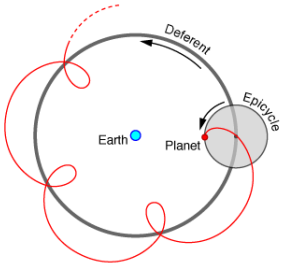
BIO311 note

(Modeling for Computational Biology)

Sequence	Method	Assessment Type (EXAM or CW)	Learning outcomes assessed (Use codes under learning outcomes.)	Duration	Week	% of final mark	Resit (Y/N/S)
#001	Assignment 1	CW	ABCDEF			25	N
#002	Assignment 2	CW	GH			15	N
#003	Written examination	EXAM	ALL	3 hour(s)		60	N

1 Kinematics

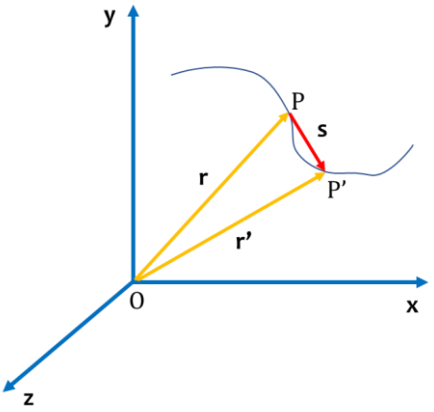


1.1 Definition of Kinematics

- A branch of dynamics that deals with aspects of motion apart from considerations of mass and force.
- It concerns the description of the motion of a point (body) or a system of points (bodies) without considering the forces that cause the motion.

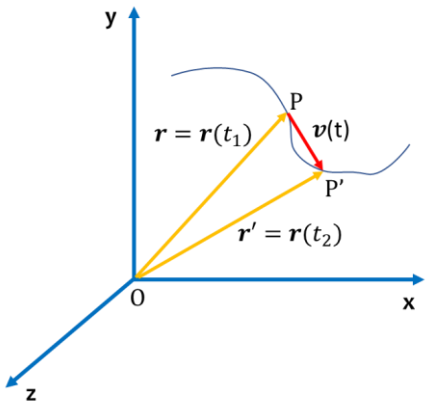
1.2 Position and reference systems

- Origin (O): The point in space that serves as the zero point for the coordinate system. All positions are measured relative to this point.
- Oriented axes: Imaginary lines that intersect at the origin and extend in specific directions. In a three-dimensional space, there are typically three axes (x, **y**, z) that are perpendicular to each other. The axes are oriented vectors with positive and negative directions.
- Scale: A scale defines how distances are measured along the axes. It determines the units of measurement.



- The position of a point P is completely determined by the position vector $\boldsymbol{r} = \boldsymbol{OP} = (x, y, z)$.
  Since our object moves in space, after some time it will be in a new position $\boldsymbol{r'} = \boldsymbol{OP'} = (x', y', z')$.
- The difference between the two position vectors after a certain time:
  $\boldsymbol{s} = \boldsymbol{\Delta r} = \boldsymbol{OP'} - \boldsymbol{OP} = (x' - x,\ y' - y,\ z' - z) = (\Delta x, \Delta y, \Delta z)$

1.3 Velocity



- As the point P moves in time, the vector position **r** is a function of time: $\boldsymbol{r} \equiv \boldsymbol{r}(t) = (x(t), y(t), z(t))$
- The average velocity $\boldsymbol{v}_{ave}$ in a time interval is defined as the displacement vector divided by the time interval in which the displacement happens:

$$\boldsymbol{v}_{ave} = \frac{\Delta \boldsymbol{r}(t)}{\Delta t} = \frac{\boldsymbol{r}(t+\Delta t)}{\Delta t} = \frac{\boldsymbol{s}}{\Delta(t)} = \frac{(\Delta x, \Delta y, \Delta z)}{\Delta t} = \frac{(x(t_2)-x(t_1), y(t_2)-y(t_1), z(t_2)-z(t_1))}{t_2 - t_1}$$

- The *velocity* (or *instant velocity*) $\boldsymbol{v}(t)$ is the time derivative of the position vector $\boldsymbol{r}(t)$: $\boldsymbol{v}(t) = \lim_{\Delta t \to 0} \frac{\Delta \boldsymbol{r}(t)}{\Delta t} = \frac{d\boldsymbol{r}(t)}{dt} = \left(\frac{dx(t)}{dt}, \frac{dy(t)}{dt}, \frac{dz(t)}{dt}\right) = (v_x(t), v_y(t), v_z(t))$

  So, the velocity of a point is a vector, whose components are the time derivatives of the components of the position

vector.

1.4 Uniform linear motion

- A body moving with constant velocity **v** is said to move according to a uniform linear motion.
- The law of motion for such body can be obtained by integrating both sides of the velocity equation using the time as independent variable:

$$\boldsymbol{v}(t) = \frac{d\boldsymbol{r}(t)}{dt}$$

$$\int_t^{t+\Delta t} \boldsymbol{v}dt' = \int_t^{t+\Delta t} \frac{d\boldsymbol{r}(t)}{dt}dt'$$

$$\boldsymbol{v}\Delta t = \boldsymbol{r}(t+\Delta t) - \boldsymbol{r}(t)$$

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boldsymbol{v}\Delta t$$

$$\boldsymbol{v} = (v_x, v_y, v_z)$$

$$\begin{cases} x(t+\Delta t) = x(t) + v_x\Delta t \\ y(t+\Delta t) = y(t) + v_y\Delta t \\ z(t+\Delta t) = z(t) + v_z\Delta t \end{cases}$$

1.5 Acceleration

- The acceleration a(t) is the time derivative of the velocity v(t):

$$\boldsymbol{a}(t) = \lim_{\Delta t \to 0} \frac{\Delta \boldsymbol{v}(t)}{\Delta t} = \frac{d\boldsymbol{v}(t)}{dt} = \left(\frac{dv_x(t)}{dt}, \frac{dv_y(t)}{dt}, \frac{dv_z(t)}{dt}\right) = (a_x(t), a_y(t), a_z(t))$$

- Since the velocity is the time derivative of the position, the acceleration is the second derivative of the position with respect to time:

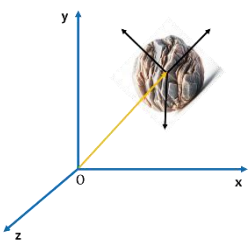
$$a(t) = \frac{d^2\boldsymbol{r}(t)}{dt^2} = \left(\frac{d^2x(t)}{dt^2}, \frac{d^2y(t)}{dt^2}, \frac{d^2z(t)}{dt^2}\right)$$

- The acceleration is a vector, whose components are the time derivatives of the components of the velocity vector.
- So here we put the Uniform linear motion equation into position related version:

$$\begin{cases} x(t+\Delta t) = x(t) + v_x\Delta t = x(t) + v_x(t)\Delta t + \frac{1}{2}a_x(\Delta t)^2 \\ y(t+\Delta t) = y(t) + v_y\Delta t = y(t) + v_y(t)\Delta t + \frac{1}{2}a_y(\Delta t)^2 \\ z(t+\Delta t) = z(t) + v_z\Delta t = z(t) + v_z(t)\Delta t + \frac{1}{2}a_z(\Delta t)^2 \end{cases}$$

1.6 Degrees of freedom and configuration space

- Degrees of freedom are the variable that needs to be known to describe a (static) system completely. A single point in 3-dimension has 3 degrees of freedom, and they are usually associated with the three components of its position vector (*x*, *y*, *z*).
- A system composed of N bodies in 3-dimension has 3N degrees of freedom and can be described by a vector in a 3N-dimensional space, which is called the configuration space: $\boldsymbol{r} = (x_1, y_1, z_1, x_2, y_2, z_2, \ \ldots, x_N, y_N, z_N)$
- Due to constraints (chemical bounds, physical delimitations, etc.), the accessible space is a subset of the 3N-dimensional space.
- For example, for a rigid body, the relative positions between the different parts of the system do not change with time. A rigid body has 6 degrees of freedom (i.e. 3 coordinates for its position and three angles that define its orientation in space) (Translation degrees of freedom & Rotation degrees of freedom).

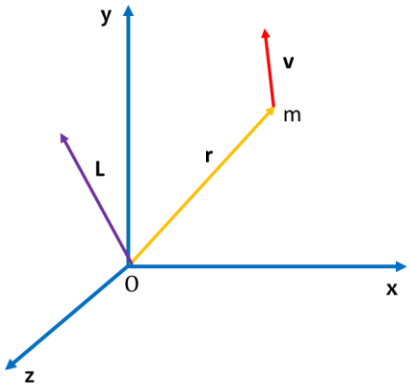


1.7 Momentum

- We define the momentum of an object of mass m moving with velocity v as the product between the mass and the velocity:

$$\boldsymbol{p} = m\boldsymbol{v}$$

- Then the total momentum for a system of N particles is the sum of the momenta of each individual particle: $\boldsymbol{p_{TOT}} = \sum_{i=1}^{N} m_i \boldsymbol{v}_i$
- The total momentum of a system is conserved if the system is isolated.



1.8 Angular Momentum

- We define the angular momentum of an object of mass m moving with velocity v as the cross product between its position vector and its momentum: $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p} = \boldsymbol{r} \times m\boldsymbol{v}$
- Then the *total angular momentum* for a system of *N* particles is the sum of the angular momenta of each individual particle: $\boldsymbol{L_{TOT}} = \sum_{i=1}^{N} \boldsymbol{r} \times \boldsymbol{p}_i = \sum_{i=1}^{N} \boldsymbol{r} \times m_i \boldsymbol{v}_i$
- The total angular momentum of a system is conserved if the system is isolated.

1.9 The Phase Space

A dynamic system of N particles with masses mi can be described when positions and velocities (or better momenta) are completely known. Position and momentum vectors of the system are both 3N-dimensional vectors, so that whole system can be considered a point in a 6N-dimensional space:

$$\boldsymbol{r} = (x_1, y_1, z_1, x_2, y_2, z_2, \ \ldots, x_N, y_N, z_N)$$
$$\boldsymbol{p} = (p_{x1}, p_{y1}, p_{z1}, p_{x2}, p_{y2}, p_{z2}, \ \ldots, p_{xN}, p_{yN}, p_{zN})$$

As for the configuration space, due to constraints, the accessible space is a subset of the 6N-dimensional space. We call this subspace the phase space.

2 Dynamics: forces, work and energy

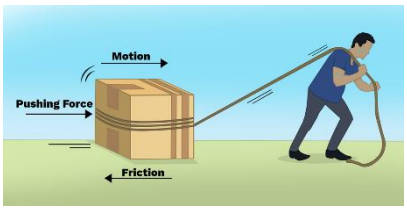
2.1 Newton's laws of motion

- A body remains at rest or in motion at a constant speed in a straight line, except insofar as a force act upon it.
- The acceleration of an object is directly proportional to the net force acting on it and inversely proportional to its mass.
- If two bodies exert forces on each other, these forces have the same magnitude but opposite directions.

$$\boldsymbol{F} = m\boldsymbol{a} = m\frac{d\boldsymbol{v}}{d\boldsymbol{t}} = \frac{d(m\boldsymbol{v})}{dt} = \frac{d\boldsymbol{p}}{dt}$$

2.2 Example of forces

- Gravity: $F = -G\frac{m_1 m_2}{r^2}$ and $\boldsymbol{F} = m\boldsymbol{g}$
- Electrostatic: $F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}$
- Friction: $F = -\mu N$



2.3 Work

We consider an object of mass m, moving according to a certain trajectory γ, under the action of external forces.

- **Definition**: we call elementary work done by the force F the quantity: $\delta W = \boldsymbol{F} \cdot d\boldsymbol{s}$

  (where ds is an infinitesimal vector displacement along γ)
- **Definition**: the work done by the force F along the trajectory γ is the integral of the elementary work along the curve γ:

$$W = \int_{\gamma} \boldsymbol{F} \cdot d\boldsymbol{s}$$

- **Example**: An object of mass *m* is left to fall with initial velocity v_0=0 from a height *h*. We want to calculate the work done by the force of gravity when the object reaches the ground.

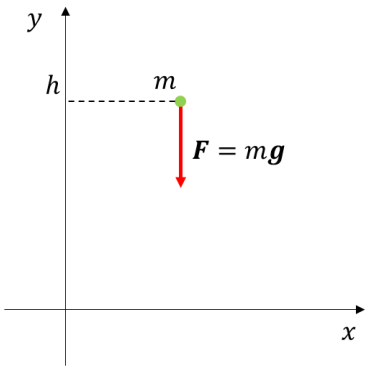
$$W = \int_{\gamma} \boldsymbol{F} \cdot d\boldsymbol{s} = \int_{\gamma} m\boldsymbol{g} \cdot d\boldsymbol{s} = \int_0^h (mg)ds = mg\int_0^h ds = mgh$$

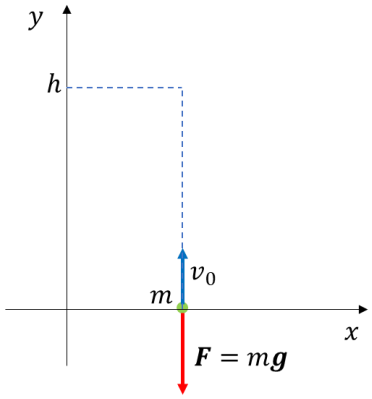
$F = mg$

g and ds are parallel and in the same direction.

The force of gravity does not depend on the position

So, the force of gravity is constant and it is always parallel and in the same direction of the motion.





Let us now consider the case in which the force and the direction of motion are antiparallel (opposite directions).
This is the case of an object leaving the ground with velocity $v_0 > 0$. We know that the object will reach a certain height h, where its velocity will be 0. At that point, it will start descending (as in the previous case).
And the work done in the ascending phase is:

$$W = \int_\gamma \boldsymbol{F} \cdot d\boldsymbol{s} = \int_\gamma m\boldsymbol{g} \cdot d\boldsymbol{s} = -\int_0^h (mg)ds = -mg\int_0^h ds = -mgh$$

(Which equals to the opposite of the work done by gravity in the descending phase)

2.4 Work and Kinetic Energy

Let us consider again the case of a falling object. We want to calculate the change in velocity when the object reaches the ground. From the equation of the uniformly accelerated motion:

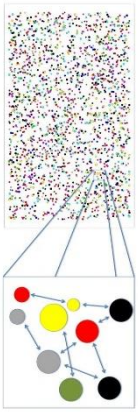
$$\begin{cases} v_f = v_0 + gt \\ h = v_0 t + \frac{1}{2}gt^2 \end{cases} \Rightarrow \quad t = \frac{v_f - v_0}{g} \quad \Rightarrow \quad h = \frac{1}{2g}(v_f^2 - v_0^2)$$

If we multiply each term of last equation for mg we obtain the following relationship:

$$mgh = \frac{1}{2}mv_f^2 - \frac{1}{2}mv_0^2$$

$$\because W = mgh \text{ and } K = \tfrac{1}{2}mv^2$$

$$\therefore W = K_f - K_i = \Delta K \quad \text{(Work-Energy theorem)}$$



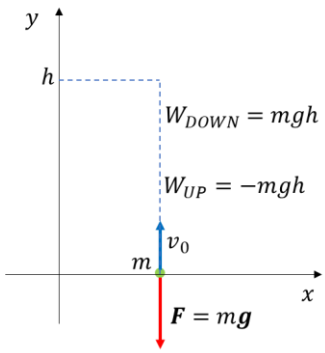
Multiple forces can act on an object at the same time. Forces sum linearly (as vectors). The total work done on the object is the sum of the work done by each force:

$$F_{NET} = F_1 + F_2 + \cdots + F_N = \sum_{i=1}^{N} F_i$$

$$W_{NET} = \int F_{NET} \cdot ds = \int (F_1 + F_2 + \cdots + F_N) \cdot ds = \int F_1 \cdot ds + \cdots + \int F_N \cdot ds = W_1 + \cdots + W_N$$

This means that the total change in kinetic energy is given by the sum of the work done by each of the forces acting on the object: $\Delta K = \sum_{i=1}^{N} W_i$

3 Potential energy and Conservation of energy



3.1 Work in closed circuits

- We have seen that if we launch an object from the ground, the work done by the force of gravity in the ascending path is equal and opposite of the work done in the descending part: $W_{UP} = -W_{DOWN}$
- When the object reaches the ground again, the net work done by the force of gravity is 0: $W_{NET} = W_{UP} - W_{DOWN} = 0$
- This can be extended to arbitrary circuits in space: The work done by the force of gravity

on a circuit is always null.

• For a generic circuit, we can break the trajectory in a sum of small displacements. At each step, we can decompose the displacement in two components, one parallel and one orthogonal to the force of gravity. But the result stays the same.

$$d\boldsymbol{s} = d\boldsymbol{s}_{\parallel} + d\boldsymbol{s}_{\perp}$$

$$W_{NET} = \int_{\gamma} m\boldsymbol{g} \cdot d\boldsymbol{s} = \int_{\gamma} m\boldsymbol{g} \cdot (d\boldsymbol{s}_{\parallel} + d\boldsymbol{s}_{\perp}) = \int_{\gamma} m\boldsymbol{g} \cdot d\boldsymbol{s}_{\parallel} + \int_{\gamma} m\boldsymbol{g} \cdot d\boldsymbol{s}_{\perp}$$

$$\int_{\gamma} m\boldsymbol{g} \cdot d\boldsymbol{s}_{\perp} = 0 \quad (\boldsymbol{g} \text{ and } d\boldsymbol{s}_{\perp} \text{ are orthogonal})$$

$$\int_{\gamma} m\boldsymbol{g} \cdot d\boldsymbol{s}_{\parallel} = mg \int_{y(A)}^{y(A)} dy = 0$$

$$W_{NET} = 0$$

3.2 Work along different paths

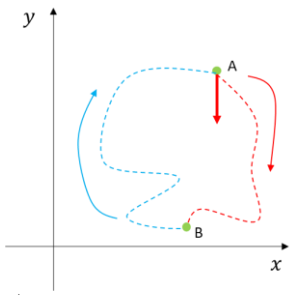
• Conclusion: The work done by the force of gravity to move from point A to point B does not depend on the chosen path.

• Suppose we are moving from point A to point B using the red path and come back to A using the cyan path. The net work done in the cycle is the sum of the two contributions and it is zero as we completed the cycle:

$$W_{NET} = W_{AB(R)} + W_{BA(C)} = 0$$

• On the other hand, the work done to go from point B to A (let's say on the cyan path) is equal and opposite to work done from A to B using the same path:

$$W_{NET} = W_{AB(R)} + W_{BA(C)} = W_{AB(R)} - W_{AB(C)} = 0 \Rightarrow W_{AB(R)} = W_{AB(C)}$$

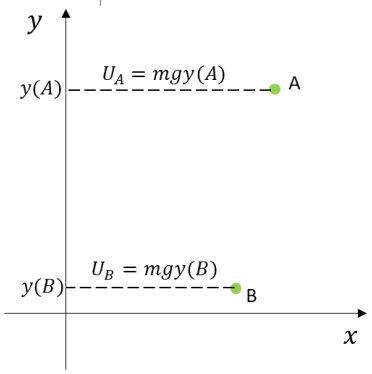


3.3 Potential Energy

Since the work done by the gravitational force on an object that moves from point A to point B, does not depend on the path, but only on the starting and ending point, we can write it as the difference of a function of the coordinates of A or B:

$$W_{AB} = U_A - U_B = -\Delta U$$

U is called potential energy (is a form of energy). It is a property of an object interacting with a force field, and it depends on the object's position (or orientation) in space. For the gravitational force, a good choice of the functional form of U is U=mgy, for any arbitrary origin of the reference system y=0.



3.4 Conservation of Energy

• The work-energy theorem tells us that the work done to an object by an arbitrary force causes a variation of the kinetic energy.

$$W_{AB} = U_A - U_B = K_B - K_A$$
$$K_A + U_A = K_B + U_B$$

• As A and B are arbitrary points in space, the relationship above is valid for any point in space. If we introduce the quantity $E_{TOT} = K + U$, the total energy of the object we have:

$$E_{TOT} = K + U = const$$

• This means that the total energy of an object moving under the effect of a gravitational field is always conserved.

3.5 Conservative forces

Conservative forces

- The work done is independent on the path.
- The work in a circuit is always zero.
- The potential energy can be defined.
- The total energy is conserved.
- Examples: gravitation, electromagnetic force, literally every fundamental force, harmonic forces.

Non-Conservative forces

- The work done depends on the path.
- The work in a circuit is not zero.
- The potential energy cannot be defined.
- The total energy cannot be defined.
- Examples: friction, viscous friction, other kind of frictions…

3.6 Relationship between Force and Potential Energy

• If we consider first the 1-dimensional case (force and displacement happens in the same line). From the definition of

elementary work we have: $\delta W = \boldsymbol{F} \cdot d\boldsymbol{s} = Fdx$

On the other hand, if the force is conservative, we can write: $\delta W = -dU$

By comparing the two equations we obtain the following relationship, we have: $F = -\frac{dU}{dx}$

- It's seeming that U is a function only of the space $U = U(x)$, this means that conservative forces do not depend on the velocity of the object on which they act and they do not change with time.
- In general, the potential energy does not depend on a single coordinate. In this case, the force is given by the opposite of the gradient of the potential energy: $\boldsymbol{F} = -\boldsymbol{\nabla U} = \left(-\frac{\partial U}{\partial x}, -\frac{\partial U}{\partial y}, -\frac{\partial U}{\partial z}\right)$
- An object experiencing an energy potential, feels a force in the direction in which the energy potential is decreasing most rapidly.

- The gravitational energy potential is: $U(y) = mgy$

The gravitational force is $F = -\frac{dU(y)}{dy} = -mg$

The choice of the zero of the potential energy is arbitrary and does not carry physical meaning. If we add an arbitrary constant to $U(y)$ there are no changes on the gravitational force: $U'(y) = mgy + c \implies F = -\frac{dU(y)}{dy} = -mg$

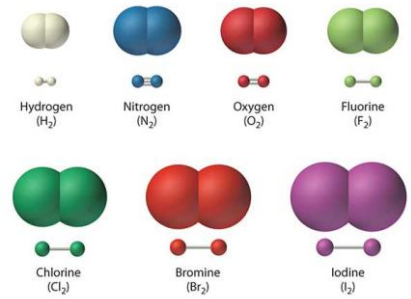
3.7 Harmonic potential

- Harmonic forces are proportional and in opposition to the displacement.
- They produce periodic motions, and their potential is quadratic.

$$F = -k(x - x_0) \;\Rightarrow\; U = -\frac{1}{2}k(x - x_0)^2$$

$$\rightarrow x(t) = A\cos(\omega t + \varphi)$$

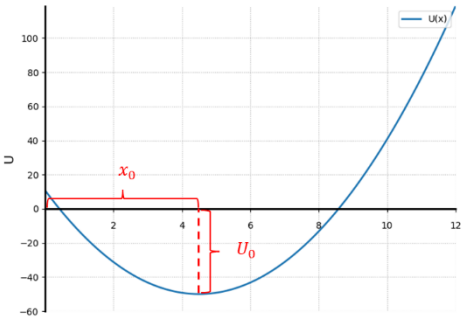
- A: amplitude of the motion
- $\omega = \sqrt{\frac{k}{m}}$
- φ: depends on the initial position and velocity



- Diatomic molecules (and chemical bonds in general) can be modeled by harmonic potentials.

$$\rightarrow U = -\frac{1}{2}k(x - x_0)^2 + U_0$$

- x_0: bond distance
- U_0: bond energy

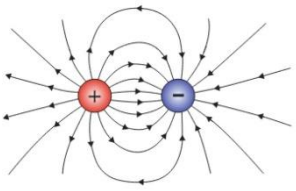


3.8 Electrostatic energy potential

- Electrostatic forces are the forces that charged particles exert on each other.
- The force is attractive if the charges are of opposite signs and repulsive if they are of the same sign.
- The force is determined by the Coulomb law.

$$\rightarrow F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2} \;\Rightarrow U = -\frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r}$$

- q_1, q_2: charges of the interacting particles
- r: distance between the two particles
- ε_0: dielectric constant



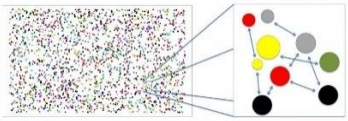
4 Thermodynamics

4.1 Thermodynamic systems

- A thermodynamic system is a system, separated by its surroundings, made of matter (or radiation) which can be studied by the law of thermodynamics.
- A thermodynamic system is not necessarily isolated and can, in principle, exchange energy or matter with its surroundings.

• Usually, we consider it made of a high number of smaller parts (proteins in a cell, atoms in a protein, etc.).
• The system is usually described by macroscopic variables that can be easily measured, while its components are described by microscopic variables that are not easily accessible.

4.2 Internal energy



We assume that no external force is acting on the system and that the velocity of the system respect to the reference of the lab is $v = 0$. Each particle will have a certain total energy, given by the sum of the kinetic energy and the potential energy due to the interaction with other particles and the surrounding (the particles collide with each other and the energy can change during the collisions):

$$E_{T,i} = K_i + U_i$$

So, we define the internal energy U of the system the sum of the energies of the particles composing the system:

$$U = \sum_{i=1}^{N} E_{T,i} = \sum_{i=1}^{N} (K_i + U_i)$$

(U is a function of the state of the system)

• The energy of a single particle is completely determined once the position and velocity of the particle is known:

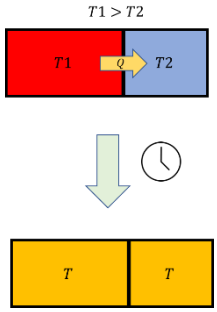
$$K_i = \frac{1}{2} v_i^2 \qquad\qquad U_i = U_i(\boldsymbol{r}) = U_i(x, y, z)$$

4.3 Pressure

Pressure p the average of the total force exerted by the fluid to the container for unit area:

$$p = \frac{F}{A}$$

4.4 Thermal equilibrium



Let us consider two thermodynamics systems of different temperatures ($T1$ and $T2$, with $T1 > T2$). We know from experience that if they are put in interaction, they will eventually reach the same temperature T, with $T2 < T < T1$. The two systems are now in thermal equilibrium.
When this happens, we say that the hotter system has transferred a certain amount of heat (Q) to the colder system.
A thermodynamic system is in thermodynamic equilibrium (with itself) if each of its parts has the same or similar temperatures and pressure.

4.5 Heat transfer and Joule experiment

• The formula of heat transfer: $Q = mc\Delta T$
• The equivalence between work and heat show that heat is a form of energy: $W = Q$
• In the famous experiment James Prescott Joule demonstrate that the work done on a fluid can increase the temperature of the fluid once equilibrium is reached.

4.6 The first law of thermodynamics ☢

The change of internal energy in a system is equal to the heat absorbed minus the mechanical work done by the system:

$$Q - W = \Delta U$$

As we have seen, ΔU is a state function (i.e. it depends only on the initial and final states of the system and not on the path chosen for moving between states). While neither Q nor W are state function, their difference is a state function.

4.7 Internal energy and temperature

• The internal energy is linked to the temperature:

$$Q = mc\Delta T \qquad\qquad W = mc\Delta T$$

• The temperature measures the average kinetic energy of the particles in the system.

For monoatomic gases: $\langle K \rangle = \frac{3}{2} k_B T$ (k_B : Boltzmann constant)

• In general, heat and work could be "absorbed" by the potential energy part of the internal energy, and not only to increase

the kinetic energy, i.e. the temperature of the system.

$$U = \sum_{i=1}^{N} E_{T,i} = \sum_{i=1}^{N} (K_i + U_i)$$

4.8 The second law of thermodynamics ☢

• The second law of thermodynamics states that the total entropy of an isolated system can never decrease over time; it can only remain constant or increase. In simpler terms, it suggests that natural processes tend to move towards a state of greater disorder or randomness.

$$\sum_{i=1}^{n} \frac{Q_i}{T_i} \leq 0$$

And the equality holds only for reversible transformations:

$$\sum_{i=1}^{n} \frac{Q_i}{T_i} = 0$$

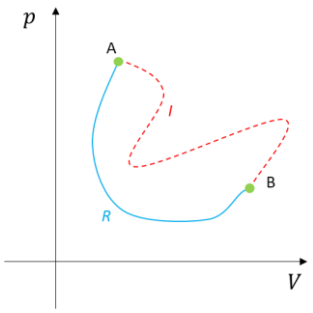
This means that the quantity $\frac{\delta Q}{T}$ is a state function when the calculation is done along reversible transformations. We call entropy (S) this state function and the entropy change is defined as:

$$S(B) - S(A) = \int_A^B \frac{\delta Q}{T}$$

• Let us consider a transformation that move us from state A to state B through an irreversible path, then bring us back through an irreversible path.

$$0 \geq \int_A^A \frac{\delta Q}{T} = \left(\int_A^B \frac{\delta Q}{T}\right)_I + \left(\int_B^A \frac{\delta Q}{T}\right)_R = \left(\int_A^B \frac{\delta Q}{T}\right)_I - [S(B) - S(A)]$$

$$S(B) - S(A) \geq \left(\int_A^B \frac{\delta Q}{T}\right)_I$$



If we consider a completely isolated system the term $\left(\int_A^B \frac{\delta Q}{T}\right)_I$ is 0, as $\delta Q = 0$, and the previous equation becomes: $S(B) \geq S(A)$

This means for any transformation occurring in an isolated system, the entropy of the final state can never be less than that of the initial state. If the transformation is reversible, the equality sign holds, and the system suffers no change in entropy.

4.9 Free Energy

From the definition of entropy and considering that T is constant we can write:

$$\mathrm{W} = -\Delta U + Q$$

$$S(B) - S(A) \geq \int_A^B \frac{\delta Q}{T} \quad \Rightarrow \quad Q = \int_A^B \delta Q \leq T[S(B) - S(A)]$$

Putting together the two equations we can see that there is an upper limit on the amount of work we can extract from a thermodynamics system, as this inequality holds:

$$\mathrm{W} = -\Delta U + Q \leq -\Delta U + T[S(B) - S(A)]$$

In other word the total amount of work that can be extracted is limited by the increase of entropy.

• For Gibbs Free Energy:

$$G = H - TS$$

- G is Gibbs free energy.
- H is Enthalpy.
- T is the absolute temperature.
- S is the Entropy of the system.

• For Helmholtz Free Energy:

$$F = U - TS$$

- F is the Helmholtz free energy.
- U is Internal Energy.
- T is the absolute temperature.
- S is the Entropy of the system.

Attention: If the free energy is a minimum, the system is in a state of stable equilibrium.

5 Statistical Mechanics

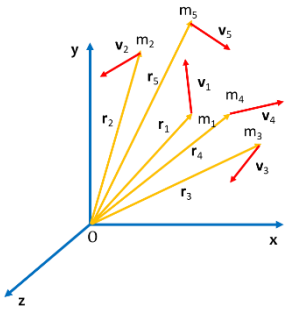
5.1 Thermodynamic variables

- **Extensive Variables:** These are properties that depend on the size or extent of the system. They are additive for subsystems and include variables such as mass (m), volume (V), internal energy (U), entropy (S), and the amount of substance (n).
- **Intensive Variables:** These are properties that do not depend on the system size or the amount of material in the system. They include temperature (T) and pressure (p).

5.2 Microstates and macrostates

The state of a system is a vector in a 6N dimensional space:

$$\boldsymbol{\sigma} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_N, \boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N) = \left(x_1, y_1, z_1, x_2, y_2, z_2, \ \dots, x_N, y_N, z_N, v_{x1}, v_{y1}, v_{z1}, v_{x2}, v_{y2}, v_{z2}, \ \dots, v_{xN}, v_{yN}, v_{zN}\right)$$



- The vector $\boldsymbol{\sigma}$ is the microstate of the thermodynamic system.
- Achieving such a detailed description of a system with many particles is practically impossible and essentially useless, as many microstates can be very similar in practice.
- A macrostate is defined by the macroscopic properties of a system, such as its temperature, pressure, volume, and the number of particles.
- A single macrostate can correspond to a vast number of microscopic configurations (or microstates) of the system's components (like atoms or molecules), which are consistent with the macroscopic properties.

5.3 Macrostates and thermodynamic variables

- Equiprobability Postulate: In an isolated system in thermodynamic equilibrium, all microstates that have the same energy are equally probable.
- Even though microstates are equally probable, macrostates are not. It is easy to understand by thinking of dice rolling. When rolling multiple dice, extreme values become more and more unlikely, and the average becomes more probable.

5.4 Ergodic hypothesis

A thermodynamic system evolves in time following some trajectory in the phase space. If each of the microstates are equally probable, the system will spend more time in macrostates that are represented by a larger number of microstates.

- Under this hypothesis, the phase average of a variable (A) will be equal to the time average:

$$\langle A \rangle = \bar{A} = \frac{\sum_{t=1}^{T} A(t)}{T}$$

phase average: average of an over all the possible microstates of the system

time average: average of an over time in a particular dynamic trajectory

- We define Ω as the number of microstates that correspond to a certain macrostate. The entropy of the macrostate is (Boltzmann formula for entropy):

$$S = k_B ln\Omega$$

5.5 Systems in thermal equilibrium

In an isolated system in thermodynamic equilibrium, all microstates that have the same energy are equally probable. But if the system is not isolated and can exchange heat (i.e. the total energy is not conserved?)

- For example, let us say that the lower part of the graph has lower energy. In this case, the probability of finding a state of total energy E will be proportional to its Boltzmann factor:

$$P(\sigma) \propto e^{-\frac{E(\sigma)}{k_B T}}$$

- The most probable macrostate will have high entropy and low energy, i.e. low free energy.
- The phase average of a variable (A) is still equal to the time average, because the time past in low energy configuration is on average higher:

$$\langle A \rangle = \frac{\sum_\sigma A e^{\frac{-E(\sigma)}{k_B T}}}{\sum_\sigma e^{\frac{-E(\sigma)}{k_B T}}} = \bar{A} = \frac{\sum_{t=1}^{T} A(t)}{T}$$

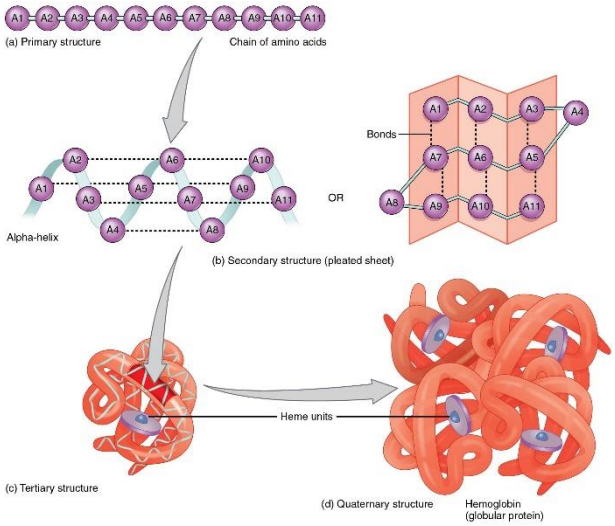
phase average: average of A over all the possible microstates of the system

time average: average of A over time in a particular dynamic trajectory

6 Structural Models

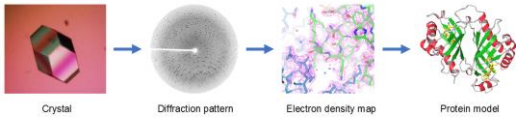
6.1 Protein structure

- **Primary structure**: Linear sequence of amino acids linked by peptide bonds (shown as beads on a string)
- **Secondary structure**: Local folding patterns stabilized by hydrogen bonds
  - Alpha helix (coiled structure)
  - Beta sheet (zigzag planar structure)
- **Tertiary structure**: 3D arrangement and folding of the entire polypeptide chain stabilized by various interactions (ionic, hydrogen bonds, disulfide bridges, hydrophobic)
- **Quaternary structure**: Complex of multiple polypeptide chains (subunits) arranged in a specific geometric pattern (e.g., haemoglobin with 4 subunits)
- Structural levels increase from primary to quaternary, with more complex folding and molecular interactions determining the final 3D shape and function

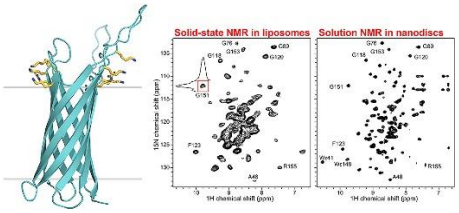


6.2 Experimental methods

- **X-Ray Crystallography**
  - Requires crystallizing the protein of interest
  - X-rays are diffracted by the repeating array of molecules in the crystal
  - Diffraction pattern allows 3D structures to be determined at atomic resolution
  - Established method but challenges in crystallization

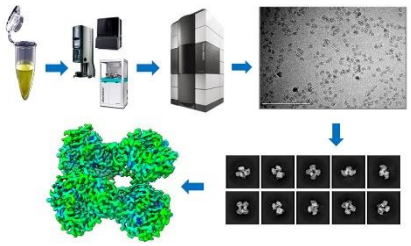


- **NMR Spectroscopy**
  - Measures magnetic properties of atomic nuclei in a protein solution
  - Provides info on atoms' chemical environment and dynamics
  - Calculates 3D structure from nuclear distances/angles
  - Can study proteins in near-native solution conditions
  - Size limits of ~50-100 kDa for traditional NMR



- **Cryo-Electron Microscopy (Cryo-EM/ET)**
  - Rapidly freezes protein samples in vitreous ice
  - Electrons scattered by frozen biomolecules recorded

- 3D maps reconstructed from 2D projection images
- Recent resolution "revolution" reaching near-atomic levels
- Can study large complexes, membrane proteins, dynamics



6.3 The Protein Data Bank and PDB files

Atom number, Atom "Type", Residue Name, Chain, Residue Number

Atom coordinates (Å)

Occupancy (fraction of time spent in that position for each atom)

Everything following the "ATOM" word indicates the properties of an atom in the protein

Element

```
ATOM      1  N   ARG A  20      24.557 -38.438  -2.219  1.00 48.18           N
ATOM      2  CA  ARG A  20      24.786 -38.383  -3.724  1.00 48.20           C
ATOM      3  C   ARG A  20      26.209 -38.780  -4.233  1.00 46.86           C
ATOM      4  O   ARG A  20      26.312 -39.492  -5.209  1.00 46.96           O
ATOM      5  CB  ARG A  20      24.328 -37.035  -4.295  1.00 48.44           C
ATOM      6  CG  ARG A  20      23.791 -37.076  -5.718  1.00 49.20           C
ATOM      7  CD  ARG A  20      23.543 -35.655  -6.309  1.00 49.60           C
ATOM      8  NE  ARG A  20      22.419 -34.974  -5.669  1.00 54.70           N
ATOM      9  CZ  ARG A  20      22.453 -33.754  -5.107  1.00 55.47           C
ATOM     10  NH1 ARG A  20      23.558 -33.001  -5.109  1.00 52.88           N
ATOM     11  NH2 ARG A  20      21.345 -33.268  -4.550  1.00 55.28           N
ATOM     12  N   SER A  21      27.281 -38.306  -3.583  1.00 45.68           N
ATOM     13  CA  SER A  21      28.663 -38.666  -3.946  1.00 43.64           C
ATOM     14  C   SER A  21      29.113 -39.741  -2.991  1.00 42.98           C
ATOM     15  O   SER A  21      28.807 -39.631  -1.810  1.00 43.65           O
ATOM     16  CB  SER A  21      29.618 -37.466  -3.846  1.00 43.49           C
ATOM     17  OG  SER A  21      29.413 -36.523  -4.890  1.00 41.49           O
ATOM     18  N   PHE A  22      29.822 -40.764  -3.483  1.00 41.35           N
ATOM     19  CA  PHE A  22      30.314 -41.886  -2.644  1.00 40.87           C
ATOM     20  C   PHE A  22      31.844 -42.145  -2.634  1.00 39.14           C
ATOM     21  O   PHE A  22      32.290 -43.097  -1.991  1.00 39.49           O
ATOM     22  CB  PHE A  22      29.550 -43.211  -2.938  1.00 41.87           C
ATOM     23  CG  PHE A  22      28.062 -43.084  -2.760  1.00 44.68           C
ATOM     24  CD1 PHE A  22      27.255 -42.664  -3.815  1.00 46.93           C
ATOM     25  CD2 PHE A  22      27.468 -43.322  -1.516  1.00 49.07           C
ATOM     26  CE1 PHE A  22      25.864 -42.500  -3.642  1.00 50.21           C
ATOM     27  CE2 PHE A  22      26.069 -43.156  -1.328  1.00 50.25           C
ATOM     28  CZ  PHE A  22      25.272 -42.759  -2.394  1.00 48.69           C
```

6.4 Methods for Structural predictions (general introduction)

- **Homology Modeling:**

  Homology modeling is a technique for predicting the three-dimensional structure of a protein based on its sequence similarity to other proteins with known structures.

  It relies on the principle that proteins with similar sequences often have similar structures.

  Software tools like MODELLER, SWISS-MODEL, and I-TASSER are widely used for homology modeling.

- **Threading or Fold Recognition:**

  Threading methods attempt to recognize known protein folds or structural motifs in the target protein sequence.

  They involve aligning the target sequence onto known protein structures and evaluating the compatibility based on various scoring functions.

  Examples of threading programs include PROSPECTOR, SPARK, and pGenTHREADER.

- **Ab initio or de novo Modeling:**

  Ab initio methods aim to predict protein structures from their amino acid sequences without relying on known structural templates.

  These methods often use physics-based energy functions or machine learning techniques to sample the conformational space and identify low-energy structures.

  Popular ab initio modeling programs include ROSETTA, QUARK, and UNRES.

- **Machine Learning and Deep Learning Methods:**

  Machine learning and deep learning techniques, such as neural networks and deep belief networks, have recently gained popularity in protein structure prediction.

  These methods can learn patterns and features from large datasets of known protein structures and sequences.

  Examples include AlphaFold, trRosetta, and the methods used in the CASP competitions.

- **Hybrid Methods:**

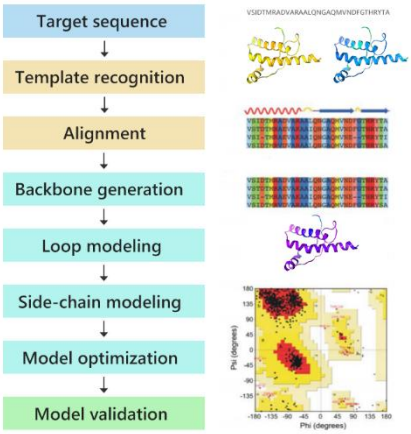
  Hybrid methods combine various computational techniques, such as homology modeling, threading, ab initio

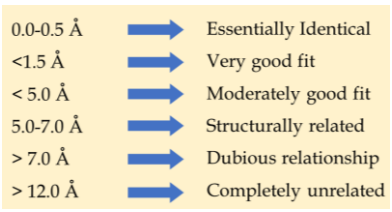
modeling, and machine learning, to improve the accuracy of structure prediction.
These methods often employ consensus or meta-server approaches, integrating the results from multiple structure prediction methods.
Examples include MULTICOM, I-TASSER, and Robetta.

6.5 Homology Modeling

Homology modeling, also known as comparative modeling, uses homologous sequences with known 3D structures for the modeling and prediction of the structure of a target sequence

- Basic assumption: two homologous proteins will share very similar structures
- Although the number of actual proteins is vast, there is a limited set of tertiary structural motifs to which most proteins belong
- Homology modeling is one of the best-performing methods and (in most cases) gives accurate predicted models



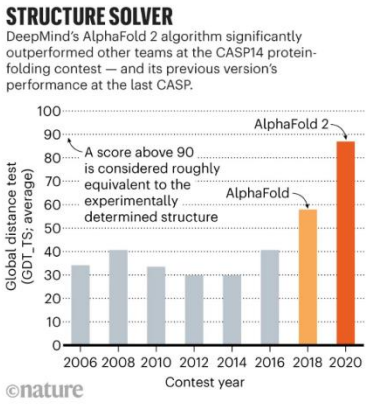


Step required to perform Homology Modeling:

1. Retrieving the sequence to be modeled (target sequence)
2. Identifying suitable templates from PDB using BLAST
3. Aligning target sequence with template sequences
4. Performing homology modeling using proper tools
5. Model quality check
6. Quantitative comparison of the model with the experimental information

(In most cases, all these steps are done automatically by the software)

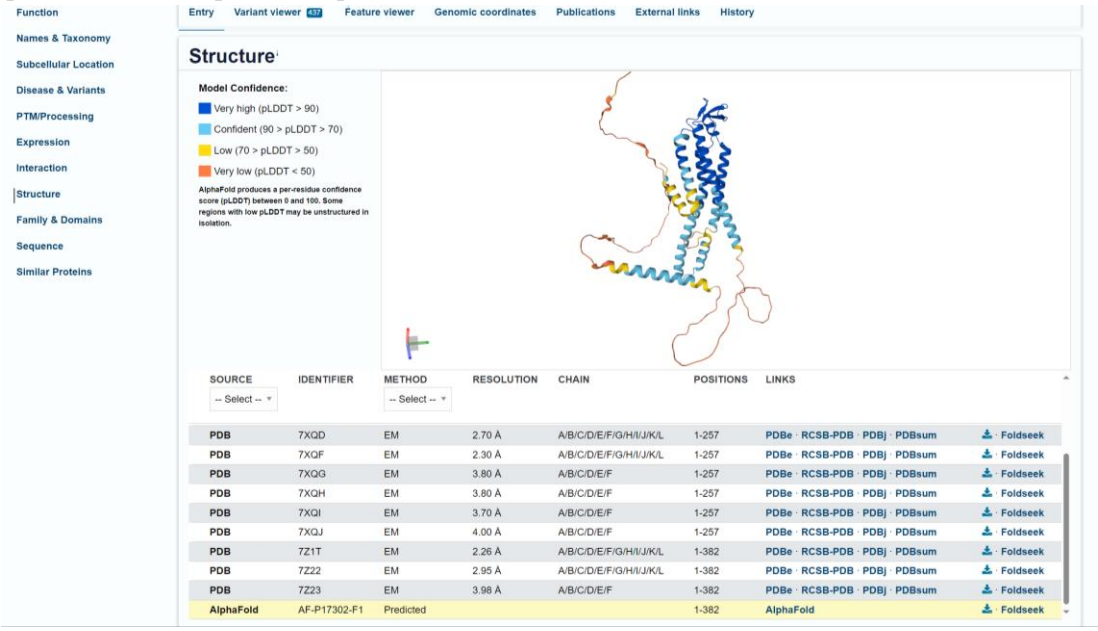
6.6 AlphaFold



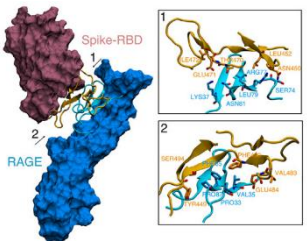
AlphaFold utilizes deep learning and advanced machine learning techniques to accurately predict protein structures. The key innovations of AlphaFold include:

1. Attention-based Neural Network: AlphaFold uses an attention-based neural network architecture that can capture long-range interactions between amino acids, which are crucial for accurate protein structure prediction.
2. Evolutionary Information: AlphaFold incorporates evolutionary information from multiple sequence alignments, which helps to improve the accuracy of its predictions.
3. End-to-End Approach: AlphaFold is an end-to-end approach that directly predicts the three-dimensional structure from the amino acid sequence without relying on intermediate steps or templates.

AlphaFold predictions on Uniprot (example):



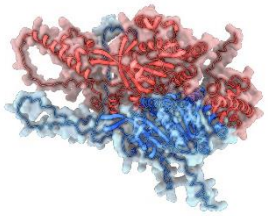
6.7 Molecular Docking



Protein-Ligand interactions:

- Ligands are usually small and more rigid than proteins (few degrees of freedom).
- Small molecules usually bind to pockets (cavities).
- Interactions are often mediated by hydrophobic contributions.
- In many practical cases, a large number of ligands need to be tested (screened).

Protein-Protein interactions:



- Proteins are large and can have important configuration changes upon binding.
- The interaction between the proteins happens on their surface.
- It is often mediated by hydrophilic residues and by surface complementarity.
- The interacting proteins are usually known (no need for screening).

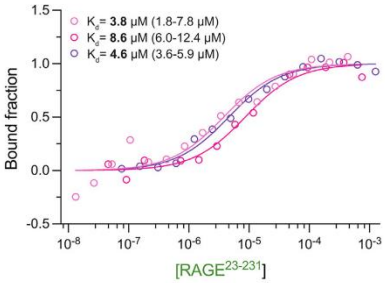
6.8 Binding Affinity and Binding Free Energy

Let us assume we have a generic chemical reaction: $A\_x\ B\_y \rightleftharpoons xA+yB$; we call dissociation constant the quantity:

$$K_D = \frac{[A]^x[B]^y}{[A_xB_y]}$$

If the reaction is 1:1 ($x = 1,\ y = 1$), the dissociation constant represents the concentration at which half of the molecules are interacting:

$$K_D = [A],\ [B] = [AB] \text{ or, equivalently, } \frac{[AB]}{[B]+[AB]} = \frac{1}{2}$$



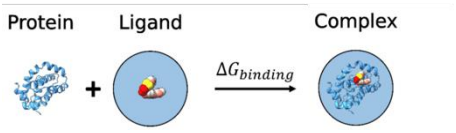
The binding affinity is linked to the binding free energy ΔG (difference of free energy between the bound and unbound state) by the following equation:

Energy measured in $KJ/mol \cdot K$

$$\begin{cases} \Delta G = -RTlnK_D \\ K_D = e^{-\Delta G/RT} \end{cases}$$

$$K_D = e^{-\Delta G/K_BT} \rightarrow \text{Energy measured in } K_BT$$

- ΔH: Energy changes associated with breaking and forming intermolecular interactions, (hydrogen bonds, van der Waals interactions, and electrostatic interactions), during the binding. Can be favorable or unfavorable. Energy changes (usually unfavorable) due to the loss of solvation of the proteins and ligands.



- ΔS: change in the configurational space accessible to the proteins and ligands upon binding (usually unfavorable). Release or reorganization of water molecules from the binding interfaces (can contribute favorably)
- ΔG is the difference of free energy between the bound and unbound state:

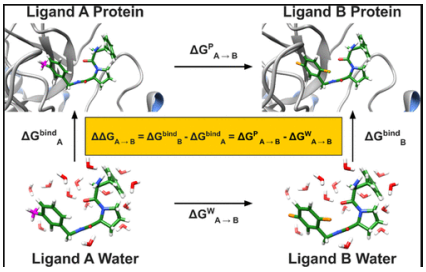
$$\Delta G = \Delta H - T\Delta S = G_B - G_U = U_B - U_U - TS_B - TS_U$$

(If no change in volume happens: $\Delta H = \Delta U + p\Delta V = \Delta U$)

Example: if we want to compare multiple ligands, we can calculate $\Delta\Delta G = \Delta G_B - \Delta G_A$, i.e. the difference between the two binding affinities

$$\frac{K_{DA}}{K_{DB}} = e^{-(\Delta G_A - \Delta G_B)/K_BT} = e^{\Delta\Delta G/K_BT}$$

If $\Delta\Delta G < 0$ B is a better binder than A (lower binding energy, better binding affinity)



6.9 Docking procedures and strategies

Binding Site Detection:

- Identify the possible binding site from the literature (list of known binding sites) or from structural information.

Scoring:

- Evaluates the quality of the docking and allow to rank different results.
- It generally uses approximate calculation and can roughly been interpreted as binding energy.

Conformational Search:

- Further analysis can be necessary to improve the quality of the results.
- It is possible to study the system dynamics to gain insight on the quality and stability of the docking.
- Usually requires simulations (see next lesson).

Rigid docking:

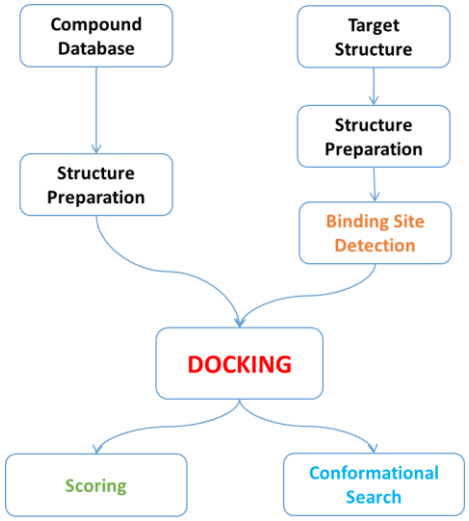
- Both molecules are assumed rigid in orientation
- Explore the best matching with minimized energy of the system
- Simple and less computational cost
- Suits for protein-protein, protein-nucleic acid

Semi-flexible docking:

- Ligand – flexible, target - rigid
- Larger but reasonable computational cost
- Suits for protein-small molecule

Flexible docking:

- Both molecules - flexible
- Tremendous computational cost
- Suits for accurate examination of sophisticated interaction

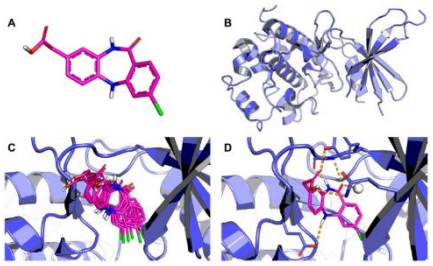


Outline of the molecular docking process:

[A] 3D structure of the ligand; [B] 3D structure of the receptor;

[C] The ligand is docked into the binding cavity of the receptor and the putative conformations are explored;

[D] The most likely binding conformation and the corresponding intermolecular interactions are identified.



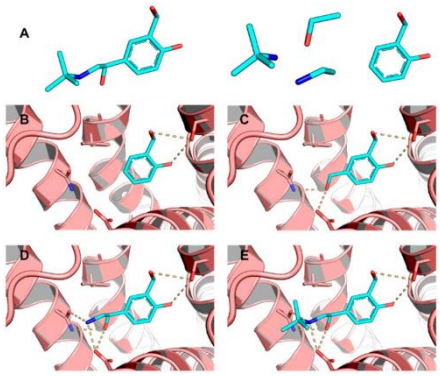
The incremental construction method:

[A] The ligand (stick representation, carbon in cyan) is broken into several fragments;

[B] The anchor fragment is docked in the binding site of the molecular target (cartoon representation, carbon in salmon);

[C] The next fragment is docked after the anchor fragment;

[D & E] The other fragments are docked sequentially to construct the entire ligand in its binding conformation. Residues in the active site are shown in stick representation (carbon in salmon). Hydrogen bonds are indicated as dashed lines.



(Reference: Ferreira et al. Molecules. 2015 Jul; 20(7): 13384–13421.)

7 Molecular Dynamics

7.1 Concepts of Molecular Dynamics

Molecular dynamics **(MD)** is a computer simulation method for analyzing the physical movements of atoms and molecules.

- We want to describe our biological system of interest from the point of view of the Physics and derive its properties from physical principles.
- From a theoretical point of view, a biological system can be regarded as a collection of nuclei and electrons, interacting through the electrostatic potential. The electron density can be written in term of a wave function, which is solution of the Schrödinger equation.

$$i\hbar\frac{\partial}{\partial t}\Psi(X,t) = \left[\frac{-\hbar^2}{2m}\nabla^2 + V(x,t)\right]\Psi(X,t)$$

$$\rho_i(X) = q_i|\Psi_i(X)|^2 \quad \text{charge density}$$

$$V_{1,2} = \int \frac{\rho_1\rho_2}{4\pi\varepsilon_0 r_{1,2}} d\,\rho_1 d\rho_2$$

This equation describes the evolution of the probability density function in time.

• Particle-wave duality: elementary particles can behave as ordinary matter or as waves, depending on how we measure their characteristics.

$\Psi(x,y,z)$: wave function

$\int |\Psi(x,y,z)|^2\, dxdydx$: probability density

Particles are described by wave function, which represent probability densities and the wave function evolves according to the Schrödinger equation.

7.2 Classical Molecular Dynamics

Each atom is treated as a "classical" object which interacts with the other atoms in the system.

The interaction is represented through a classical potential.

The time evolution of the system is determined by the second Newton's law.

$$\boldsymbol{F} = m\boldsymbol{a}$$

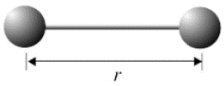
Since the electronic configuration is frozen in the ground state, we can simplify the form of the energy of interaction in many different ways:

1) A charge is assigned to each atom, and it is equal to the average charge of that atom.
2) Interactions between bonded atoms reproduce chemical and geometrical properties of the molecule, in particular regarding bond distances, angles and strength.
3) Non-bonded atoms interact exclusively through Coulomb and Lennard-Jones potentials.

7.3 Chemical bond properties

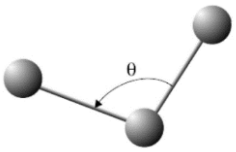
The properties of the chemical bonds depend on the electronic configuration of the atoms in the molecule, and not only on the kind of atoms. Atoms oscillate around their equilibrium average distances; thus, chemical bonds are treated as *harmonic oscillators*. The spring constant depend on the atoms involved and on their electronic configurations

$$U_r = k_r(r - r_0)^2$$



Two consecutive bonds form an angle. For this angle, we assume a harmonic approximation: the angle can oscillate around the equilibrium configuration with a well-defined angular strength

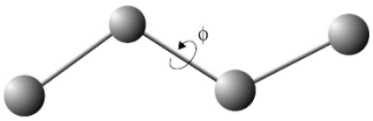
$$U_\theta = k_\theta(\theta - \theta_0)^2$$

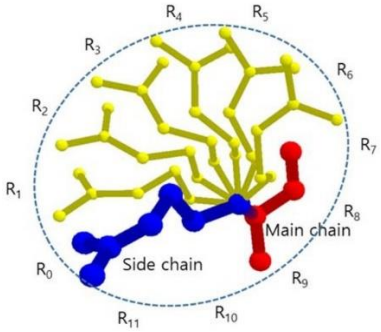


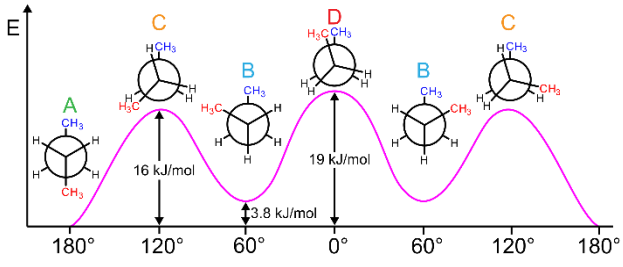
Four consecutive atoms define three bonds and two planes. The two planes describe an angle which is called *dihedral angle*. Due to steric and electrostatic interaction, not all the angular position are allowed for the dihedrals. Several minima of the can exist for a single dihedral angle, corresponding to different possible configurations (*isomers of rotation* or *rotamers*).

• Dihedral angles allow biological macromolecules to be flexible and assume many different configurations.

$$U_\varphi = \sum C_k \cos(\varphi - \varphi_{0,k}))$$



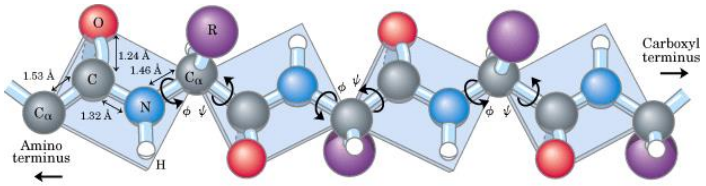




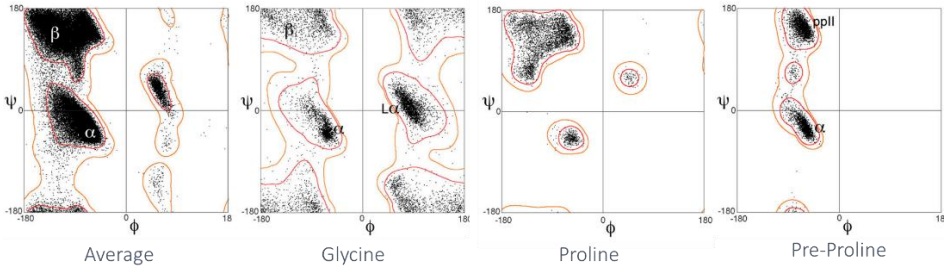
Relative conformation energy diagram of butane as a function of dihedral angle

7.4 Ramachandran Plots

- The relative position of two consecutive amino acids is described by two dihedral angles (φ and ψ).
- Not every combination of φ and ψ is possible, due to steric and electrostatic hindrance.
- The values of the two dihedral angles is linked to the local secondary structure.

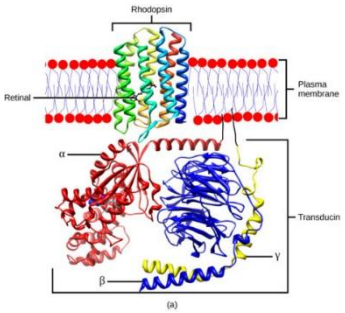


- Different amino acids have different preferred orientation (which explain their secondary structure propensity).



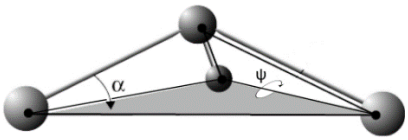
The *cys-trans* isomerism, is particular important in chemistry and biochemistry. It refers to cases in which a dihedral angle can have two different values. Switching from one to the other can induce large configurational changes.

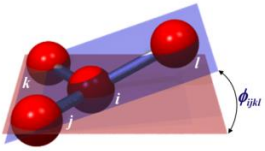
Rhodopsin is a subfamily of photosensitive GPCRs. It contains a *retinal*, a molecule that can transition between cys and trans configurations upon the absorption of a *photon*. The change in the retinal configuration causes the activation of the GPCR signal, allowing the organism to sense light.



Carbon rings and other similar structure display planar configurations due to the specific electronic configurations of their atoms. Improper dihedral angles need to be introduced to keep such atoms in the same plane.

$$U_\alpha = k_\alpha(\alpha - \alpha_0)^2$$





Atoms that are not connect by stable chemical bonds are considered non-bonded, and they interact through long range potentials:

- Coulomb interaction: the charge represents the average charge on the atom due to the distribution of the electrons in that particular molecule.

$$U_C = \frac{q_i q_j}{4\pi\varepsilon_0 r}$$

- Lennard-Jones potential: It is the sum of an attractive part (quantum effect) and a repulsive part which represent the effect of electronic clouds in close proximity.

$$U_{LJ} = \left(\frac{A}{r^{12}} - \frac{B}{r^6}\right)$$

- Hydrogen bonds: an empirical potential that take explicitly into account of the possibility of creation of hydrogen bonds.

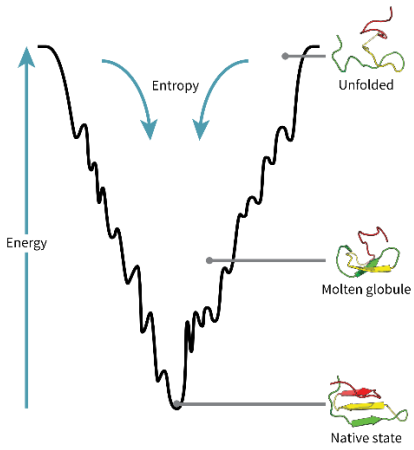
$$U_{HB} = \left(\frac{C}{r^{12}} - \frac{D}{r^{10}}\right)$$

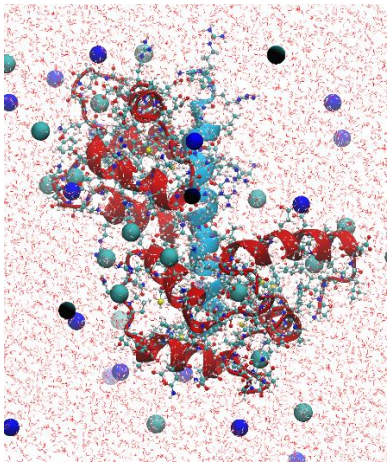
Below is one example of a classical force fields:

$$U = \sum_{bonds} k_r(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} C_k \cos(\varphi - \varphi_{0,k})) + \sum_{impropers} k_\alpha(\alpha - \alpha_0)^2 + \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{i<j} \left(\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6}\right) + \sum_{h.b.} \left(\frac{C}{r_{ij}^{12}} - \frac{D}{r_{ij}^{10}}\right)$$

7.5 Energy Minimization

The starting model to be analyzed with MD simulations is usually not at an energy minimum and not equilibrated. E.g. X-ray structure is in nonnative conditions (dehydration, close packing etc.), and do not have hydrogen atoms. Homology models can have side chains in erroneous positions and overlapping atoms. Before starting a simulation, it is necessary to perform energy minimization.

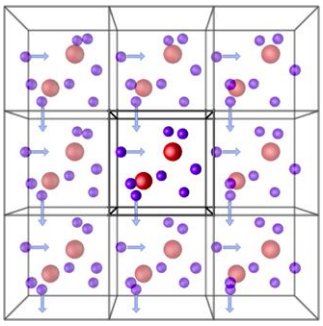


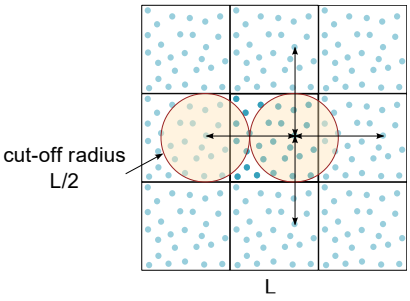


- Start from a PDB file of your system of interest (downloaded from the Protein Data Bank or obtained by modeling). In this example we see calmodulin interacting with the C-terminus of PMCA pump.
- You need to have the position of all atoms in the system. Hydrogen atoms and loops can be missing from the PDB file.
- These parts can be reconstructed with standard methodologies, but the reconstruction tends to produce clashes with high energy. Perform Energy Minimization before starting.
- Create a realistic environment: include ions, water molecules (and the cell membrane if necessary).

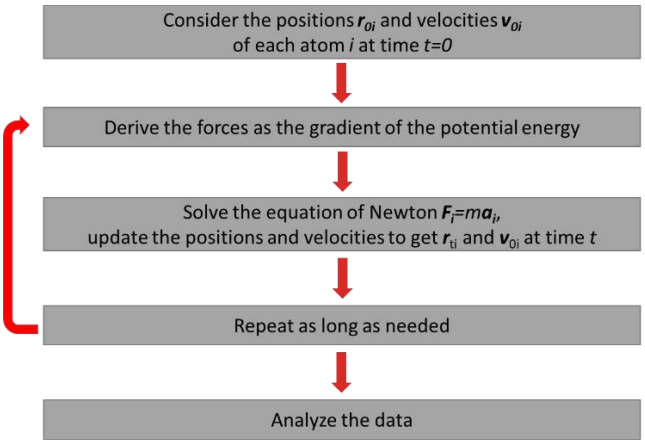
7.6 Periodic Boundary Conditions

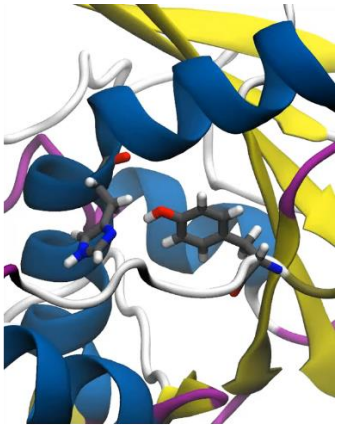
- An isolate box of water would evaporate very fast, losing contacts with the protein. To solve this issue periodic boundary conditions are introduced. The system is repeated identical infinite times in the 3-dimensional space.
- This is not necessary when working with implicit solvent.Since there exist infinite copies of the system, the total charge of the system can be only **zero** (good) or **infinite** (bad)
- Furthermore, periodic boundary conditions create a system which always have infinite atoms, and atoms can in principle interact with their replicas. Interactions must be calculated only within a cutoff distance.





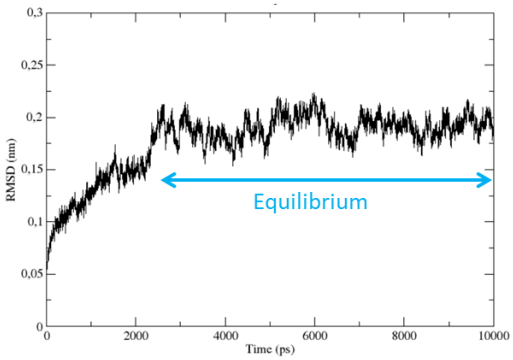
7.7 Molecular Dynamics algorithm





7.8 RMSD and RMSF

RMSD



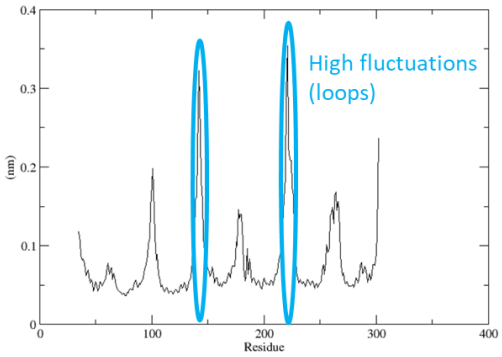
$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i(t) - \boldsymbol{r}_i(0))^2}$$

Root Mean Squared Deviations

It is a measure of the average displacement of a group of atoms from their initial/reference positions over the course of a molecular dynamics (MD) simulation trajectory. It is calculated as a function of time/frame and provides insight into the overall structural deviation during the simulation.

The RMSD is expected to increase initially and then typically plateaus once equilibrium is reached, oscillating around a relatively constant value (of a few Å).

RMSF



$$RMSF(i) = \sqrt{\frac{1}{T}\sum_{t=0}^{T}\left(\boldsymbol{r}_i(t) - \boldsymbol{r}_{i,AVE}\right)^2}$$

Root Mean Squared Fluctuations

It quantifies the mobility of individual atoms or residues by measuring their time-averaged deviation from a reference position (usually the initial structure or average structure) over the entire MD trajectory.

It is calculated as a function of the residue index along the protein sequence. Higher RMSF values correspond to more flexible regions, while lower values indicate rigid or well-packed regions.

8 In Molecular Simulations

8.1 Concepts of Molecular Dynamics