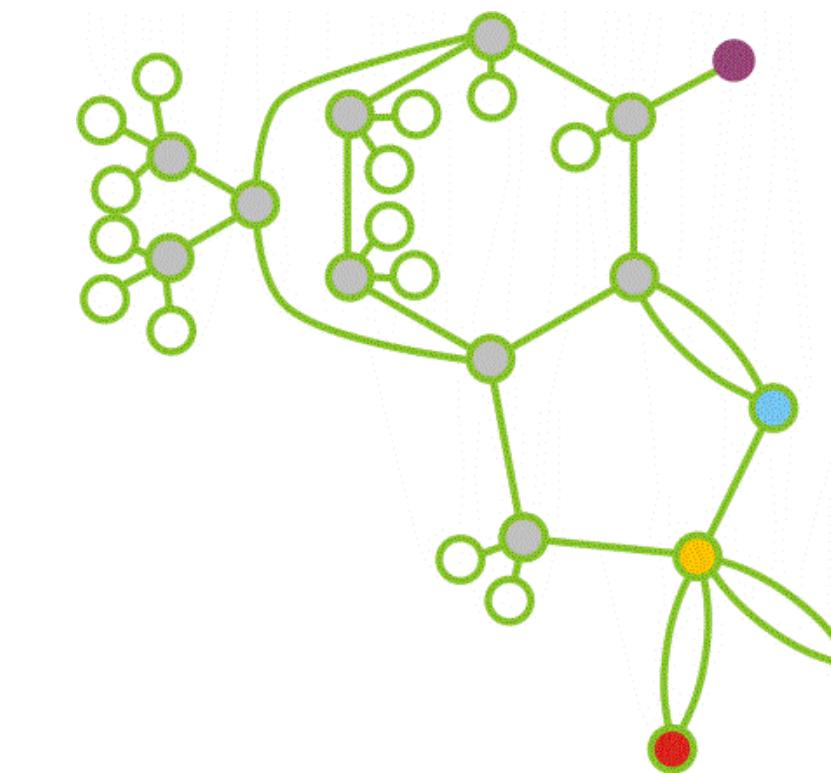
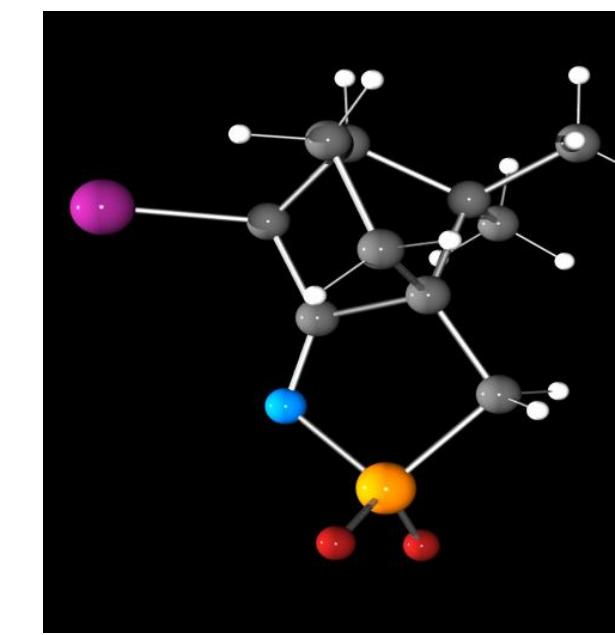
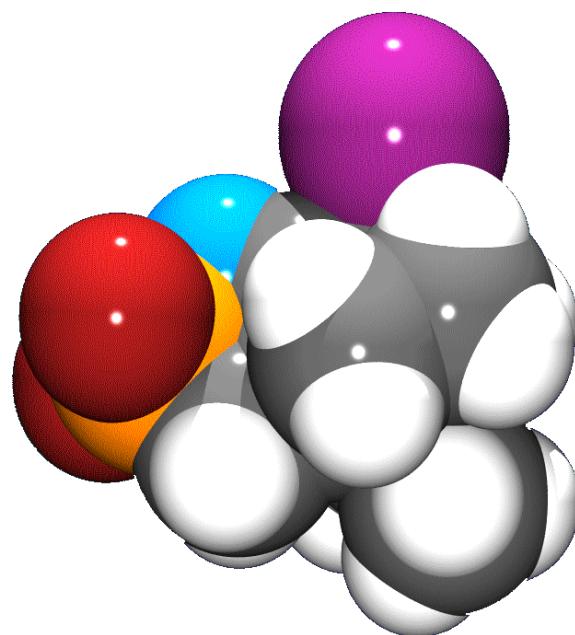
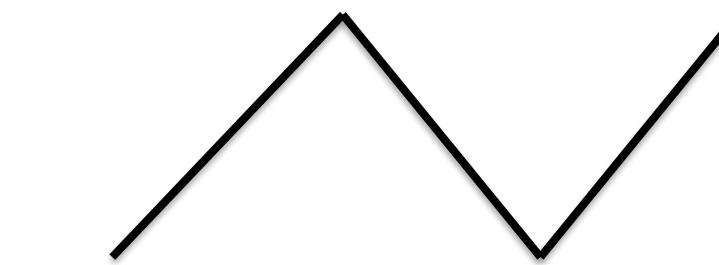
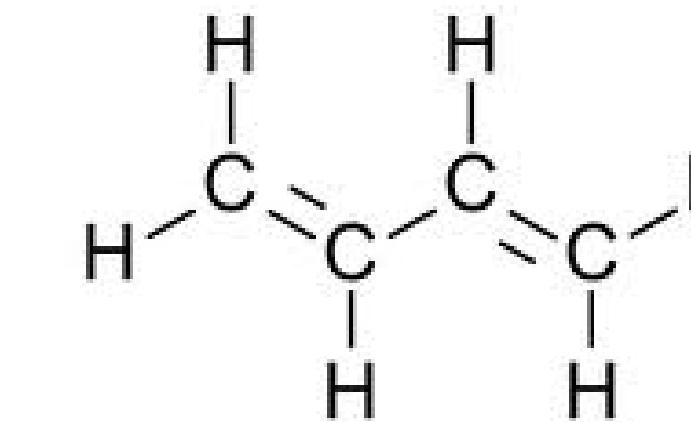
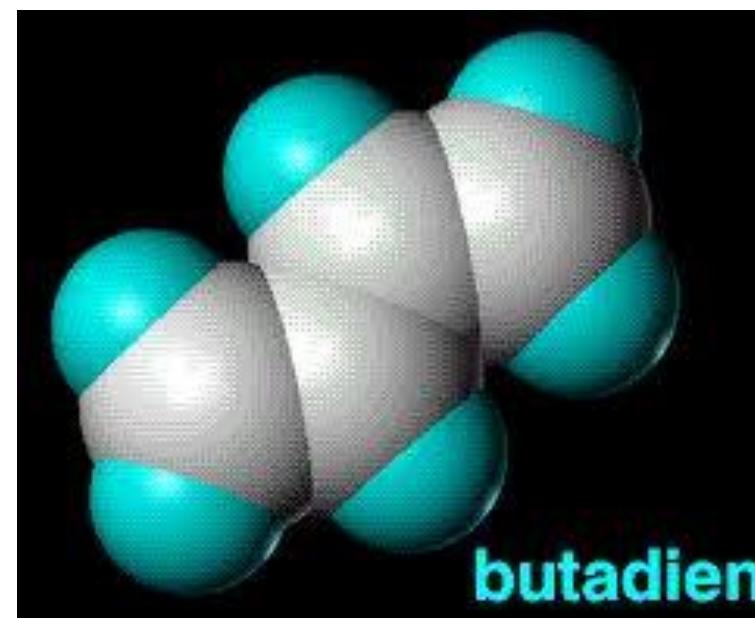
The Fifth, who chanced to touch the **ear**,
Said: “E’en the blindest man
Can tell what this resembles most;
Deny the fact who can
This marvel of an Elephant
Is very like a **fan!**”

The Sixth no sooner had begun
About the beast to grope,
Than, seizing on the **swinging tail**
That fell within his scope,
“I see,” quoth he, “the Elephant
Is very like a **rope!**”

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

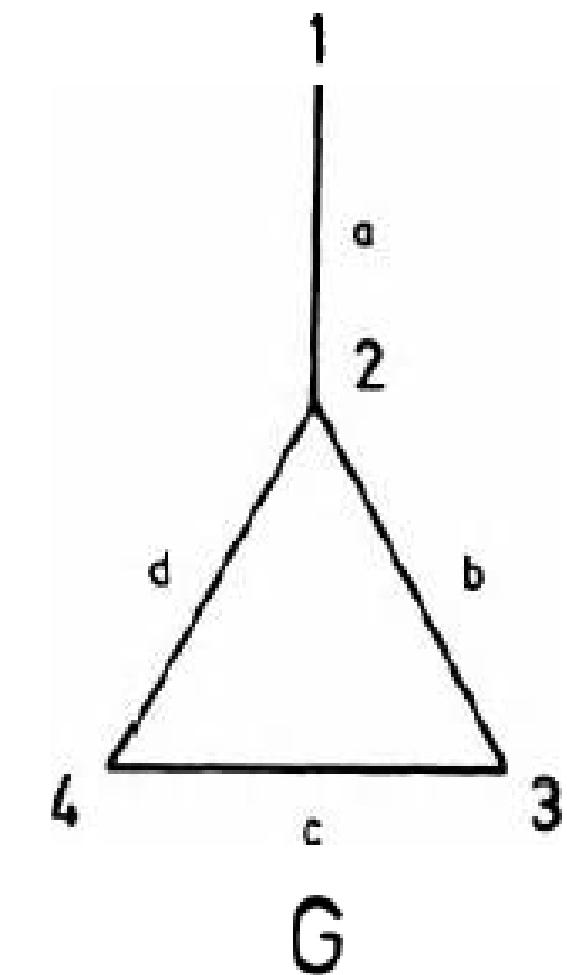
- John Godfrey Saxe (1816-1887)

- molecular (structural) graphs (often: hydrogen-supressed)
- degree of a vertex = valence of atom

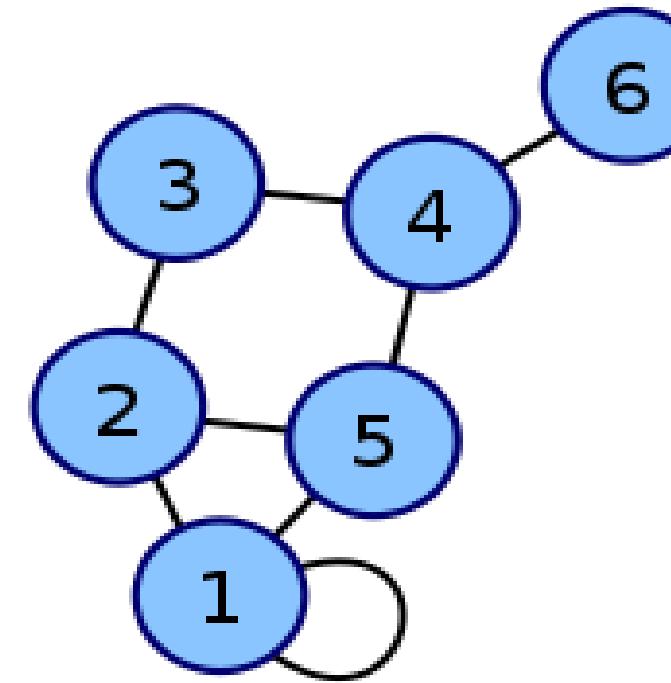


The *vertex-adjacency matrix*, denoted by \mathbf{A} , of a vertex-labeled connected simple graph G with V vertices is a square $V \times V$ matrix, which is determined by the adjacencies of vertices in G

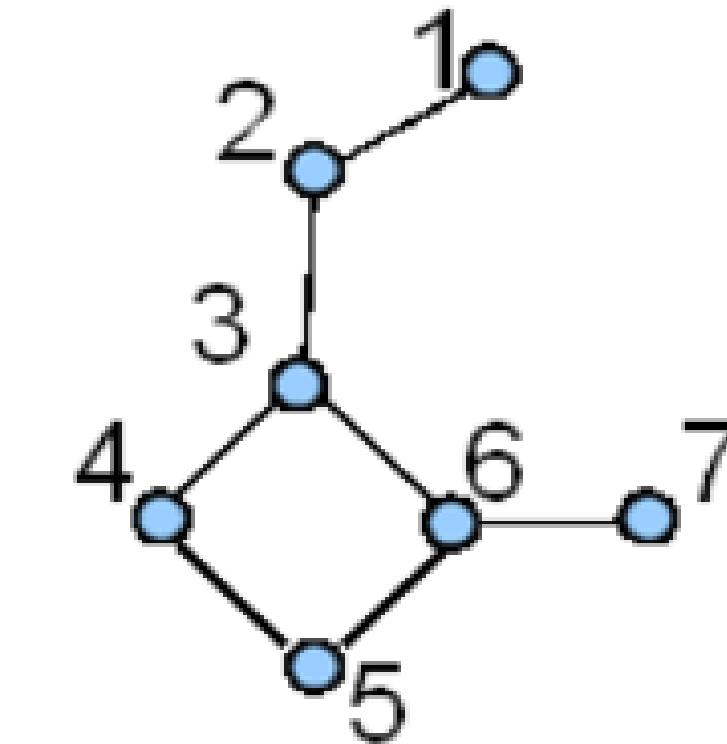
$$[A]_{ij} = \begin{cases} 1, & \text{if vertices } i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}$$



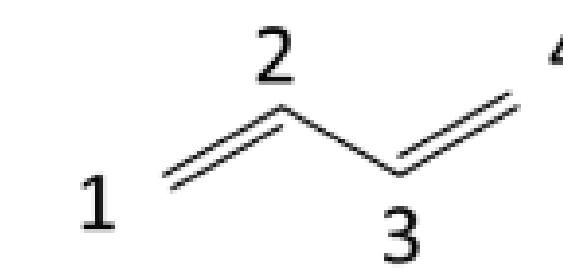
$$A(G) = \begin{pmatrix} \textcircled{1} & \textcircled{2} & \textcircled{3} & \textcircled{4} \\ \textcircled{1} & 0 & 1 & 0 & 0 \\ \textcircled{2} & 1 & 0 & 1 & 1 \\ \textcircled{3} & 0 & 1 & 0 & 1 \\ \textcircled{4} & 0 & 1 & 1 & 0 \end{pmatrix}$$



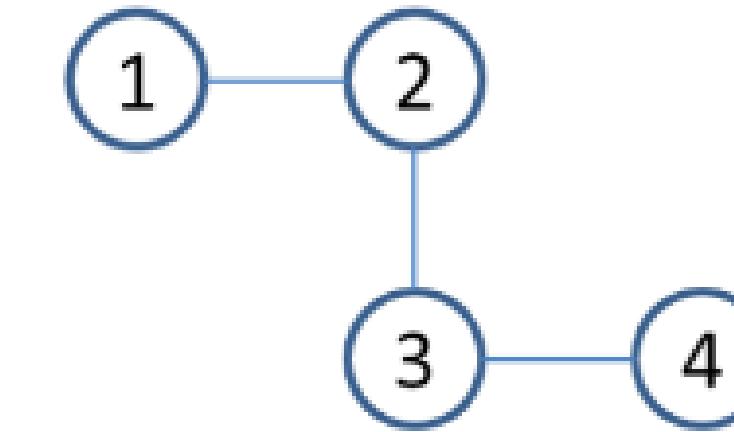
$$\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$



$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$



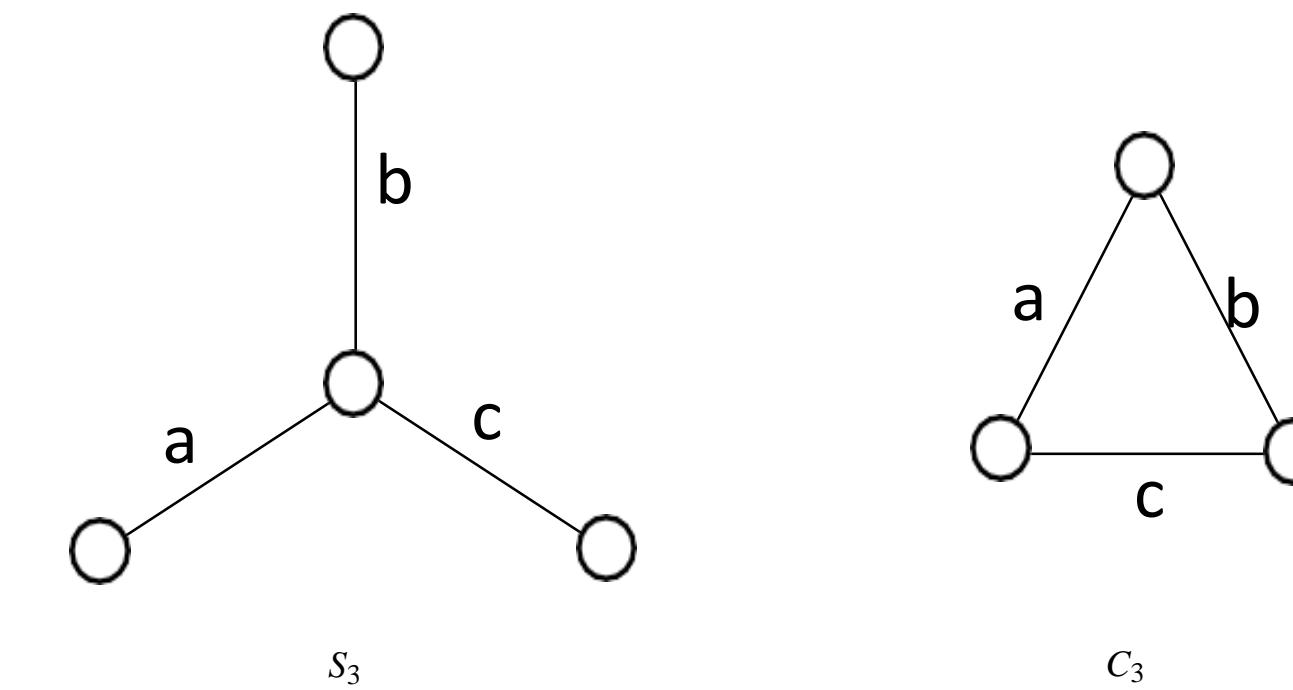
molecule



graph

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

adjacency matrix

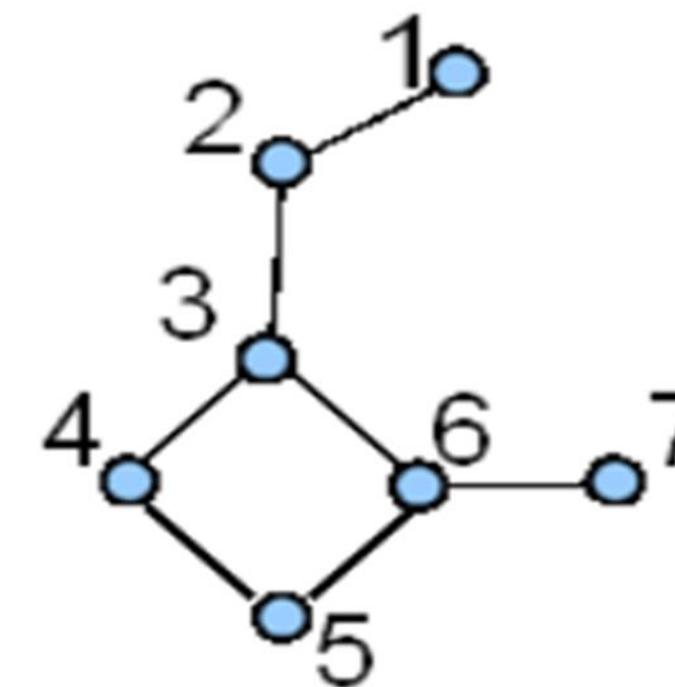


$$[E] = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

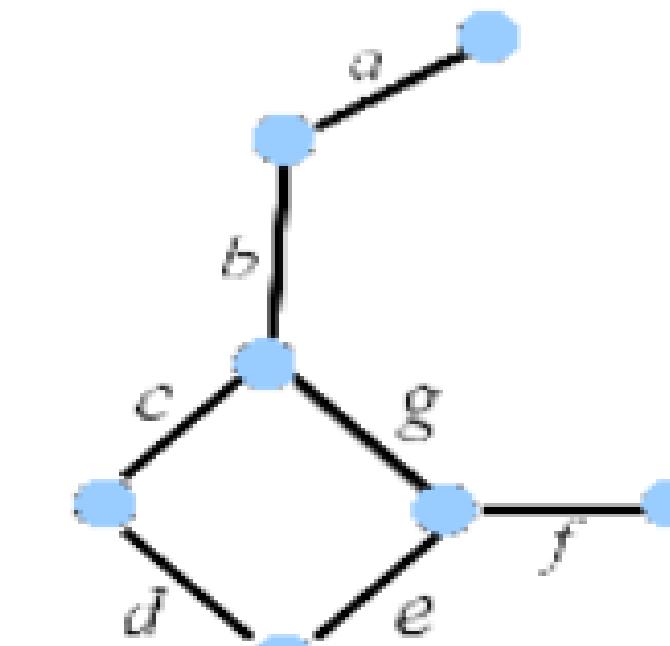
The *edge-adjacency matrix*, denoted by E , of an edge-labeled connected graph G is a square $E \times E$ matrix which is determined by the adjacencies of edges

$$[E]_{ij} = \begin{cases} 1, & \text{if edges } i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}$$

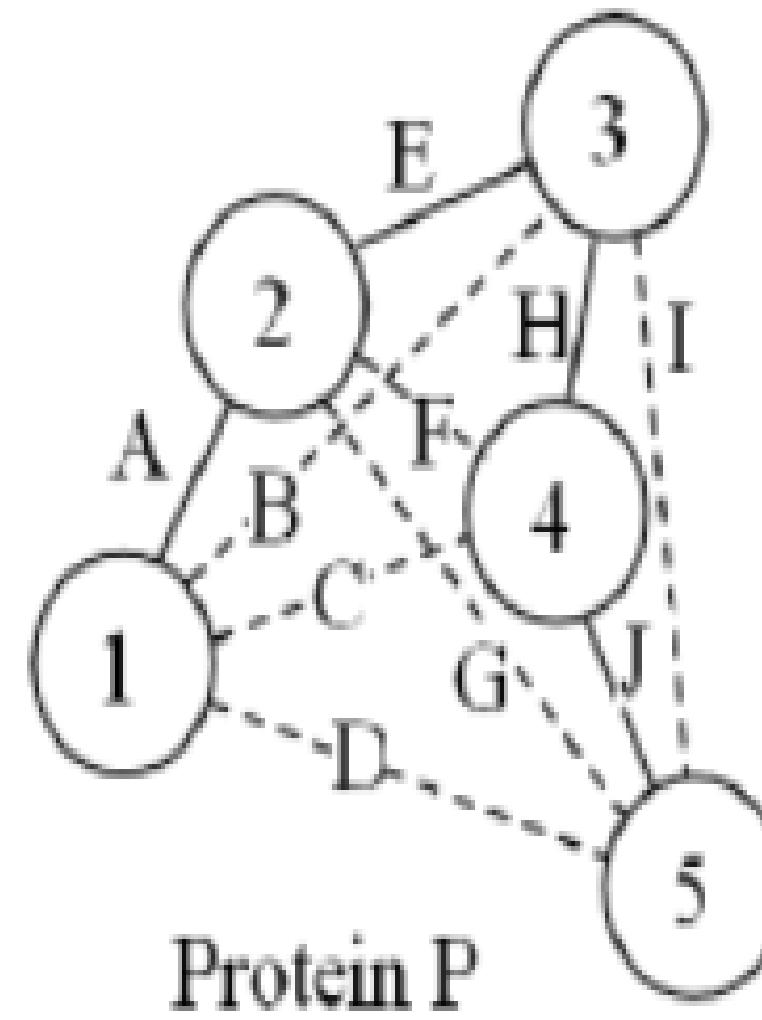
It should be noted that the vertex-adjacency matrix *uniquely* determines a graph, but the edge-adjacency matrix *does not*, that is, there are known graphs with *identical* edge-adjacency matrices. A pair of nonisomorphic graphs – the three-point star S_3 and the cycle on three vertices C_3 – possessing identical edge-adjacency matrices.



$${}^v\mathbf{A}(G_1) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

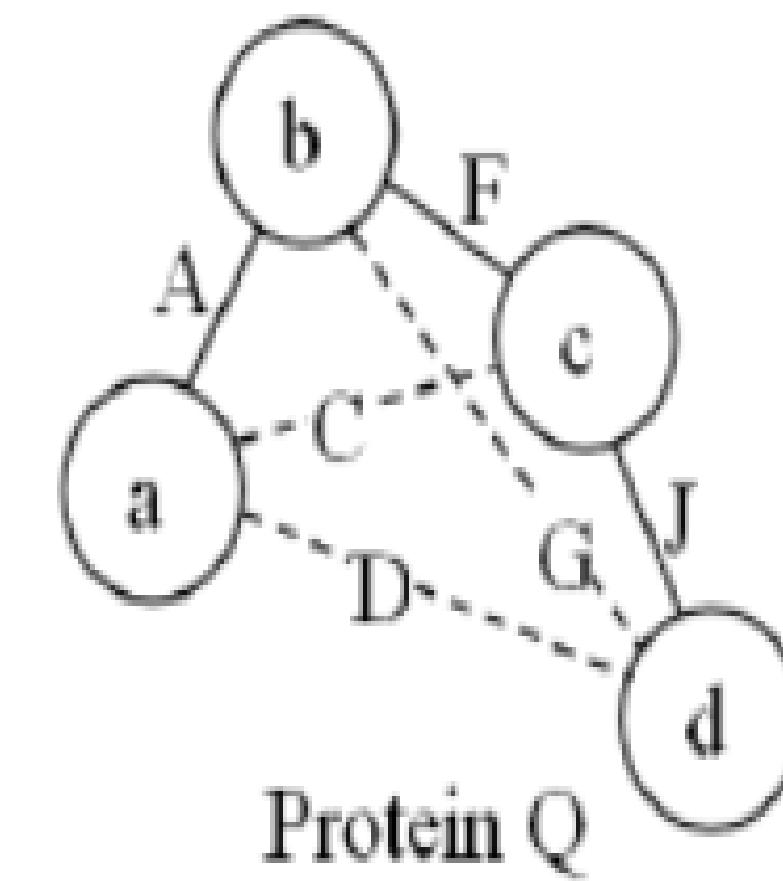


$${}^e\mathbf{A}(G_1) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$



	1	2	3	4	5
1	0	A	B	C	D
2	A	0	E	F	G
3	B	E	0	H	I
4	C	F	H	0	J
5	D	G	I	J	0

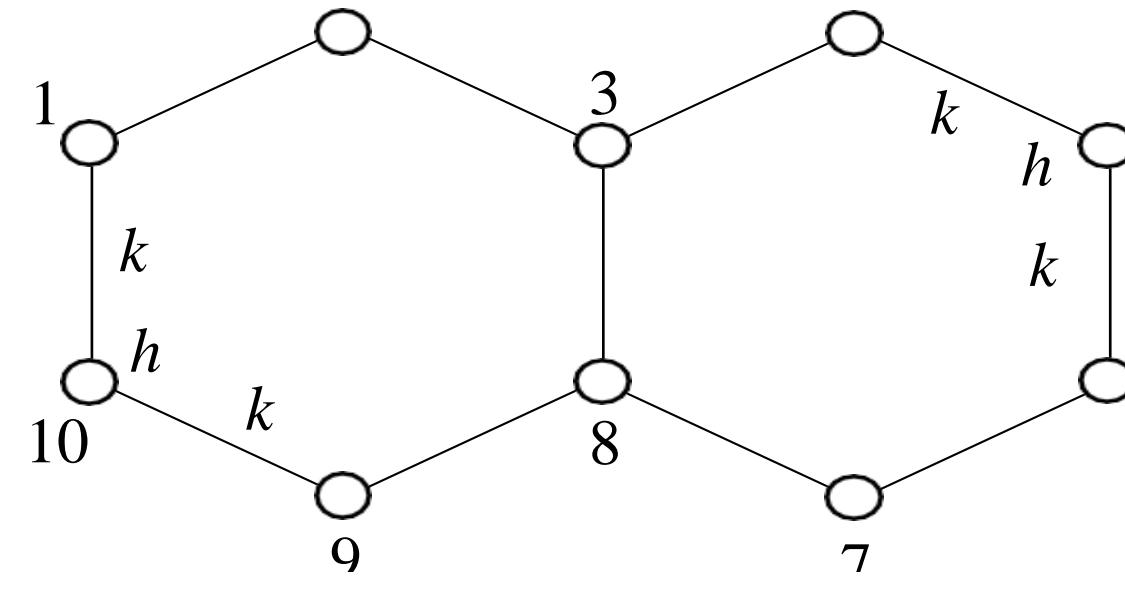
Distance matrix of P



	a	b	c	d
a	0	A	C	D
b	A	0	F	G
c	F	0	0	J
d	D	G	J	0

Distance matrix of Q

Vertex-Adjacency Matrix of Weighted Graphs



$$[A] = \begin{cases} k & \text{if the edge } i-j \text{ is weighted} \\ 1 & \text{if the edge } i-j \text{ is not weighted} \\ h & \text{if the vertex } i \text{ is weighted} \\ 0 & \text{otherwise} \end{cases}$$

The parameters h and k depend, respectively, on the chemical nature of the corresponding atoms and bonds in a molecule. Some people select for them the values of the Hückel parameters for heteroatoms and heterobonds.

$$[A] = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & k \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & k & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & k & h & k & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & k & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & k & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ k & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & h \end{bmatrix}$$

A topological index, sometimes also known as a graph-theoretic index, is a numerical invariant of a chemical graph, some examples of topological indices include the Balaban index, Harary index, molecular topological index, and Wiener index.

Unless otherwise stated, hydrogen atoms are usually ignored in the computation of such indices

The topological index of a graph is defined by $TI = | A + D |$

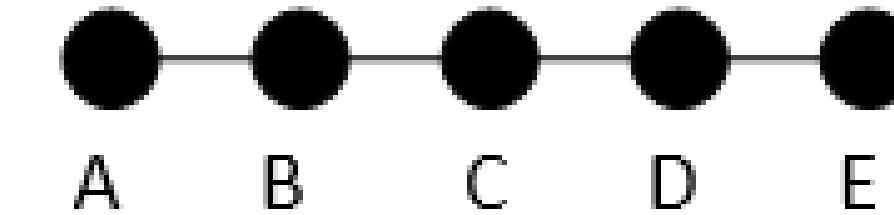
where A is the adjacency matrix, D is the graph distance matrix, and $| A + D |$ denotes the determinant of the matrix addition

Zagreb Index Z one counts the connections from each vertex (node, carbon).

The first Zagreb index $M_1(G)$ is equal to the sum of squares of the degrees of the vertices

The second Zagreb index $M_2(G)$ is equal to the sum of the products of the degrees of pairs of adjacent vertices of the underlying molecular graph

Method of calculation: If δ_i is the degree of vertex i then



$$M_1 = \sum_{i=1}^n \delta_i^2$$

$$M_2 = \sum \delta_i \delta_j$$

$$M_1 = 1^2 + 2^2 + 2^2 + 2^2 + 1^2 = 1 + 4 + 4 + 4 + 1 = 14$$

$$M_2 = 1 \times 2 + 2 \times 2 + 2 \times 2 + 2 \times 1 = 2 + 4 + 4 + 2 = 12$$

Zagreb group indices were introduced to characterize branching

Wiener Number W is the total distance between all carbon atoms (sum of the distances between each pair of carbon atoms in the molecule, in terms of carbon-carbon bonds).

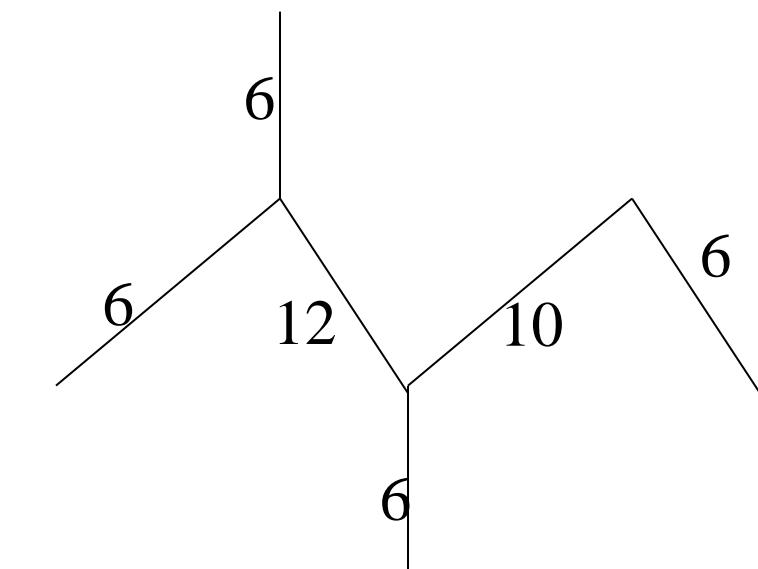
The smaller this number, the larger is the compactness of the molecule.

Method of calculation: Multiply the number of carbon atoms on one side of any bond by those on the other side; W is the sum of these two values for all bonds.

W can also be obtained by simply adding all the elements of the graph distance matrix above the main diagonal.

Or This topological index is defined as the half-sum of the elements of the distance matrix.

$$W = 6 + 6 + 12 + 6 + 10 + 6$$

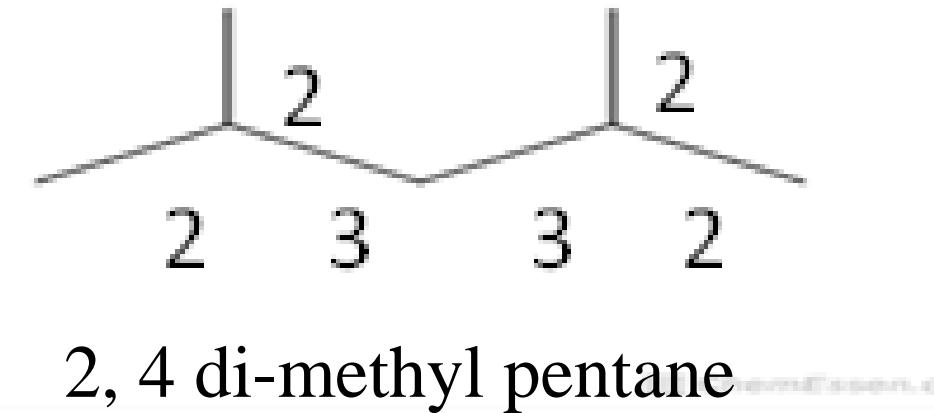


Weiner showed that the index value is closely correlated with the *boiling point of a series of alkanes*, also observed that it is correlated with other physical properties such as *density, surface tension and viscosity*.

Platt Number F Is equal to the total sum of edge-degrees in a graph G. The *edge-degree* of an edge e , $D(e)$, is the number of its adjacent edges. This index was named the *Platt number*.

Method of calculation: The Platt number of a graph G is defined by

$F(G) = \sum_{i=1}^M D(e_i)$, The Platt number, thus represents the first neighbors sum.



$$F = 2 + 2 + 3 + 3 + 2 + 2 = 14$$

Predicting physical parameters (molar volumes, boiling points, heats of formation, heats of vaporization) of alkanes

Largest Eigenvalues x_i Is equal to the highest eigenvalues of the characteristic polynomial

Method of calculation:

$$P(G;x) = \det |xI - A| \quad \text{or} \quad \det |A - xI|$$

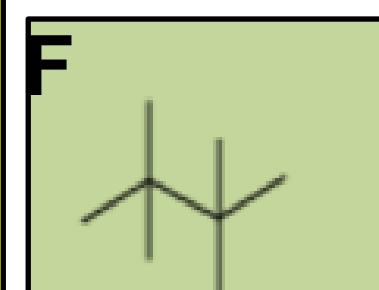
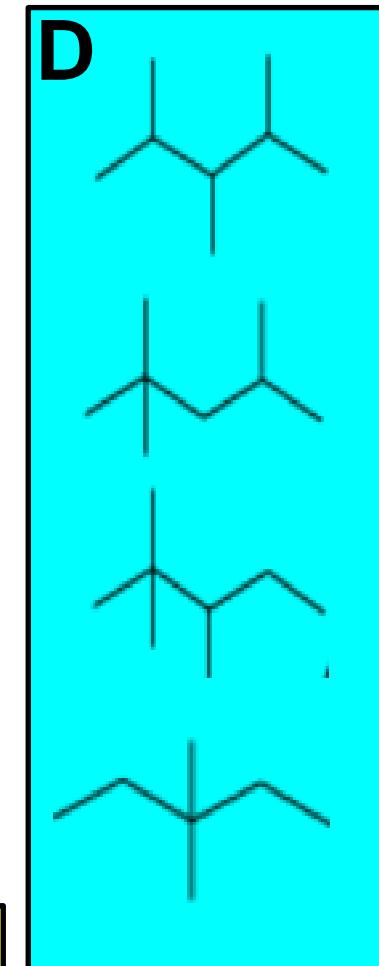
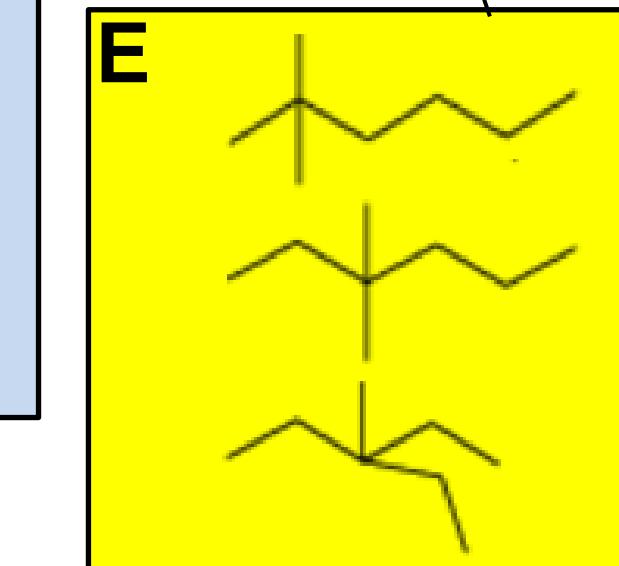
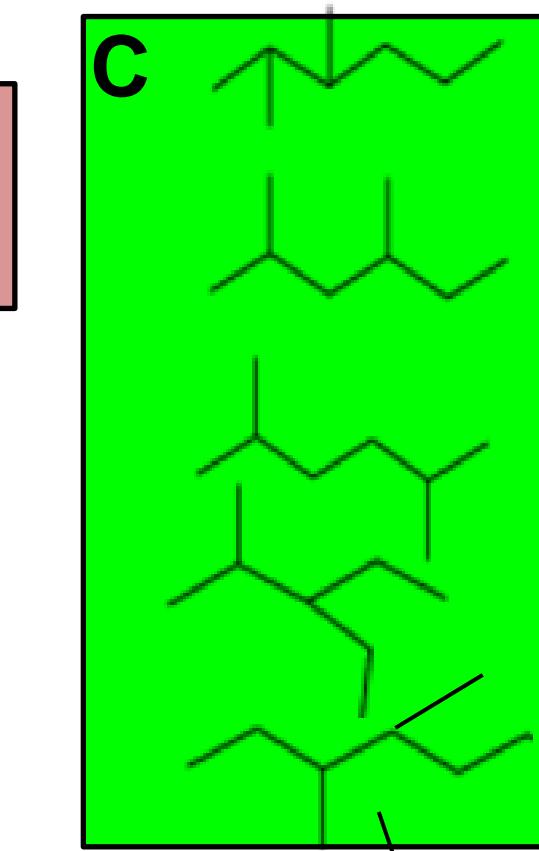
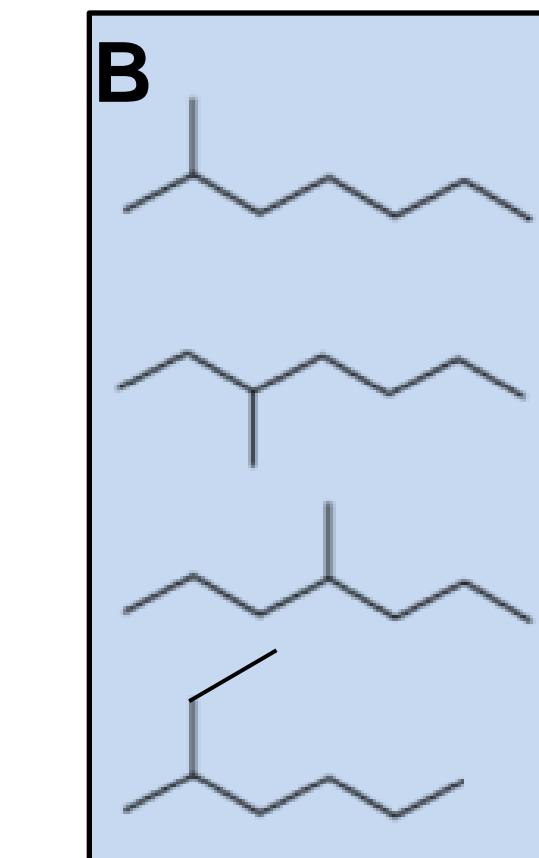
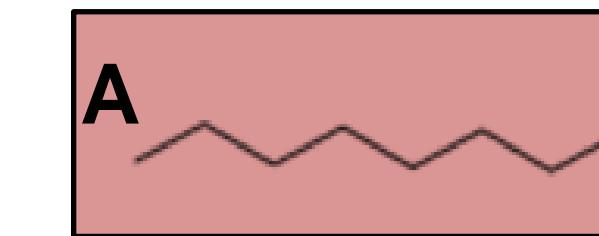
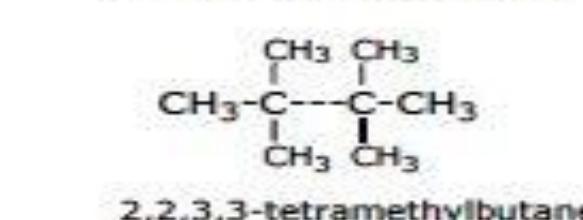
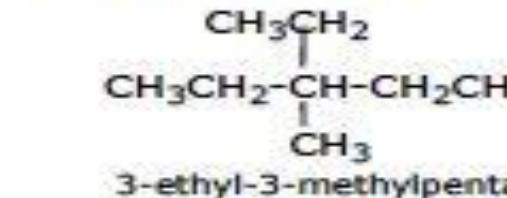
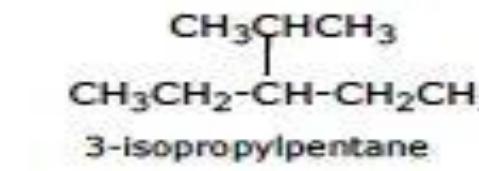
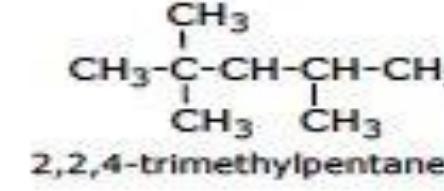
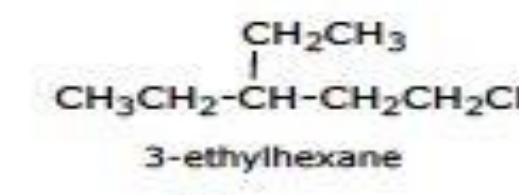
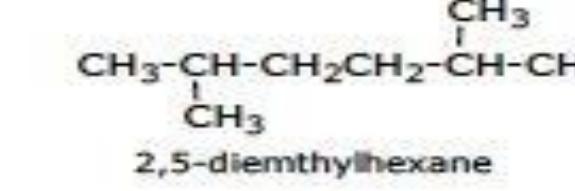
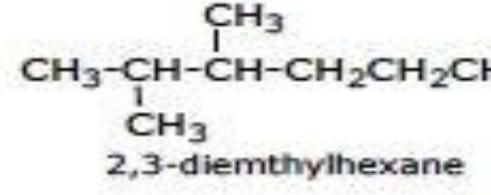
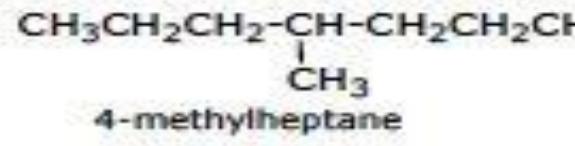
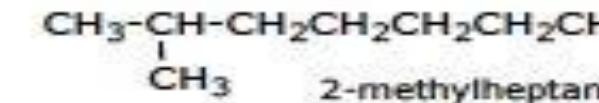
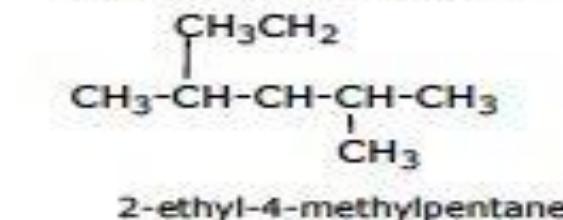
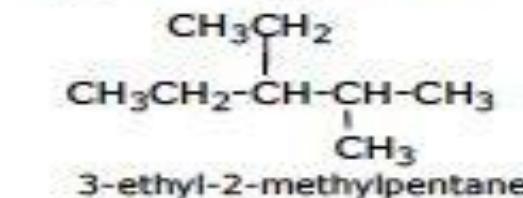
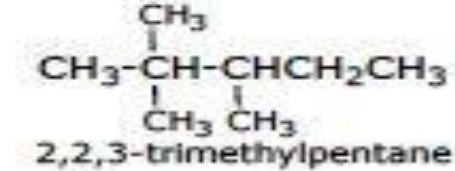
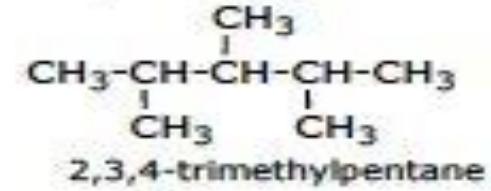
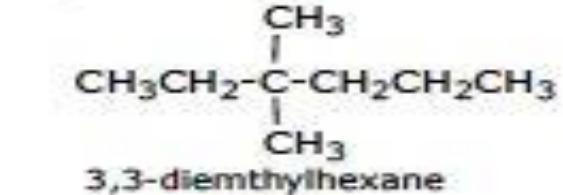
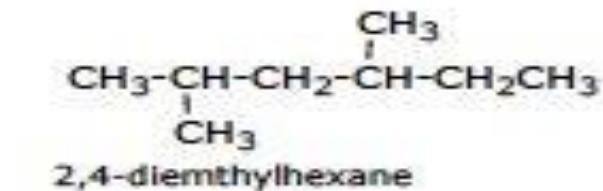
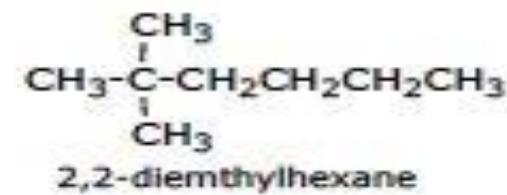
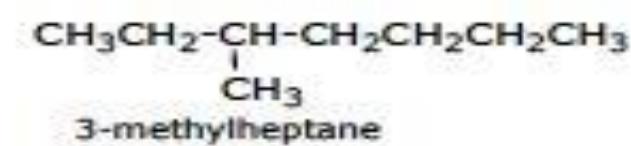
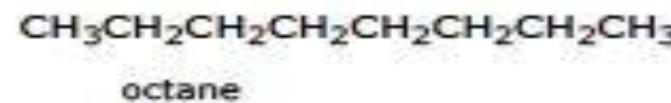
solve the characteristic polynomial $P(G;x) = 0$

Branching

Alkane Tree	x_i
	1.848
	1.932
	1.970
	2.000
	2.000
	2.053
	2.101
	2.136
	2.175

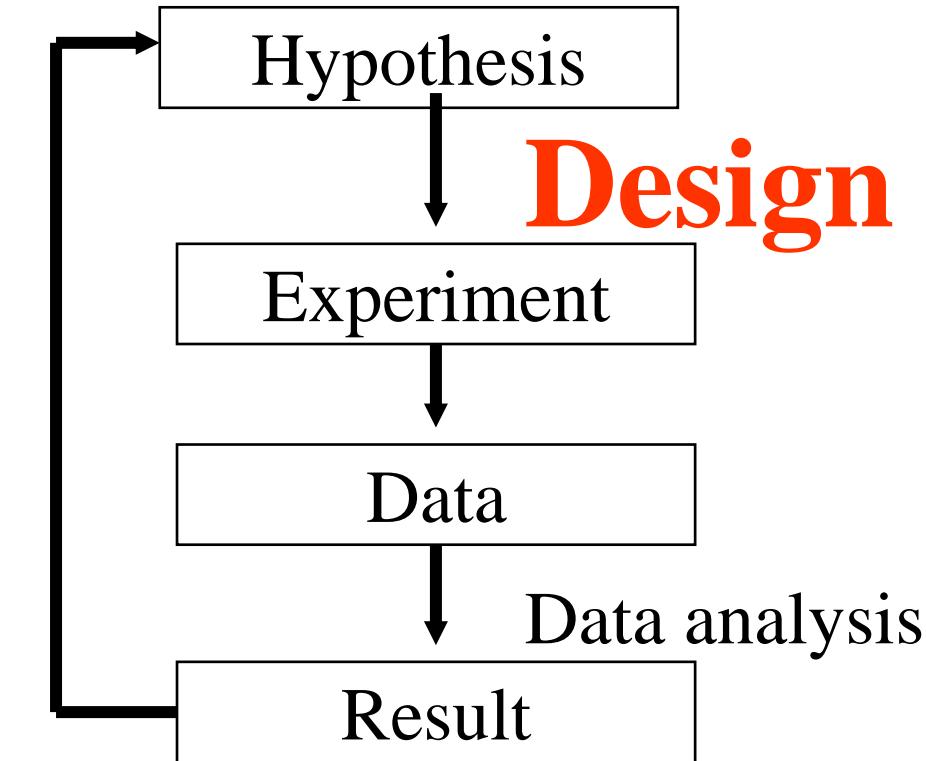
The number of terminal methyl groups are considered for the subgrouping of the molecules**

C₈H₁₈: 18 isomers

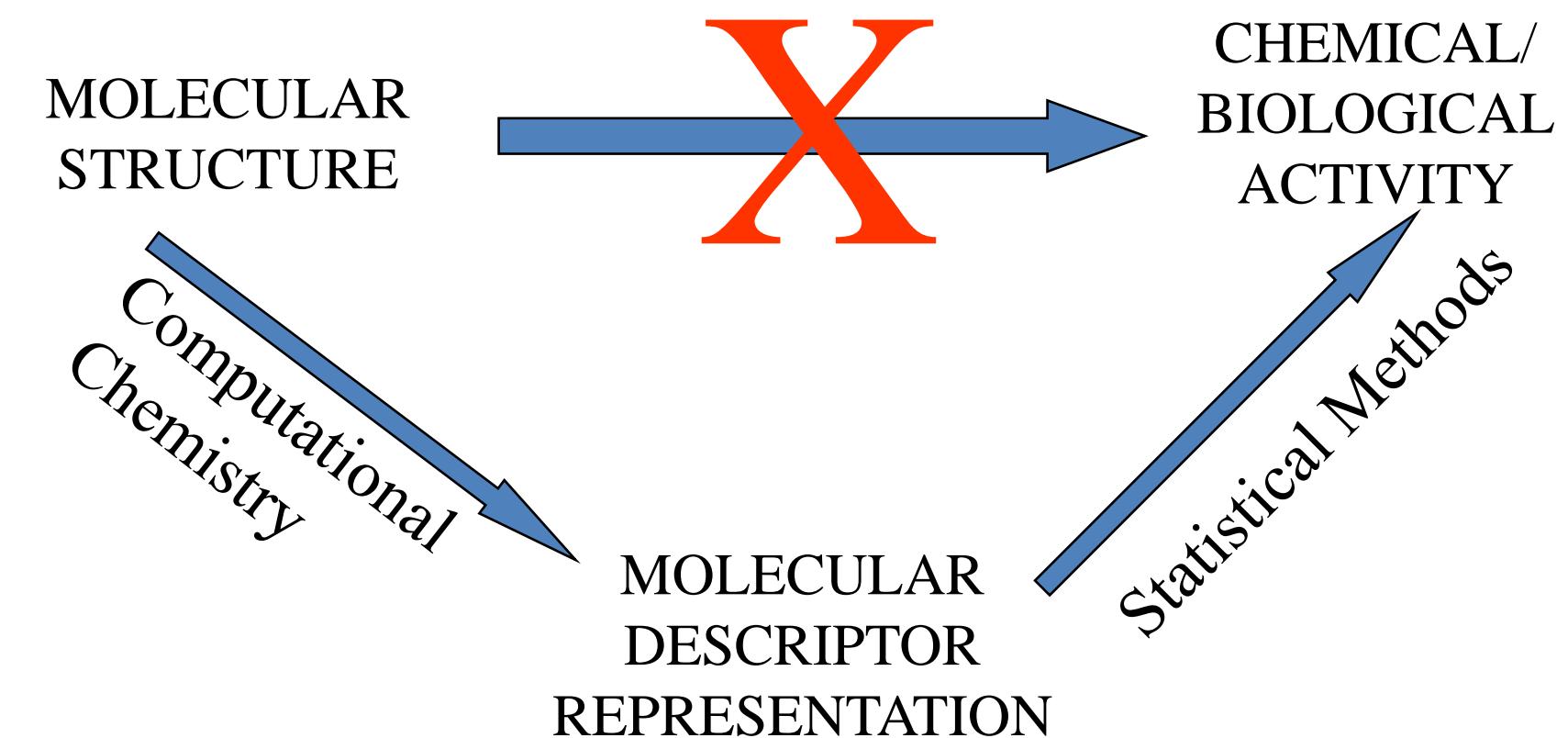


**H.Narumi and H.Hosoya, Bull. Chem. Soc. Jpn., 53, 1228 (1980)

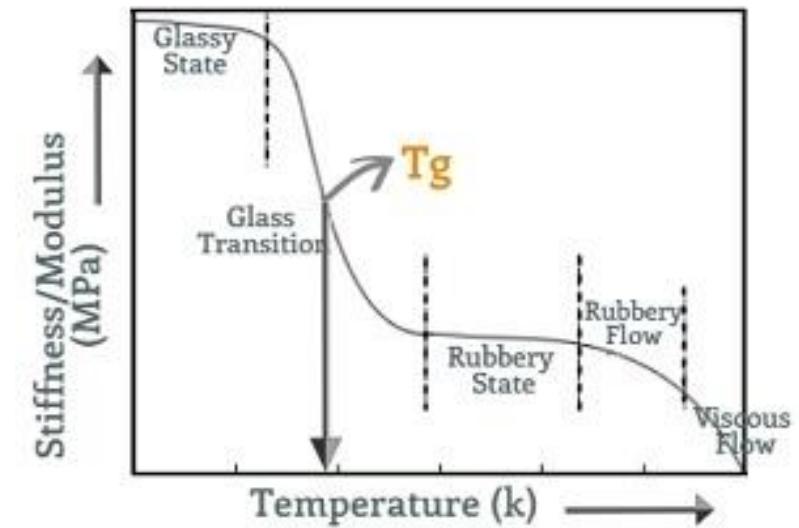
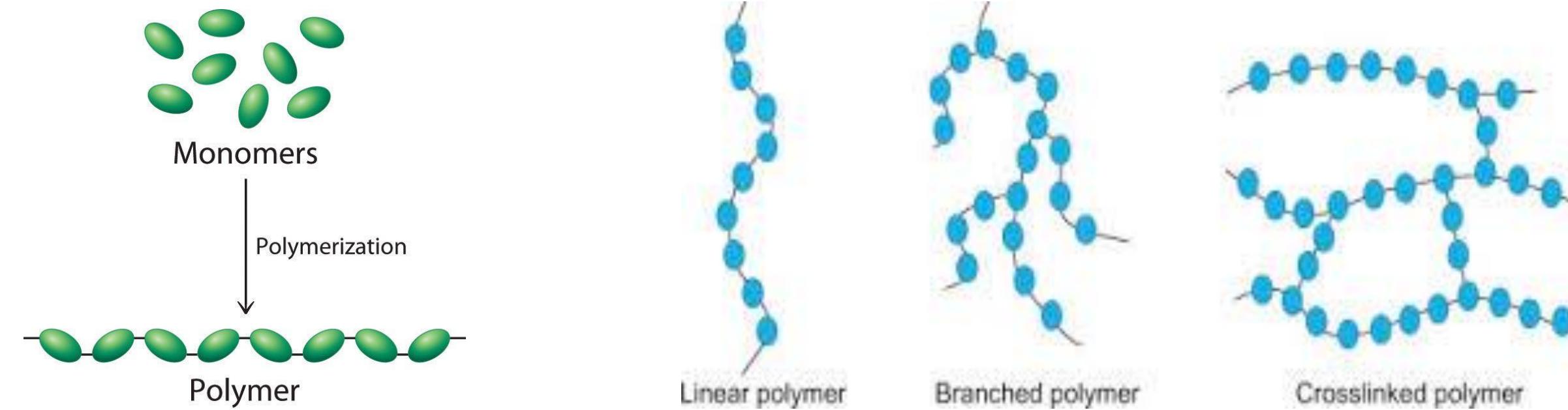
Traditional Hypothesis Driven Research Paradigm



Structure-Activity Relationships



Predictive Model for Glass Transition Temperature



Glass transition temperature:

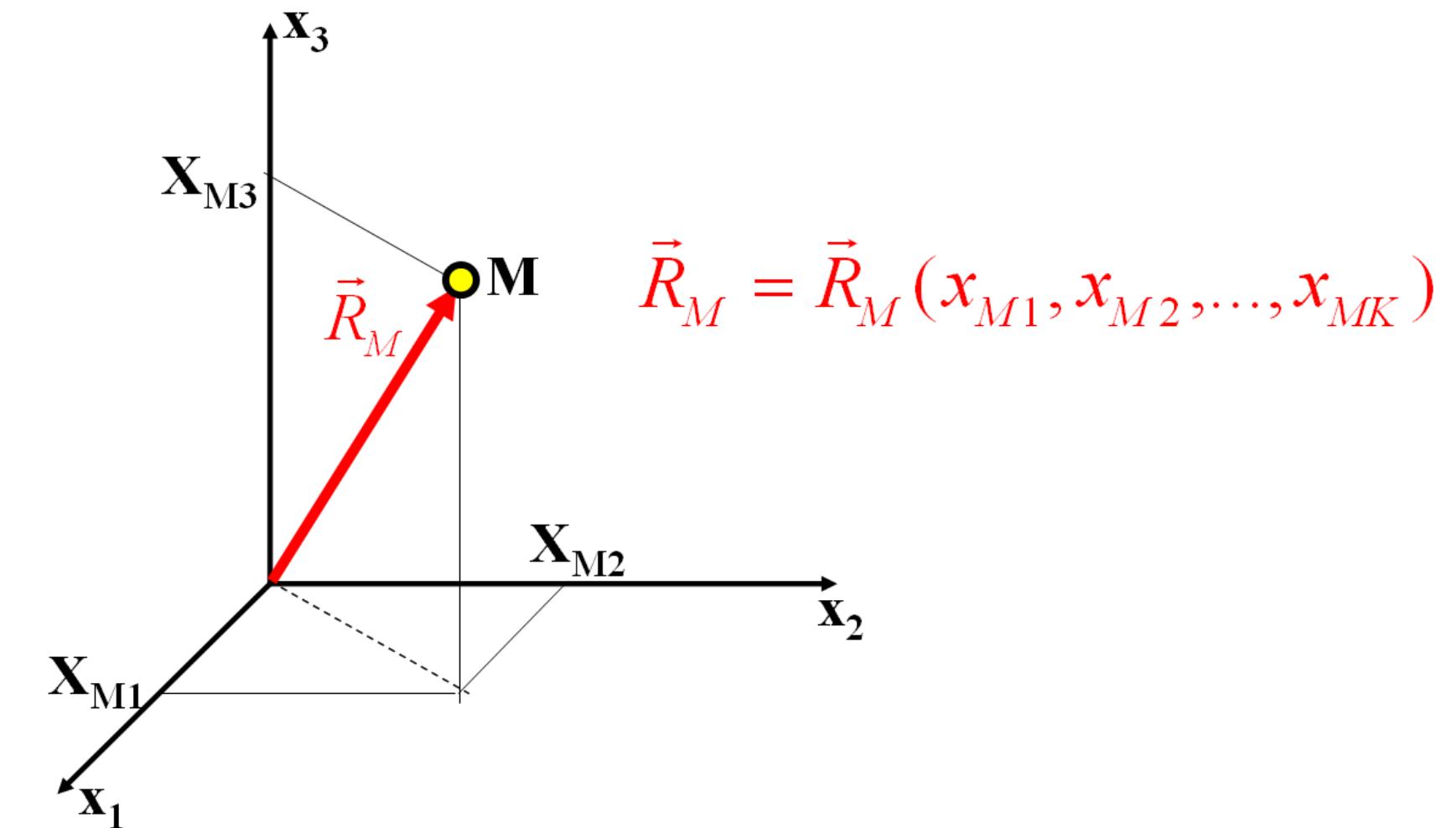
T_g of a non-crystalline material is the critical temperature at which the material changes its behavior from being glassy to rubbery

- A predictive tool designed to predict the glass transition temperature of polyesters.
- Based on monomer unit the topological indices were calculated and used in the transfer function.

Datasets are represented by a matrix of molecular descriptors

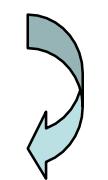
Samples (Compounds)	Variables (descriptors)			
	X_1	X_2	...	X_m
1	x_{11}	x_{12}	...	x_{1m}
2	x_{21}	x_{22}	...	x_{2m}
...
n	x_{n1}	x_{n2}	...	x_{nm}

Compounds represented by vectors in a multidimensional descriptor space

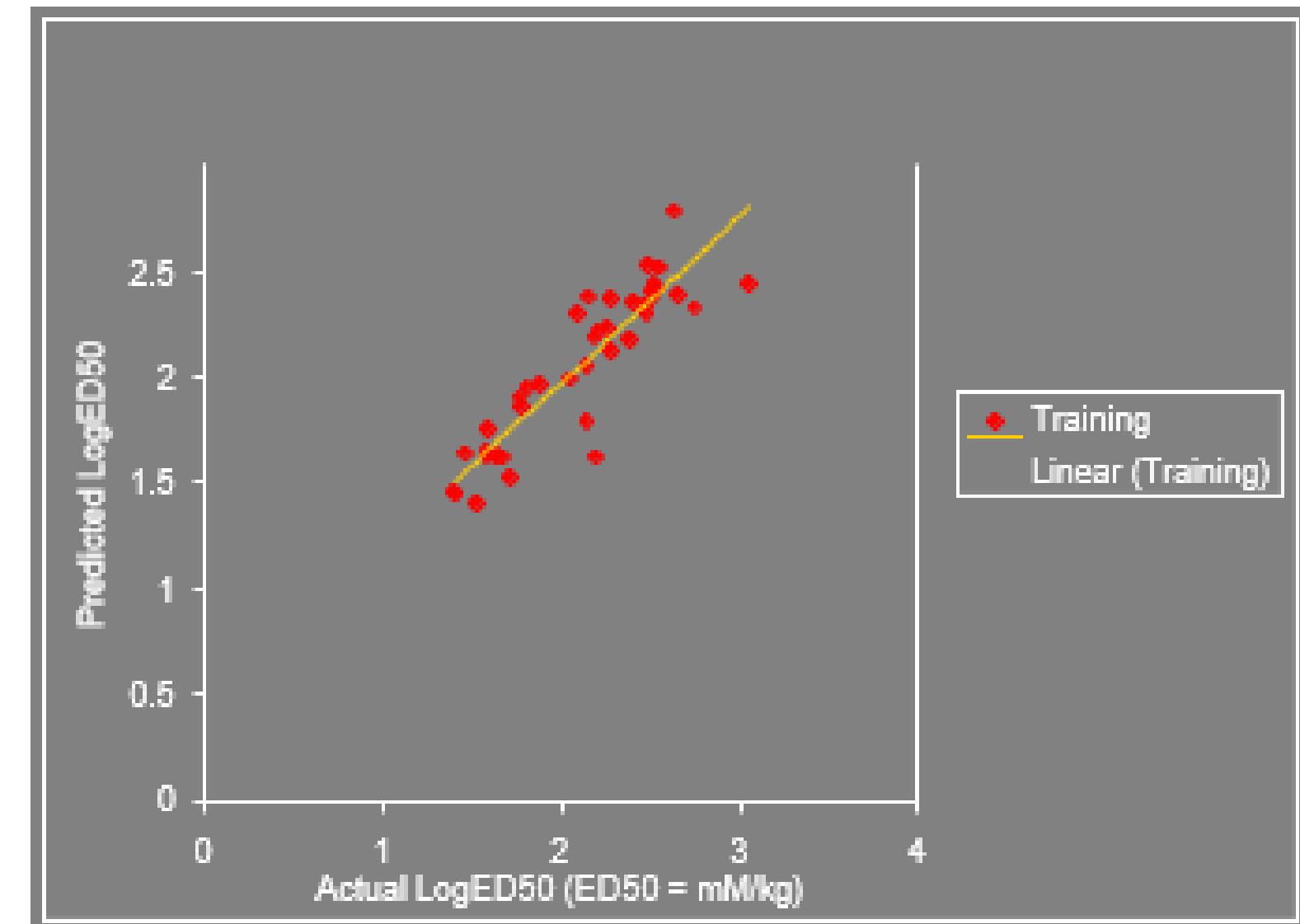


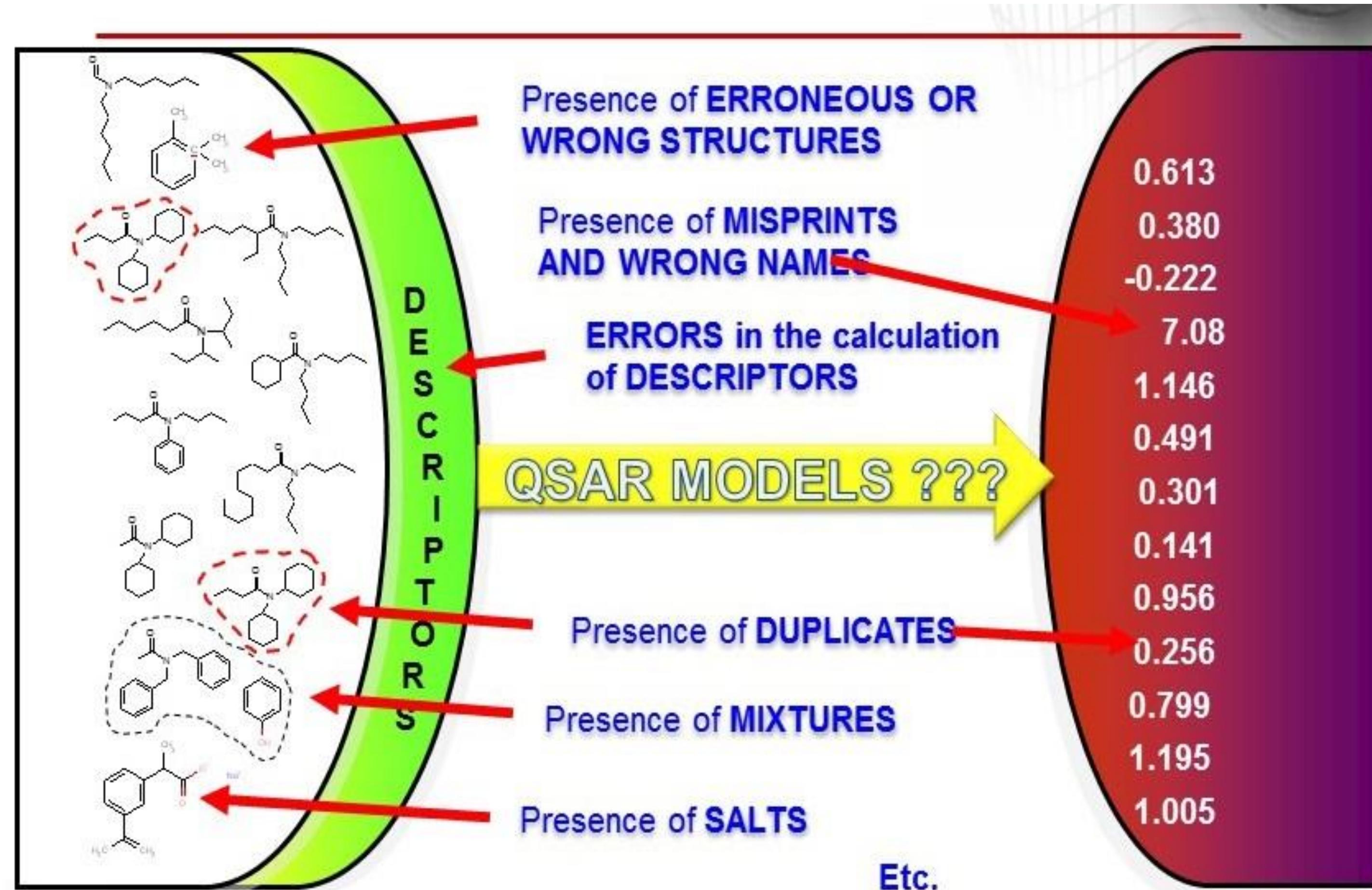
Establish quantitative relationships between descriptors and the target property capable of predicting activities of novel compounds

Chemistry	Bioactivity (IC ₅₀ , K _d ...)	Cheminformatics (Molecular Descriptors)				
Comp.1	Value1	D ₁	D ₂	D ₃		D _n
Comp.2	Value2	"	"	"		"
Comp.3	Value3	"	"	"		"
Comp.N	ValueN	"	"	"		"



$BA = F(D)$ (linear,
e.g., $-\text{LogIC}50 = k_1D_1 + k_2D_2 + \dots + k_nD_n$) or non-linear, e.g. k nearest
neighbors





Sources of Molecular Structures

Structures of molecules can be determined experimentally or predicted computationally. For small volatile molecules, such as carbon monoxide, and water, analysis of rotational lines in the microwave or infrared spectra of gaseous samples provides very accurate geometries. For larger volatile molecules, gas electron diffraction data are commonly used in structure determination, typically in conjunction with analysis of rotational spectrum. As a result of such efforts, bond lengths of nearly every stable diatomic and triatomic molecule are known with accuracy better than 0.005 angstroms. Experimental structures of many tetra-atomic and larger molecules are also known with similarly good accuracy. Structure determination of larger molecules in the gas phase becomes challenging as the spectra become more crowded and weaker due to low vapor pressure of larger molecules. The analysis is further complicated by the presence of multiple conformations in flexible molecules. However, structures of most organic molecules and of many biological macromolecules can be determined accurately in the solid state using X-ray or neutron diffraction.

Computational chemistry could, in principle, give accurate molecular structures for molecules of any size. In practice, however, accuracy comes with a steep price in computational time, and quantum chemical computations that provide bond lengths with accuracy better than 0.005 angstroms are only feasible with small molecules. Such rigorous methods are not routinely applicable for larger molecules, such as nucleotides, peptides, and many drug molecules. In such cases, one can use approximate computations that may yield structures with useful accuracy. Unfortunately, the approximate methods may also produce erroneous results. For example, a rather popular computational scheme designated as B3LYP/6-311G* overestimates the bond length of diatomic chlorine by 0.07 angstroms.

Molecular Structures via SMILES Servers

Several web sites generate 3D molecular structures from the SMILES string using the program CORINA. CORINA uses built-in tables of standard bond lengths and angles to create a reasonable model for small or rigid molecules. However, the model geometry for larger and flexible molecules is likely to be quite different than the most prevalent geometry in aqueous solution. One such site that generates 3D model structures is Online SMILES Translator by National Institutes of Health. Practice creating a molecule of ethanol using this service. Start the Structure Editor and sketch an ethanol molecule consisting of two bonds and oxygen; do not worry about hydrogens at carbons. Hit Submit Molecule and notice that the SMILES string for ethanol was generated. On the right-side panel, select PDB and 3D and hit Translate. Right-click on the link to download the molecule and save it as ethanol.pdb into your directory. Examine the file with text editor.

Sketching 2D Molecular Structures

Molecular Structure Files

Molecular structures can be stored in computer files. Minimally, one needs to define the atom type and its Cartesian coordinates to uniquely describe molecule's structure. However, largely due to historic reasons, we currently have several different file formats that represent the same structural data slightly differently.

Most chemists are well familiar with drawing 2D molecular structures and several programs allow effortlessly draw 2D representations of three-dimensional molecules. Two of the most popular 2D chemical diagram editors for Windows and Mac OS systems are ChemDraw from CambridgeSoft and MDL Draw from Elsevier MDL. Students can download a fully functional free chemical drawing program MDL Isis/Draw from Elsevier MDL website after registration. Some chemical drawing tools allow generation and export of 3D coordinates of the drawn molecule. The JME Molecular Editor allows to sketch simple molecules on-line and export these structures into the SMILES string. In more modern times various model building programs help in this task, and programs such as [Babel](#) can convert between different formats.

- Within the Born-Oppenheimer approximation, the geometry of a molecule at zero absolute temperature corresponds to the minimum of the total energy.
- The process of finding the coordinates that minimize the energy is known in the trade as *geometry optimization*. It is usually the first step in the calculation.

Energy minimization procedures

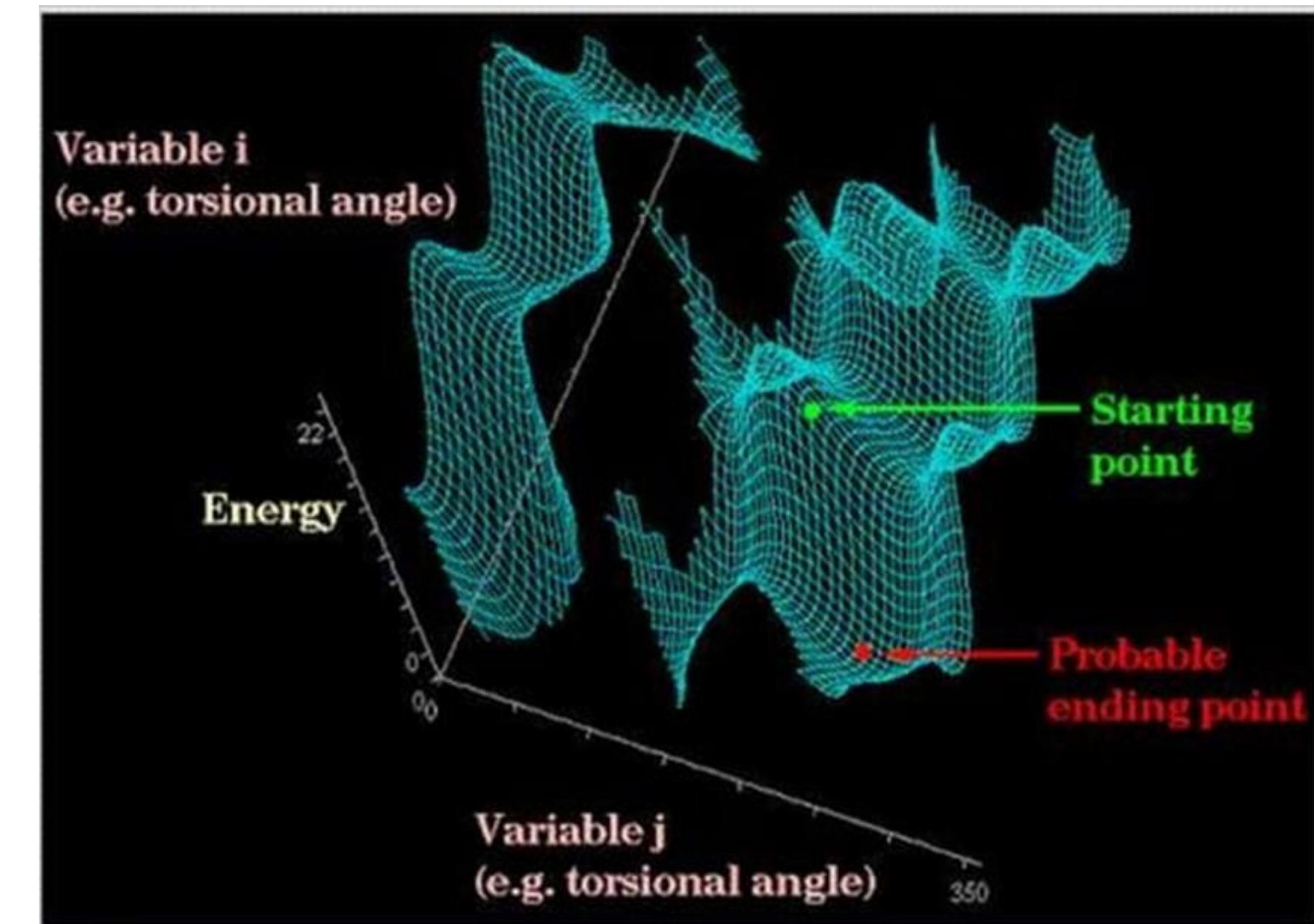
1. Conformational energy searching
2. Energy minimization
3. Minimisation algorithms

Conformational energy searching

- Energy is a function of the degrees of freedom in a molecule: bond, angle, dihedrals
- Conformational energy searching is used to find all of the energetically preferred conformations of a molecule
- This is mathematically equivalent to locating all of the minima of the energy function of the molecule
- The possible conformations for a molecule lie on an n-dim lattice, with n being the number of degrees of freedom.

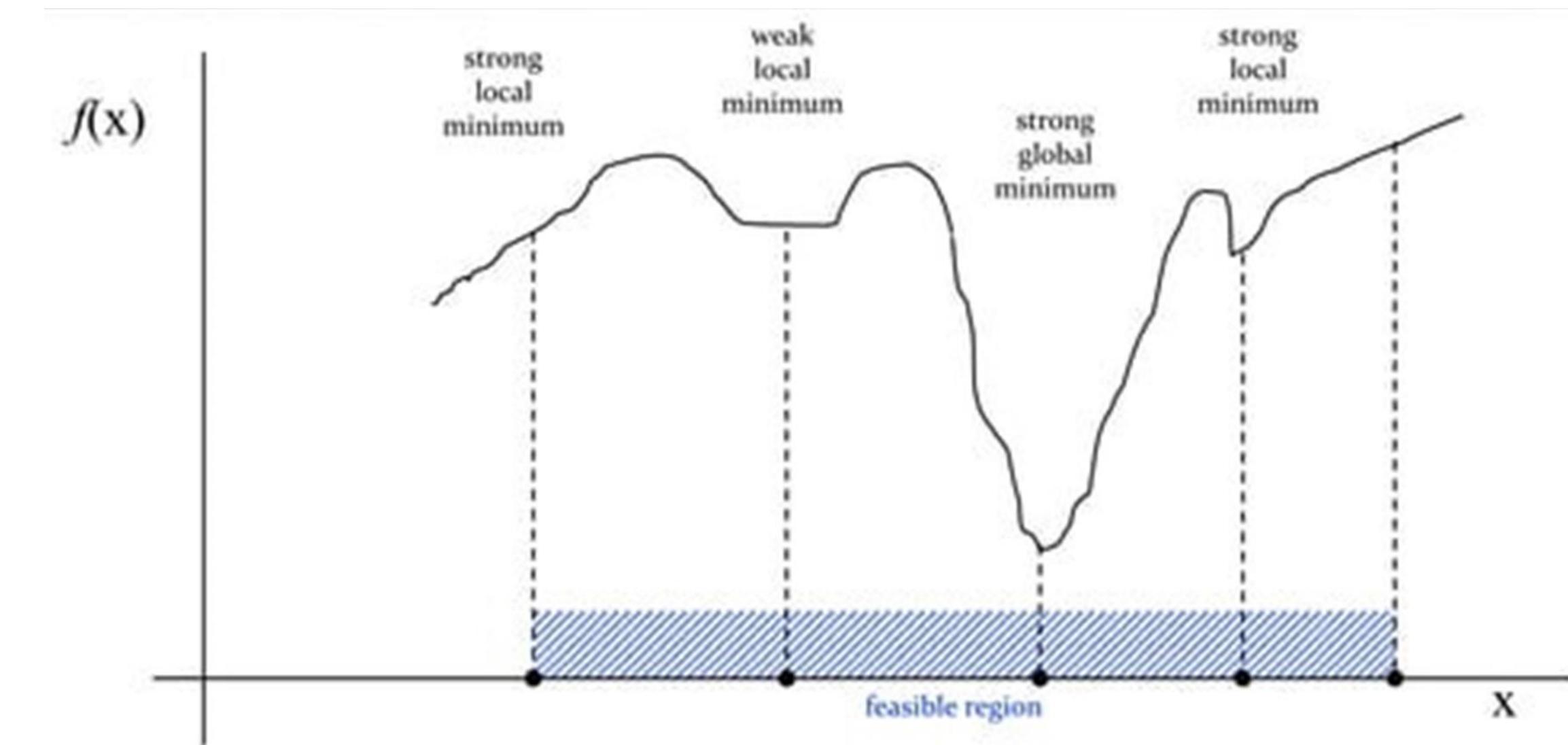
Energy Minimization

- Minimization algorithms are designed to head down-hill towards the nearest minimum
- Remote minima are not detected because this would require some period of up-hill movements.
- Minimization algorithms monitor the energy surface along a series of incremental steps to determine a down-hill direction



Minimization Algorithm

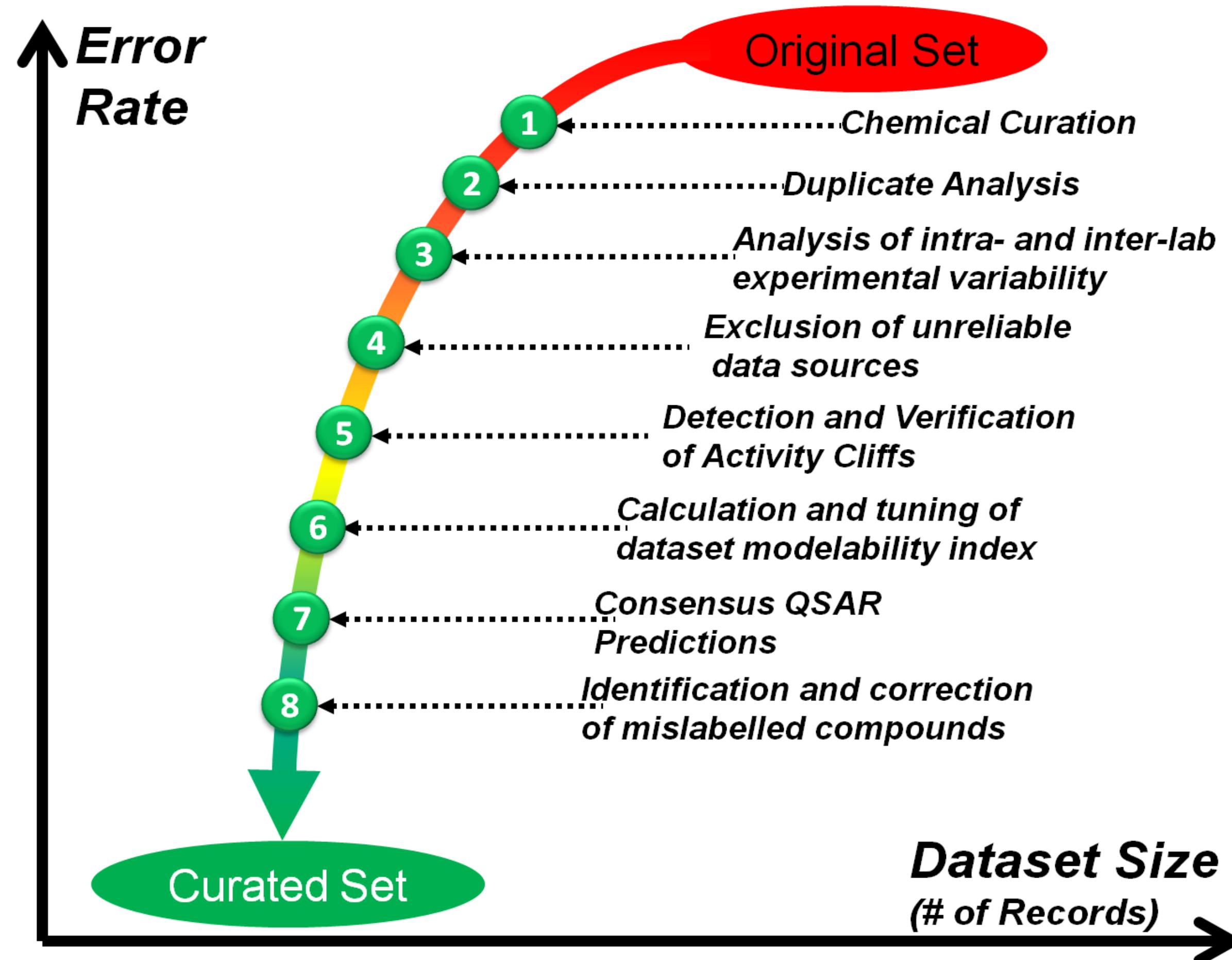
- Simplex algorithm
- Steepest descent minimizer
- Conjugate gradient and Powell minimizer
- Newton-Raphson and BFGS minimizer



Which of the minima is found depends on the starting point



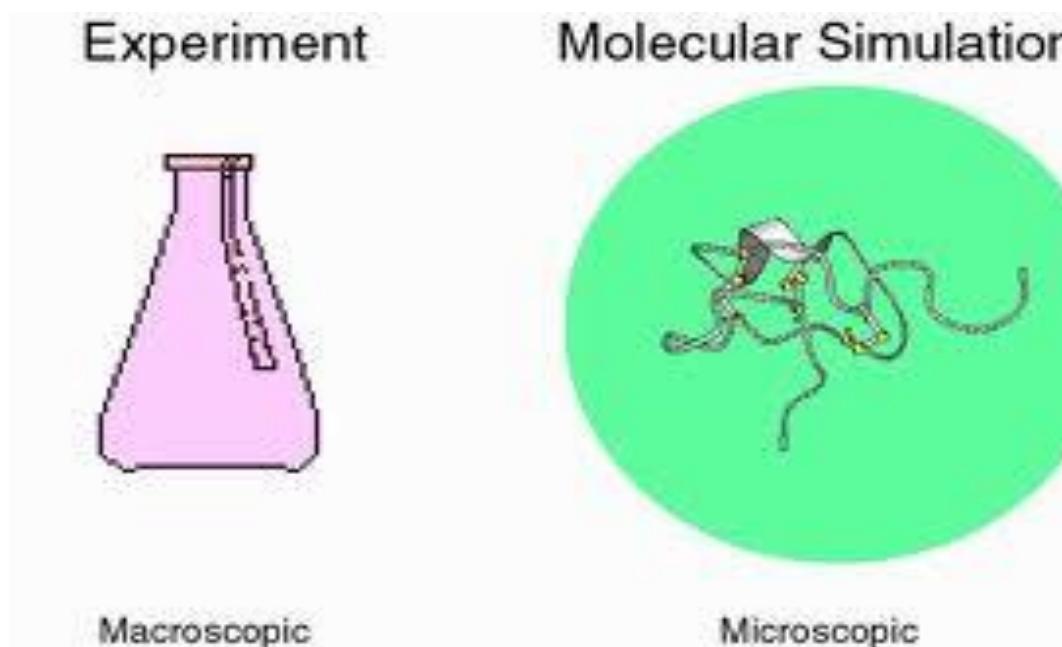
Thank you



How can we study the molecular properties ?

Graph Theory : Topology

Molecular Dynamics : A computer simulation of physical movement of atoms and molecules.



Computer simulation act as a bridge between microscopic length and time scales and the macroscopic world of the laboratory. A computer simulation of physical movement of atoms and molecules. This computational method calculates the time dependent behaviour of a molecular system.

In MD simulations, one follows the trajectories of N particles according to Newton's equation of motion:

$$m_i \frac{d^2}{dt^2} \mathbf{r}_i = \mathbf{F}_i, \quad \mathbf{F}_i = -\nabla_i U(\mathbf{r}_i, \dots, \mathbf{r}_N)$$

where $U(\mathbf{r}_1, \dots, \mathbf{r}_N)$ is the potential energy function consisting of bonded and non-bonded interactions.

The *Hückel matrix* in the Hückel theory of conjugated systems considered as the augmented vertex-adjacency matrix.

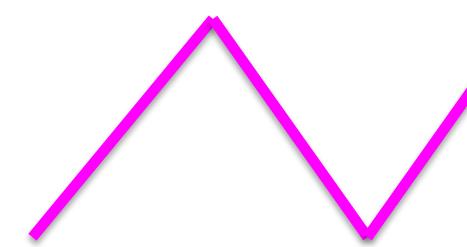
The eigenvalue equation for a matrix A can be expressed as $A - \lambda I = 0$

The solution to this equation are the eigenvalues $\lambda_i = a_{i,i}$ ($i = 1, 2, \dots, n$)

The delocalization energy is the extra stabilization that comes from letting the electrons spread over the whole molecule.

$$\text{---} \quad \begin{vmatrix} X & 1 \\ 1 & X \end{vmatrix} = X^2 - 1 \quad (\pm 1)$$

$$E_{\Pi} = 2(\alpha + \beta)$$



$$\begin{vmatrix} X & 1 & 0 & 0 \\ 1 & X & 1 & 0 \\ 0 & 1 & X & 1 \\ 0 & 0 & 1 & X \end{vmatrix} = X^4 - 3X^2 + 1 \quad (\pm 1.618, \pm 0.618)$$

$$E_{\Pi} = 2(\alpha + 1.618\beta) + 2(\alpha + 0.618\beta) = 4\alpha + 4.472\beta$$

$$\text{delocalization energy} = (4\alpha + 4.472\beta) - 2(2\alpha + 2\beta) = 0.472\beta$$

Ordering of Octane Isomers and Topological Indices

- The eigenvalues of eighteen octane isomers are conjugated ($\pm \lambda_1, \pm \lambda_2, \pm \lambda_3, \pm \lambda_4$)
- The number of terminal methyl groups are considered for subgrouping of the isomers
- The maximum eigenvalue of graph depicts the branching character
- The ordering of isomeric molecule and generation of topological descriptors done using eigenvalues

A – (- λ_1, λ_4)

B – (- λ_2, λ_4)

C – (- λ_3, λ_4)

D – (- λ_4, λ_4)

E – (- λ_5, λ_4)

F – (- λ_6, λ_4)

Two descriptor designed as $(\lambda_a + \lambda_b)$ & $(\lambda_a * \lambda_b)$

Swarna M Patra, Department of Chemistry

Hückel Matrix

The *Hückel matrix* in the Hückel theory of conjugated systems considered as the augmented vertex-adjacency matrix.

In the Hückel theory, the secular equation solved for a given conjugated system is given by

$$\det | -e_i \mathbf{S} + \mathbf{H} | = 0; i=1, \dots, V$$

Where \mathbf{H} is the Hamiltonian matrix, \mathbf{S} is the overlap matrix, e_i is the set of eigenvalues and V is the number of π -electrons in a conjugated molecule. The Hamiltonian matrix and the overlap matrix in the Hückel theory are simplified by using the set of approximations known in the quantum-chemical literature as the Hückel approximations

where \mathbf{I} is the unit $V \times V$ matrix,

$$\mathbf{H} = \alpha \mathbf{I} + \beta \mathbf{A}, \mathbf{S} = \mathbf{I}$$

α is an atomic parameter (atomic Coulomb integral) and

β is a bond parameter (resonance integral).

The eigenvalue equation for a matrix A can be expressed as $\mathbf{A} - \lambda \mathbf{I} = 0$
The solution to this equation are the eigenvalues $\lambda_i = a_{i,i}$ ($i = 1, 2, \dots, n$)

$$\det(A - \lambda I) = 0.$$

$$A = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & a_{n,n} \end{bmatrix}, \quad \det(A - \lambda I) = \det\left(\begin{bmatrix} a_{1,1} & 0 & \dots & 0 \\ 0 & a_{2,2} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & a_{n,n} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\right)$$

$$\begin{aligned}
 &= \det \begin{bmatrix} a_{1,1} - \lambda & 0 & \dots & 0 \\ 0 & a_{2,2} - \lambda & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & a_{n,n} - \lambda \end{bmatrix} \\
 &= (a_{1,1} - \lambda)(a_{2,2} - \lambda) \cdots (a_{n,n} - \lambda) = 0
 \end{aligned}$$



https://people.chem.ucsb.edu/kahn/kalju/chem126/public/molecules_getting.html

Thin Film Deposition Facility @ IDRRC RVCE



Thermal Evaporation

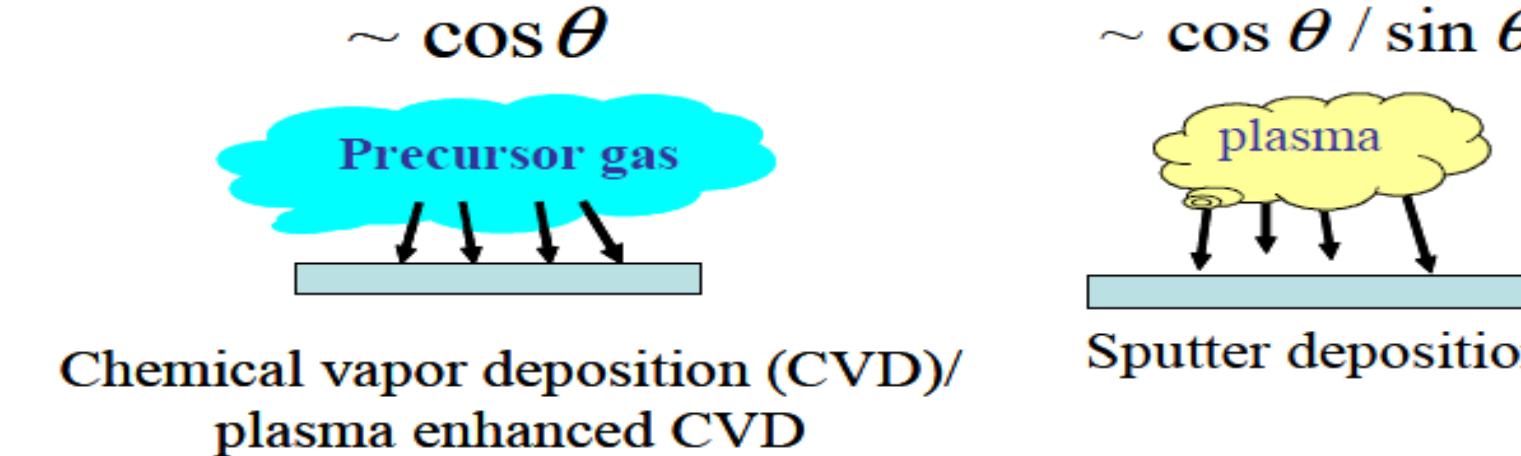
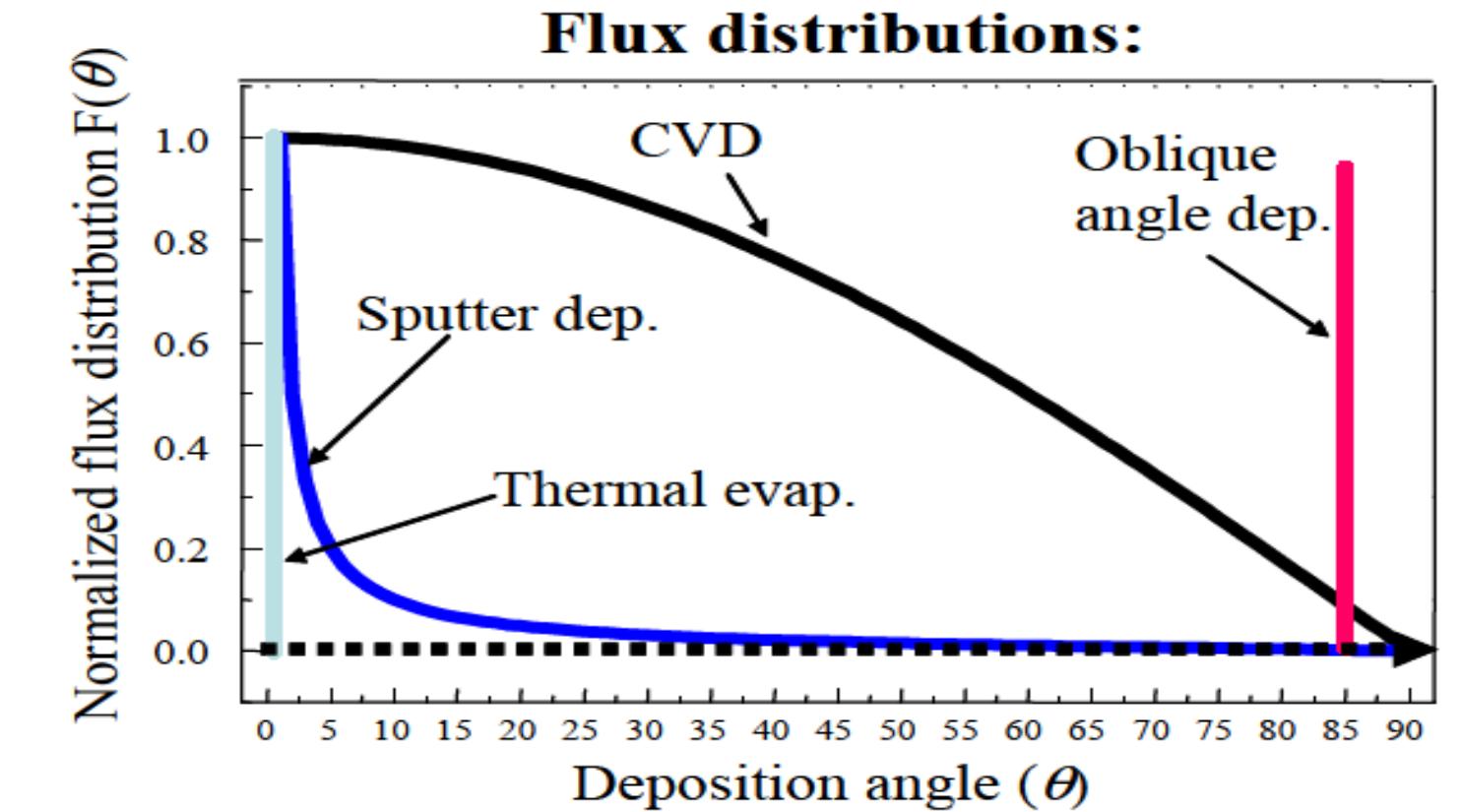
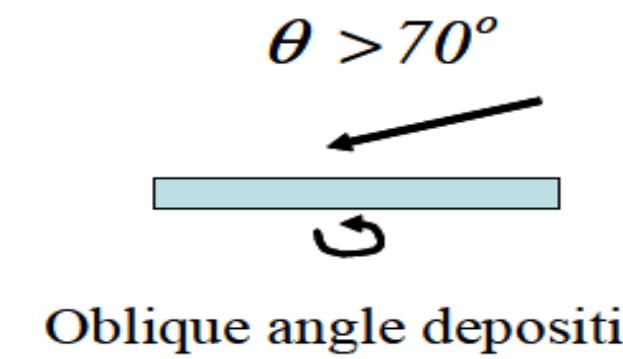
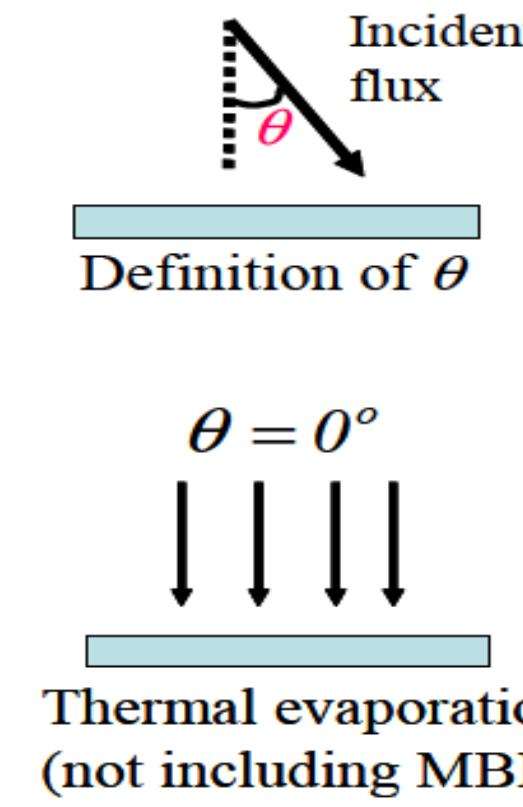


Sputtering



Plasma Enhanced Chemical Vapour Deposition

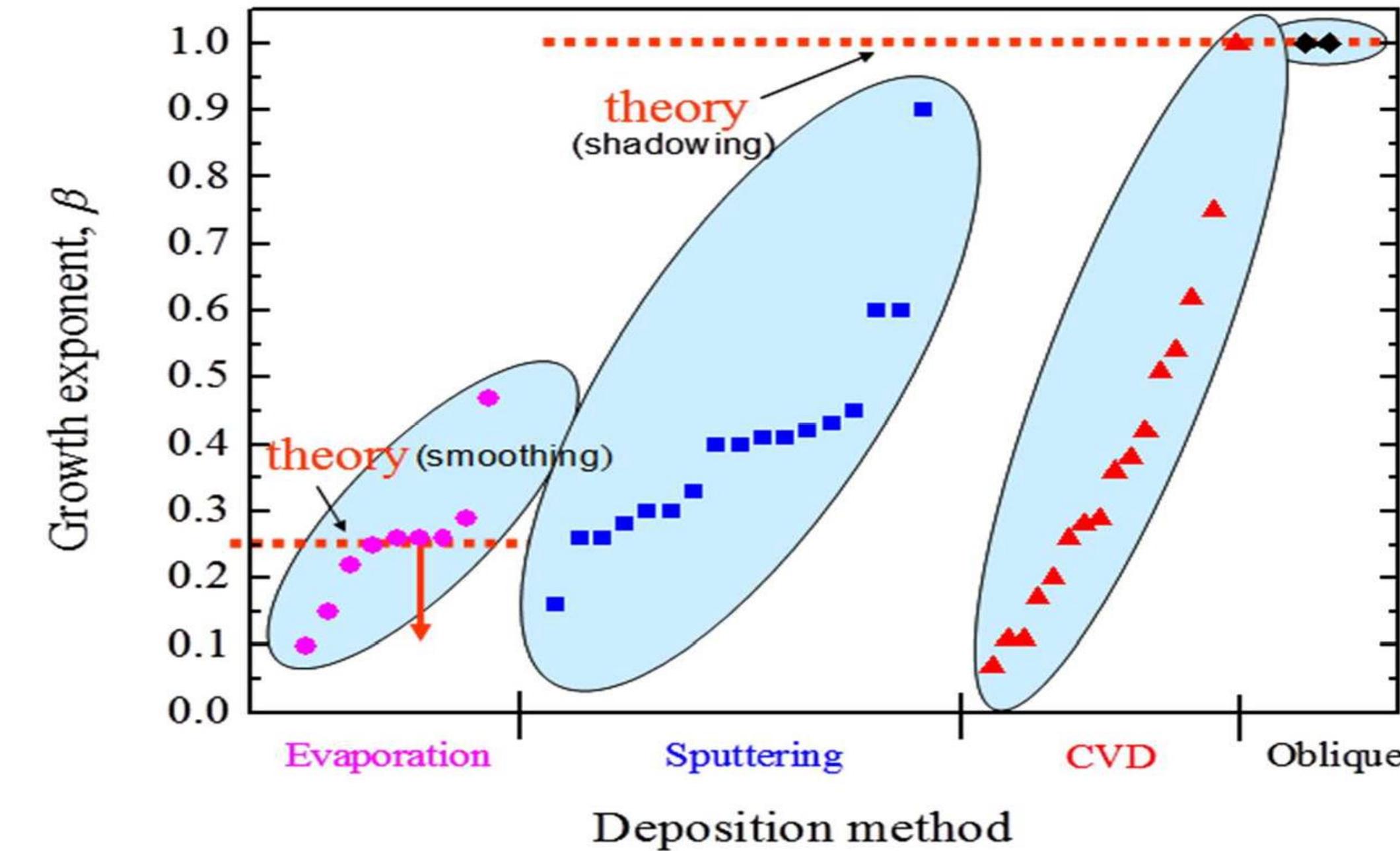
Commonly Used Thin Film Deposition Techniques



	0°	30°	45°	60°	90°
$\sin \theta$	0	$\frac{1}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{\sqrt{3}}{2}$	1
$\cos \theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	$\frac{1}{2}$	0
$\tan \theta$	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$	Not defined
cosec θ	Not defined	2	$\sqrt{2}$	$\frac{2}{\sqrt{3}}$	1
sec θ	1	$\frac{2}{\sqrt{3}}$	$\sqrt{2}$	2	Not defined
cot θ	Not defined	$\sqrt{3}$	1	$\frac{1}{\sqrt{3}}$	0

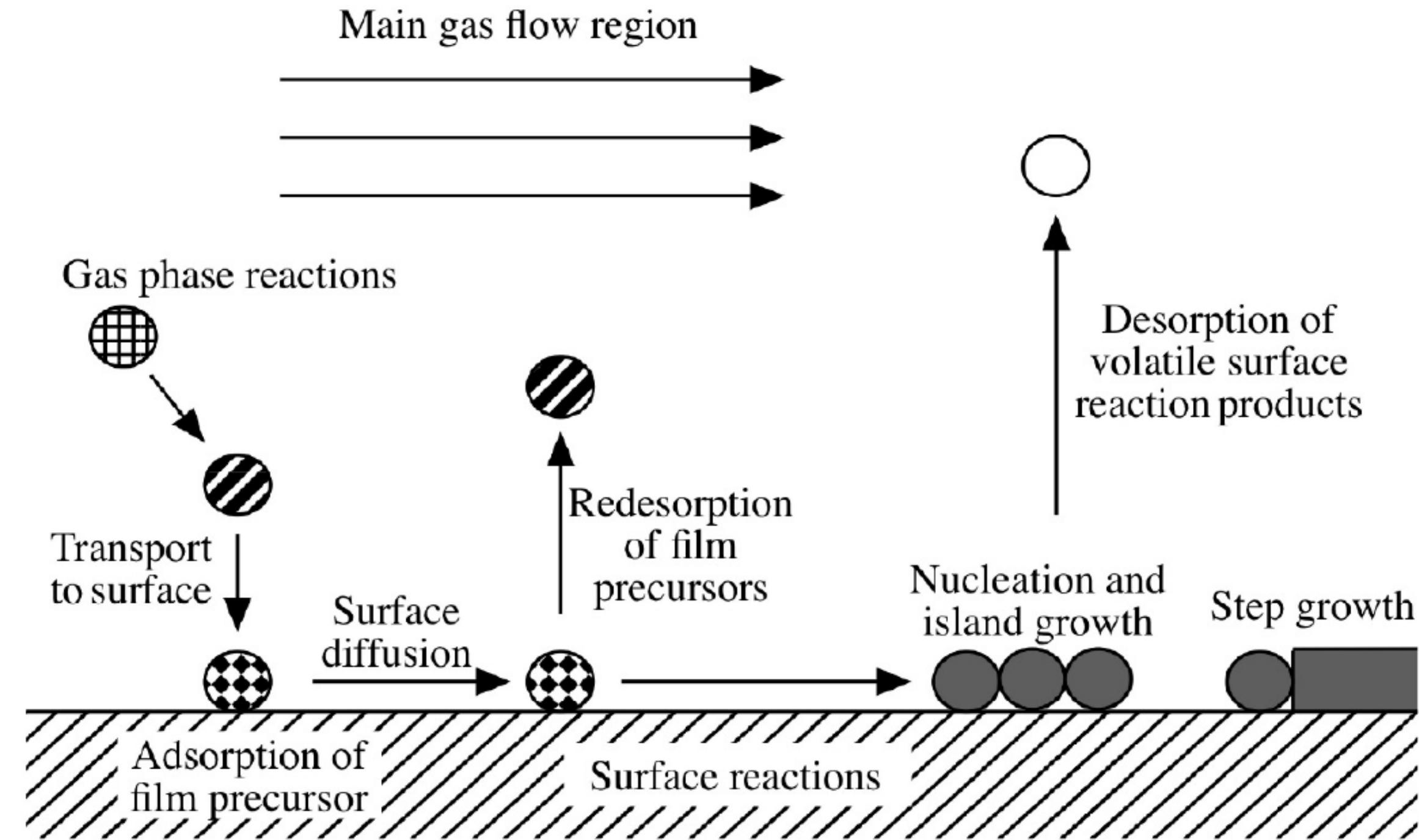
The ‘ θ ’ is defined as the angle between the surface normal and the direction of the incidence beam of atoms. The atoms that are deposited on the lattice surface and the path they take during the process can be depicted using graph theory. Various properties of these atoms are used to form the topology of the network to study various properties of thin film growth.

Growth Exponent For Thin Film Deposition Techniques



- Experimentally obtained values of growth exponent reported in the literature for different deposition techniques is compared to the predictions of common thin film growth models in dynamic scaling theory.
- Root-mean-square roughness (RMS) grows as a function of time in a power law form, where β is the “growth exponent” ranging between 0 and 1. $\beta = 0$ for a smooth growth front and $\beta = 1$ for a very rough growth front.

Fundamental of CVD transportation & Reaction Processes



Graph Theory → Network Statistics → Thin Film Growth

Shadowing Effect - Particles that approach the surface at an angle can preferentially land on higher surface points before reaching the target point.

Sticking Coefficient - The particles can stick to the surface with a probability. Lower sticking coefficients particles go through higher number of re-emissions and can connect many surface points.

Adatom - The incident particle is deposited onto the surface.

Clusters - Particles that do not stick are re-emitted and can travel to another surface point before settling down at a point on the lattice. These particles later take form into various clusters representing the hills and valleys of the rough surface.

Dynamic morphology is dependent on sticking coefficient, angle of incidence of particle, re-emission, and shadowing effect.

Thickness of substrate is proportional to growth time.

Network Models

- Enhance the topology of the network of thin film growth by treating *clusters* forming on the film surface as different nodes and establish a different “cluster-based” network model of the growth.
- A network can be compared to its initial stage and after removing some prominent nodes or hubs removed from the network.
- Better understanding the behavior of ad-atoms movement on the surface of a thin film substrate.
- Better understand the shadowing and smoothing effects of the thin film. Understanding could be used to better engineer the thin film growth
- Observations show the cluster-based model is better suited to study global behaviors of thin film growth and particularly the shadowing phenomenon

Network Behavior in Thin Film Growth Dynamics

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science and Engineering By
Satish K. Badepalli, 2017, University of Nevada, Reno

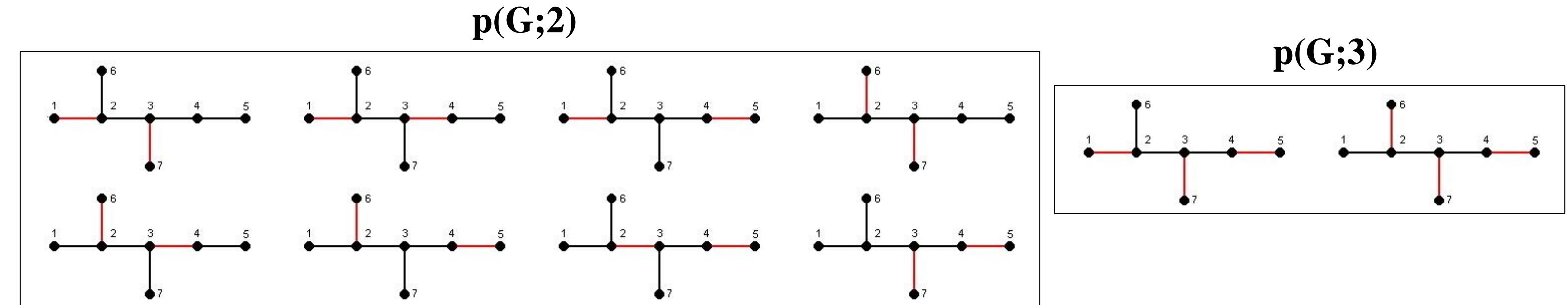
Dynamic scaling theory and Modelling

- Conventional statistical methods cannot be used to describe the complex phenomenon of surface morphology formation in thin film growth due to its intractability.
- There have been numerous modeling and experimental proposed based on dynamic scaling analysis.
- Experimental results and predictions of growth models have varied considerably. Experimentally reported values of growth exponent β are far from agreement with the predictions of these growth models
- The Monte Carlo simulations reproduced the morphologies and exponents predicted by the well-known Kardar-Parisi-Zhang (KPZ) model in dynamic scaling theory, for the case where small sticking coefficient particles were re-emitted multiple times and resulted in a conformal growth.
- Like experiments, it was not possible to capture a “universal” growth behavior using Monte Carlo simulation.
- Graph theory allows researchers to model the systems in to networks and study their characteristics both locally and globally.

Hosoya Topological Index Z is obtained by counting the k disjoint edges in a graph. Z counts all sets of non-adjacent bonds in a structure. where N is the number of vertices in the molecular graph and $p(G;i)$ is the number of selections of i mutually non-adjacent edges in G. By definition, $p(G;0) = 1$, and $p(G;1)$ is the number of edges in G.

Example: 2, 3-dimethylpentane

$$\begin{aligned} p(G;0) &= 1 \\ p(G;1) &= 6 \\ p(G;2) &= 8 \\ p(G;3) &= 2 \end{aligned}$$



The Hosoya Index of G:

$$Z(G) = p(G;0) + p(G;1) + p(G;2) + p(G;3) = 1 + 6 + 8 + 2 = 17$$

Connectivity Index

(Milan Randic, A.T. Balaban)

$$\chi = \sum (R_i R_j)^{-1/2}$$

χ is constructed from the row sums R_i and R_j of the adjacency matrix using the algorithm $(R_i R_j)^{-1/2}$ for the contribution of each bond (i,j)

χ is a bond additive quantity where terminal CC bonds are given greater weight than inner CC bonds.

$$\frac{1}{\sqrt{(1.3)}} = \frac{1}{1.732} = 0.577$$

$$\frac{1}{\sqrt{(3.3)}} = \frac{1}{3} = 0.333$$

$$\frac{1}{\sqrt{(3.2)}} = \frac{1}{2.449} = 0.408$$

$$\frac{1}{\sqrt{(1.2)}} = \frac{1}{1.414} = 0.707$$



The Randic Index of G:

$$c(G) = 1 * 0.7071 + 3 * 0.5774 + 1 * .4082 + 1 * .3333 = 3.1807$$