

STRUCTURAL BIOLOGY

❖ INTRODUCTION

STRUCTURAL BIOLOGY is a branch of molecular biology, biochemistry and biophysics concerned with:

1. molecular structure of macromolecules
2. how macromolecules acquire the structure they have
3. how alterations in their structures affect their function

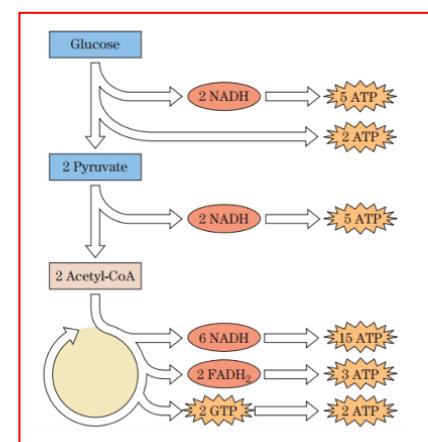
The final aim is to understand molecular mechanisms at molecular and atomic level, by studying detailed structures.

As we know, protein structures are important because **PROTEIN FUNCTION IS MADE BY STRUCTURE**.

❖ THE RESPIRATORY CHAIN

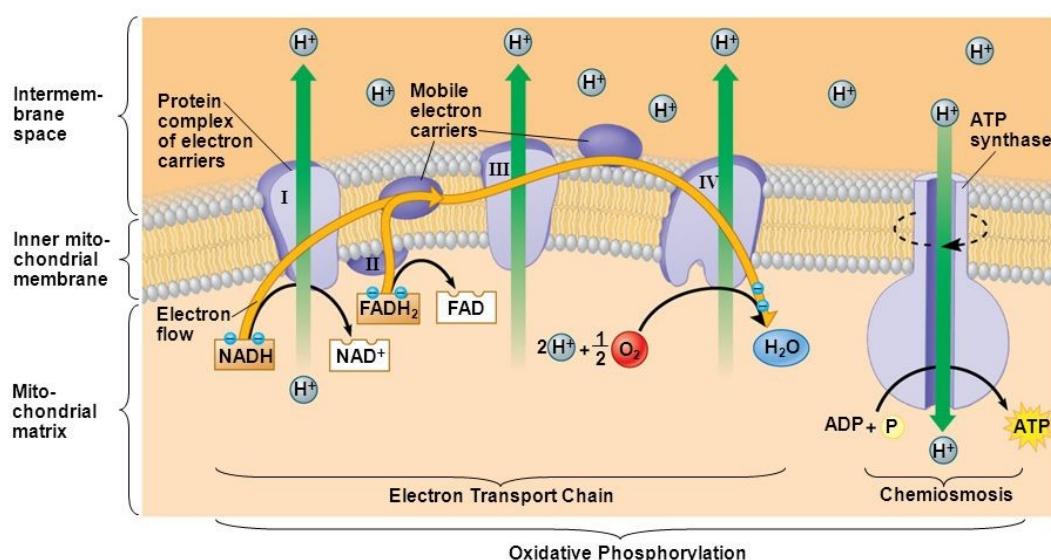
ENERGY METABOLISM is a series of reactions that occurs within the cells where:

1. **complex metabolites** are made into **monomeric units**,
2. monomeric units are transformed into a common intermediate, **acetylCoA**
3. acetyl group is oxidized to CO_2 via the citric acid cycle (TCA) with **reduction of NAD^+ and FAD**
4. Reoxidation of NADH and FADH_2 by O_2 by the oxidative phosphorylation system (OXPHOS) **yields H_2O and ATP** thanks to the formation of an electrochemical gradient.



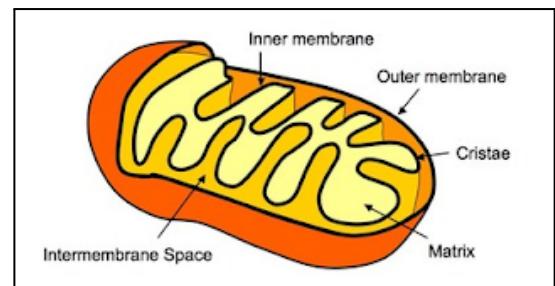
All oxidative steps in the degradation of macromolecules converge at the final stage of cellular respiration, **which involves the reduction of O_2 to H_2O with electrons donated by NADH and FADH_2 , which are oxidized into NAD^+ and FAD** .

Respiration occurs equally in light or darkness.



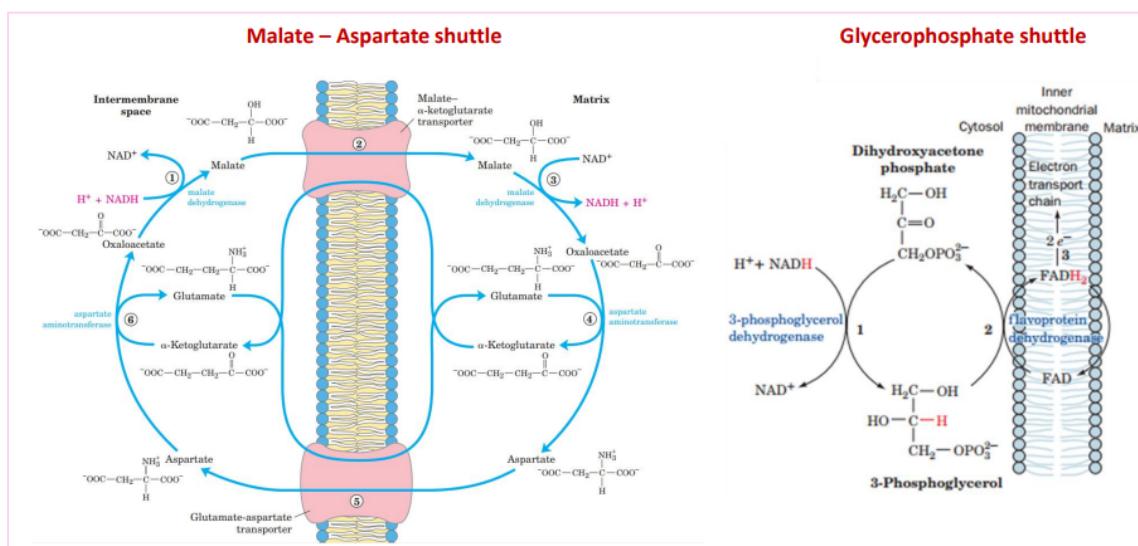
Respiration occurs in both **bacteria** and **eukaryotes**. In eukaryotes respiration happens in **mitochondria**, at the level of the *cristae* of the **inner mitochondrial membrane** (which are basically *invaginations*).

The **inner mitochondrial membrane** is **not permeable to ions** which is needed to keep the ion gradient stable. It is also extremely rich in proteins.

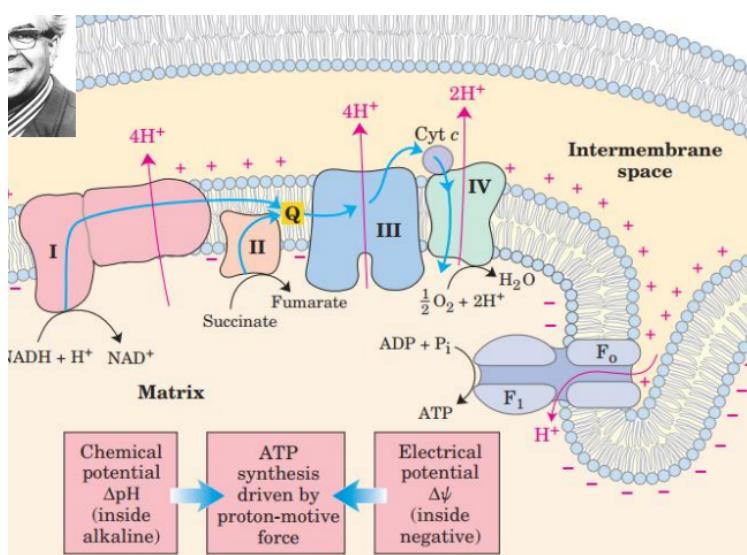


NADH enters the mitochondria and is then oxidized in order to reduce O₂ into H₂O. As we said the membrane of the mitochondria is not permeable, so NADH cannot just go in. **NADH enters the mitochondria** through:

- malate-aspartate shuttle
- reduction of dihydroxyacetone phosphate into 3-phosphoglycerol, which takes up the electrons carried by NADH. 3-phosphoglycerol is then oxidized back into dihydroxyacetone phosphate: this reaction leads to the reduction of FADH₂ inside the inner mitochondrial membrane. FADH₂ is then oxidized and it gives out electrons to the electron transport chain.



In the electron transport chain there are **FIVE EMBEDDED COMPLEXES** and two **ELECTRON CARRIERS**: coenzyme Q (or ubiquinone) and cytochrome C.



The redox reactions, through which electrons are produced, are coupled with the pumping of protons into the intermembrane space from the matrix.

These protons are then used to favor ATP synthase because they need to be pumped back into the matrix. Their movement allows ATP production.

The **ELECTROCHEMICAL GRADIENT**, that ensures the passing of ions through the ATPase complex, has two different components:

- **the chemical gradient**, or difference in concentration (of protons) across a membrane,
- **the electrical gradient**, or difference in charge across a membrane. This gradient is given by the positive environment created in the inner membrane and negative environment in the matrix

The electron carriers of the respiratory chain are organized into membrane embedded supramolecular complexes that can be physically separated. Separation happens through gentle treatment of the inner mitochondrial membrane with specific detergents, allowing the resolution of four unique electron carrier complexes and ATP synthase.

TABLE 19–3 The Protein Components of the Mitochondrial Electron-Transfer Chain

Enzyme complex/protein	Mass (kDa)	Number of subunits*	Prosthetic group(s)
I NADH dehydrogenase	850	43 (14)	FMN, Fe-S
II Succinate dehydrogenase	140	4	FAD, Fe-S
III Ubiquinone cytochrome <i>c</i> oxidoreductase	250	11	Hemes, Fe-S
Cytochrome <i>c</i> [†]	13	1	Heme
IV Cytochrome oxidase	160	13 (3-4)	Hemes; Cu _A , Cu _B

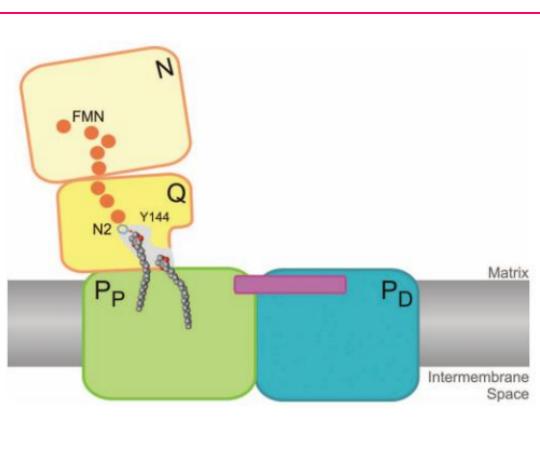
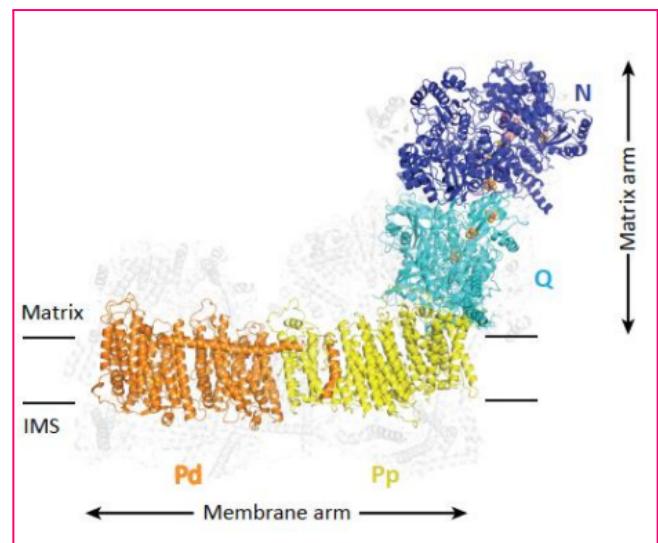
*Numbers of subunits in the bacterial equivalents in parentheses.

[†]Cytochrome *c* is not part of an enzyme complex; it moves between Complexes III and IV as a freely soluble protein.

❖ COMPLEX I

RESPIRATORY COMPLEX I (CI):

- ❖ is a **huge protein complex**, and it's also **hydrophobic**, so its crystal structure was difficult to obtain at first. It's actually the biggest respiratory complex with **43-46 subunits** (900-1000 kDa) depending on the species. There are **14 core subunits** present both in bacteria and in mammals.
- ❖ It's an **L shaped molecule**
- ❖ It's composed of two parts, an **hydrophilic** one that is in the mitochondrial **matrix**. And an **hydrophobic** one that instead is found in the inner mitochondrial **membrane**.
- ❖ It takes one NADH molecule to oxidize **ubiquinone** which causes the **pumping of four hydrogen protons**



- ❖ There are **three main modules**: N Q and P.
 - **N module** for NADH oxidation: electrons are given from NADH and go into iron-sulfur clusters
 - **Q module** for ubiquinone reduction: the module continues with some iron-sulfur clusters before reaching the ubiquinone site.
 - **P module** for proton pumping. In particular there are two different modules: P_p and P_d.
- The precise roles of accessory subunits are not yet fully understood, necessary for CI stability

The hydrophilic (matrix) arm has 7 main subunits, the hydrophobic (membrane) one also has 7. The **two arms are the unions of two different enzymes**, which worked together up until a certain point in evolution, where they integrated.

- **Membrane arm subunits** (with proton translocation activity) are similar to Mrp/H⁺ antiporters.
- **Matrix arm subunits** (which have oxidoreductase activity), are similar to nickel iron (Ni-Fe) hydrogenase, which are soluble. This also reflects how the matrix arm is the hydrophilic one.

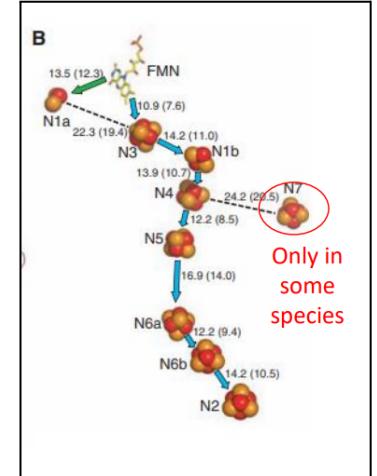
Of course, increasing the complexity of the species also increases the number of subunits and the size. Supernumerary subunits are concentrated on each side of the membrane domain, and around the lower section of the hydrophilic domain.

❖ MATRIX ARM

The subunits present in the matrix arm are *hydrophilic*, and are encoded in eukaryotes by **nuclear DNA**.

There are a series of **COFACTORS**:

- **FLAVIN MONONUCLEOTIDE (FMN)** which is the **first acceptor of NADH**. It is able to keep two electrons from NADH and can be present in three different forms: oxidative, reduced, and radical (which is an intermediate state).
- **8/9 IRON-SULFUR CLUSTERS** (**[2Fe-2S]** and **[4Fe-4S]**) that are coordinated by cysteines. They are able to **carry only one electron** at the time, one by one. Because FMN carries two electrons at a time, the transfer done by Fe-S clusters must be very fast. These clusters are coordinated by different subunits and the electron goes through all of them. The number of subunits might vary depending on the species.



Some of these **SUBUNITS** are:

NDUFVI has four domains.

- **N terminal domain** is wrapped around the Rossmann-fold domain, and contains three α -helices.
- **Rossmann-fold domain** binds **FMN** and it contains the **NADH binding site**, hence being responsible for its oxidation. This is an alpha-beta domain, with beta sheets twisted by alpha helices.
- **Ubiquitin-like domain**
- **C-terminal helical bundle** coordinates cluster **N3**

NDUFV2 is necessary for coordination of clusters in particular of **N1a**. N1a provides an **antioxidant activity**, and its structure is similar to Thioredoxin, which is involved in reactive oxygen species oxidification.

N1a is positioned out of the normal electron path. When an enzyme isn't working properly, N1a plays an important role: if electrons pile up in an environment full of oxygen, ROS is produced, which is dangerous for the cell. N1a acts as is able to reverse the electron transfer having an antioxidant functions. (RET vs FET).

NDUFS1 is also necessary for coordination like subunits **NDUFS7**, **NDUFS8** and **NDUFS2**. NDUFS8 and NDUFS2 interact extensively and take part of the CoQ binding/reaction chamber

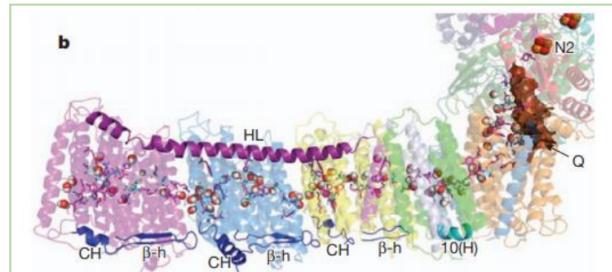
Finally, **NDUFS3** is necessary for the structure but it's not involved in the catalysis.

N1a - NDUFV2
N3 and FMN - NDUFV1
N1b, N4, N5, and N7 - NDUFS1
N6a/b - NDUFS7
N2 in NDUFS8
FMN accepts 2e- simultaneously from NADH and transfers them one at a time to the Fe-S clusters
N1a seems to work as an antioxidant

QUINONE BINDING SITE

The **QUINONE BINDING SITE** is a cavity where quinone is reduced being able to carry two electrons.

The Q site is at the interface between the hydrophilic and the hydrophobic arms: it is formed by NDUFS2, NDUFS8 and ND1



Structural rearrangements are necessary to allow the CoQ head to move in and out of the cavity.

- Most of the cavity is hydrophobic with different polar residues. (NDUFS2, NDUFS8)
- The head of quinone is hydrophilic, this means that the internal face is necessarily hydrophilic. (ND1)

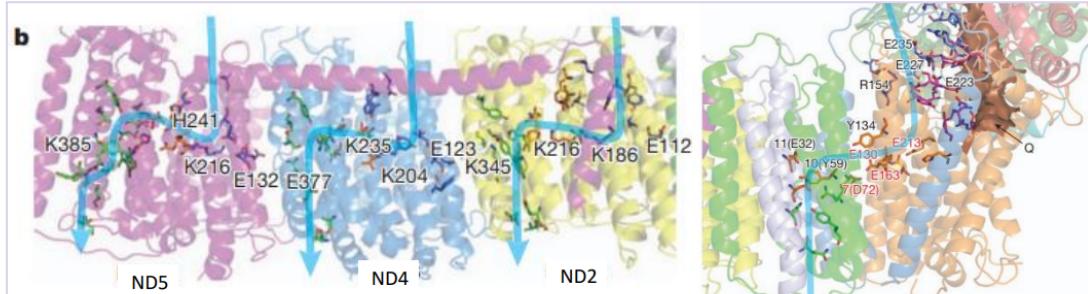
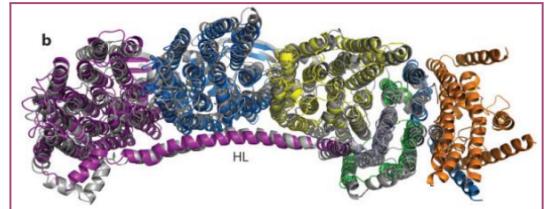
❖ MEMBRANE ARM

The subunits present in the matrix arm are *hydrophobic*, and are encoded in eukaryotes by **mitochondrial DNA**.

Membrane arm is mostly made of membrane **helices positioned across the membrane**. The arm is curved which is important for the interaction with complex 3.

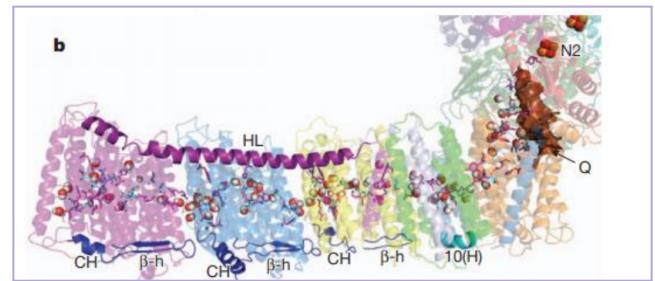
The subunits, which derive from antiporters, seem to present half channels that allow the protons to cross the membrane. There are **four hypothetical proton paths**. The subunits are very similar structurally speaking, being superimposable.

- **ND5**, which contains an important transversal helix we will talk about later.
- **ND4**
- **ND2**
- the last channel on the right is made of **ND1, ND3** and **ND6** subunits.



In the middle of the membrane there are (purple) **amino acids with hydrophilic residues** that are charged. Them being charged inside the membrane is uncommon, as the membrane is hydrophobic.

The presence of hydrophilic residues is probably due to the transporting of ions, however they still shouldn't be in a **horizontal dispositon**. This is because the transfer of protons is vertical, not horizontal.



Together with polar amino acids we also have **water molecules** embedded in the membrane, following the horizontal path of hydrophilic amino acids, which is again not really common as we are in a lipid environment.

One of the main features of the membrane arm is the presence of a very **long transversal helix**, **HL**, in the **ND5 subunit**, which is present in the middle of the membrane. Its tip is instead in a vertical position, far away from the rest of the (purple) subunit.

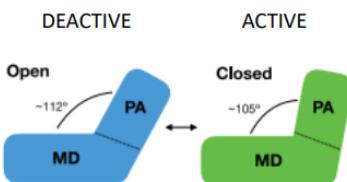
The alpha helix in the membrane is of **70 amino acids** which is not common so it wasn't predictable until they actually studied the structure. Also the helix is horizontal instead of vertical which is very peculiar.

HL helix was initially thought to stabilize the complex when translocation of protons happens. Now they think this might not be it because the mechanism would propose a movement of charges across the membrane, and this model isn't as definite for now. Its function might not be structural, and has yet to be confirmed.

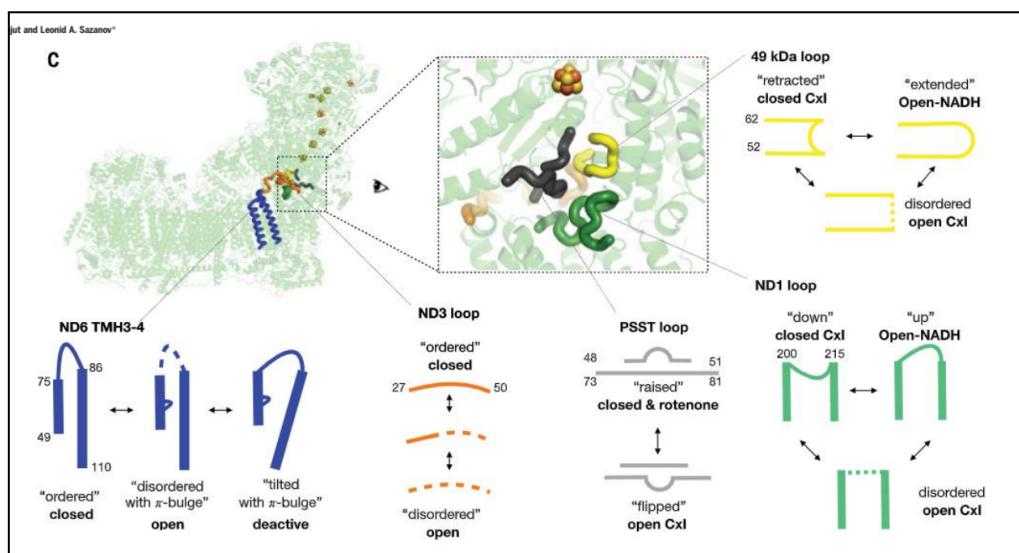
❖ OVERALL STRUCTURE

The overall structure of complex one can be found in **two conformations**, which **differ because of the matrix arm**.

- **open conformation (deactive)**, which has a wider angle, corresponds to the binding of NADH.
- **closed conformation (active)**, which has a smaller angle, corresponds to the **non-binding of NADH**.



Electron transfer is coupled with proton transfer across the membrane thanks to **small structural changes that occur in the protein**. The following are the structural changes:



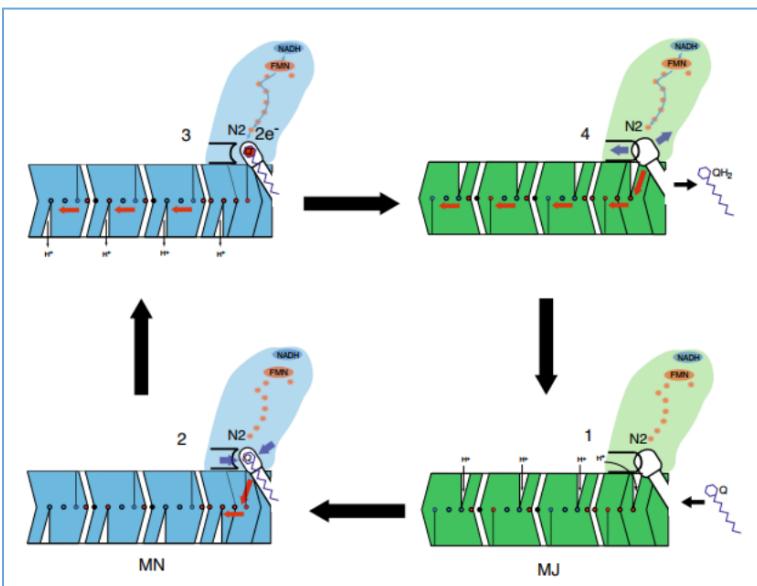
- In **49kDa** (NDUFS2) subunit, the **loop** is extended when the structure is open (NADH bound), the loop is retracted when the structure is closed (no NADH bound), and when there is no NADH present in the mixture, the structure is also closed, but the loop is disordered. In this subunit we have the quinone binding site.
- In **ND1**, the **loop** is facing down in the closed conformation (no NADH bound), while it's facing up in the open conformation (NADH bound). When NADH is not present at all we have a disordered loop.
- **PSST** (NDUFS2) subunit, the **loop** can be either raised or flipped
- **ND3** is fundamental for the working of the complex, containing a cysteine. It can be either ordered, or partially or completely disordered.
- **ND6** can also have an ordered (closed), disordered with π bulge (open), and tilted with π bulge conformation (deactive).

Bulge is where the head of the quinone should be. Thanks to small conformational changes, ghd bulge can either be flipped in or flipped out. It's not a big change but there are some differences that **deny the binding**. We can't talk about an occlusion, as it is not complete.

❖ MECHANISM OF ACTIONS: TWO HYPOTHESES

There are two hypotheses on the **coupling mechanisms from protons to electrons**.

1. INITIAL HYPOTHESIS: FOUR DIFFERENT STATES



- In **STATE1**, CoQ is not bound. CI is reduced, **NDUFS2 β1-β2 loop** with H384 is **flipped** into the Q cavity (black arc)
- in **STATE2**, CoQ enters in the CoQ binding site. The **loop retracts** because CoQ is present in the right conformation. The cavity "tightens" around the headgroup of CoQ, in order to perfectly fit the quinone.
- in **STATE3**, **2e- from NADH arrive** to CoQ via cluster N2, quinone is reduced, becoming **quinol**. Electrostatic interactions with charged residues in the E-channel initiate changes. The overall **4 protons are ejected** into the periplasm (IMS): **one for each antiporter subunit**.
- When the protons are released, former amino acids induce another conformational change. In **STATE4**, the enzyme is quickly reduced and key NDUFS2 β1-β2 loop with H384 **flips out**. Quinol is ejected into the lipid bilayer.

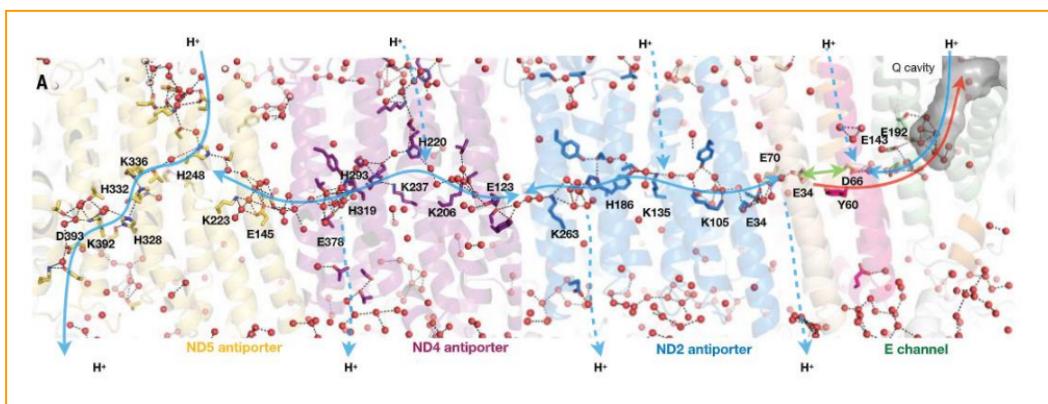
2. NEW HYPOTHESIS

In a new hypothesis, the model has been completely changed.

Taking a look at the position of water molecules and hydrophilic amino acids in the membrane arm, we can see a **very clear path**:

There are three different entry points, and one exit point in subunit ND5. Exit points in other subunits are not as clear. This means that coming out of other holes for protons is actually pretty hard.

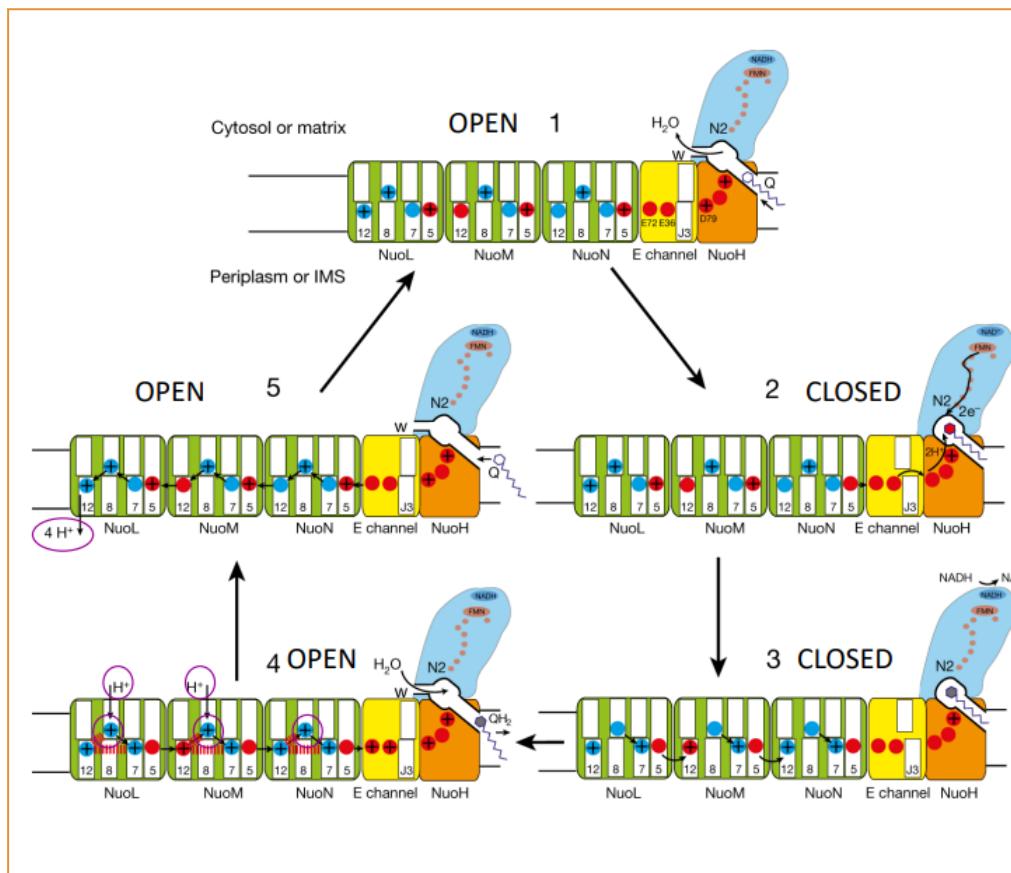
There is a new model where there is only one way to get out for protons: four protons all exit from simply one point.



According to this new model, individual steps that involve conformational changes are around the quinone binding site. There is a change in the **electrostatic interactions** between all the charged residues, which are responsible for signaling. In this case it's not a conformational change but an electrostatic one that couples translocation.

This model has five different steps:

- **STEP 1:** open conformation is present at the beginning: residues are charged, ion pairs are uncharged and quinone is yet to be present.
- **STEP 2:** when quinone binds to its binding site. This causes the formation of a close conformation, where aa are uncharged.
- **STEP 3:** with NADH oxidation, quinone is reduced and two H⁺ are taken from the central axis to complete the reaction, which results in the **redistribution of protons**. This favors the forming of an open, inactive, conformation
- **STEP 4:** transfer of charges which favored the release of reduced quinone. One water molecule enters the site and two more protons are uptaken from the water molecule.
- **STEP 5:** a series of H⁺ transfers along the central axis is initiated. It results in a total of 4H⁺ ejected from NuoL and the cycle restarting



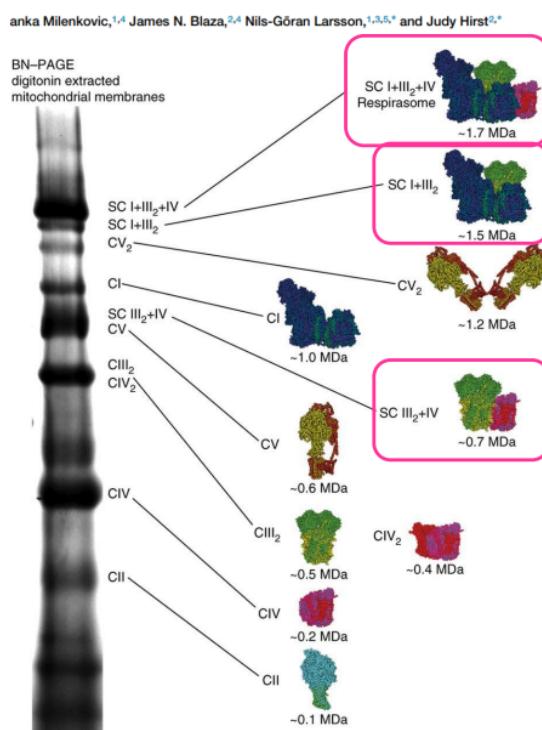
These two hypotheses are both sort-of valid; scientists are still debating.

Structural biology is in continuous evolution, especially because we are no longer using crystallography, which studies structures in only one conformation, but other techniques such as cryo EM.

SUPERCOMPLEXES

❖ INTRODUCTION

The advances in protein separation and purification show that respiratory complexes can be organized in supercomplexes



BLUE NATIVE GEL

Besides the **isolated complexes** we can see at the bottom of the gel, we also see other bands at the top, which contain **more than one complex**.

- Respiratory **complex 4** can be found either alone (at the bottom), or working as a dimer.

The most common **supercomplexes** (or respirasomes) observed are **Complex I/III, Complex I/III/IV, and Complex III/IV**.

We can see some examples in the gel:

- we have one supercomplex formed of two complexes 3 and one complex 4.
- we have one supercomplex formed by complex 1 and two complex 3.
- we have one supercomplex formed by complex 1 two complex 3, and one complex 4.

At the beginning we didn't think that supercomplexes could exist, because complexes would usually work as monomers. Then, thanks to in vivo studies, the presence of supercomplexes was confirmed.

In E. Coli supercomplexes have never been found. In most eukaryotic mitochondria however, there are many **supercomplexes with different stoichiometries**. In fact we can say that:

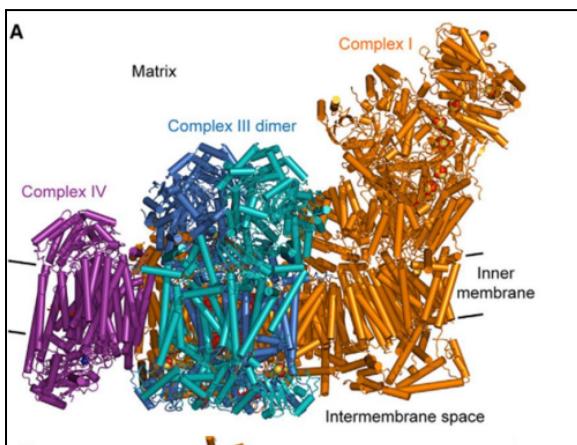
- 85%-100% of CI
- 55%-65% of CIII
- 15%-25% of CIV

are found in supercomplexes. That is to say that, for example, it is almost impossible to find complex I on its own, while it is very probable to find complex IV alone.

- Complex II has never been found in super complexes
- Complex V can have super molecular functions.

The technique used to separate supercomplexes is blue-native polyacrylamide gel: it allows **separation of membrane proteins in their native form**.

❖ WHAT IS THE ROLE OF SUPERCOMPLEXES?



Complexes can work alone, and are already pretty big, so why would they form a big complex together?

As we can see the arm of **complex 1** is curved and therefore fits perfectly with **complex 3** and **complex 4**. CIV is positioned on the toe of CI, and interacts with both CI and CIII.

Complexes one and three maintain more or less the same positions however **complex IV has been seen in at least three different configurations**: tight, intermediate, loose. The reason for this is not known, probably there's a functional meaning.

In the **SUBSTRATE CHANNELING THEORY**, the proteins are all near each other to favor transmitting electrons from one complex to the other.

However this is *probably* not the reason:

- substrates are unconstrained and free to move in any direction
- quinone is so hydrophobic that it can fuse in the mitochondrial membrane, which means that there is no need for the complexes to stay so close together.

Super complexes have possibly **OTHER FUNCTIONS** such as:

1. **ROS production is decreased.**
2. **the presence of supercomplexes favors the assembly of other complexes:** for example the presence of complex I helps the formation of complex III with its membrane arm. The formation of the complexes is modular, and they aggregate when they're not completely built.
3. **Regulation of the speed of the respiratory chain** according to the energy requirements of the cell. However recent studies are in contrast with this hypothesis
4. the formation of supramolecular complexes **prevents of protein aggregation**, which is seen from the cell as a trigger for cell death. Having many proteins, if You don't form supramolecular complexes proteins aggregate by themselves. Protein aggregation is

In the 50s, the **SOLID STATE MODEL** was created. According to it, substrates are channeled directly from one enzyme to the next, with the enzyme components assembled into a **huge supramolecular energy-converting machine**

In the 80s, **RANDOM COLLISION MODEL** said that **complexes only exist as isolated** and not aggregated.

In the 2000s, the **PLASTICITY MODEL** was created as a mix of solid and collision models. **Both isolated forms and respirasomes exist at the same time** and the quantity depends on the energy requirements of the cell. This model says the supercomplexes form as **isolated complexes first** and then join together to form different supercomplexes.

- for example, first complexes I and III are created independently, then they get together in a subsequent step.

In 2022 the **COOPERATIVE MODEL** was created. It takes into account that:

1. There is **no plasticity for complex I**, which means it is only present at super complex levels,
2. The assembly of respirasomes happens with the formation of **pre-complexes**, which, once they join together, can finish the formation of complexes.
 - for example, first complexes I and III join together when they are not completed, after they joined ultimation of complex I can happen.

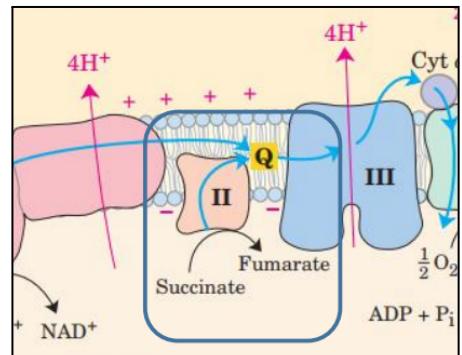
This means that probably supercomplexes happen because they garant the biogenesis of respiratory complexes.

COMPLEXES II AND III

❖ COMPLEX II

Complex II is also named **succinate dehydrogenase (SDHA)**.

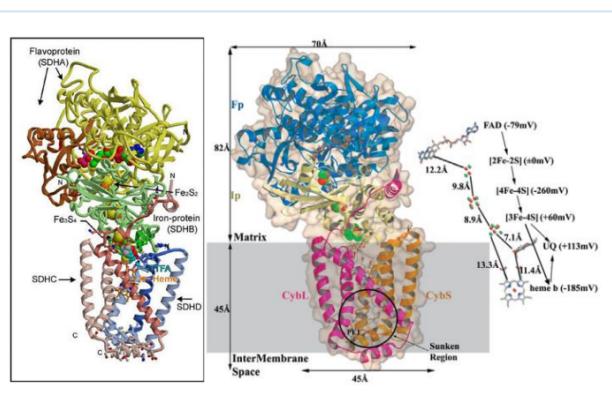
- It's the **same enzyme of the Krebs cycle**, being part of both processes: this means that it is a direct **link** between **Krebs Cycle** and **aerobic respiration**.
- Complex II is the only respiratory complex which **doesn't pump protons** into the intermembrane space. Rather, it **takes electrons from succinate and fumarate** (which can come from various pathways) and **gives them to ubiquinone**.
- It's an integral membrane protein.



We can therefore say that it couples two distinct chemical reactions:

- the reversible **oxidoreduction of succinate and fumarate** (soluble domain)
- the **oxidoreduction of CoQ** (membrane-spanning domain)

This means that succinate coming from Krebs Cycle, is converted into fumarate, which is oxidized in order to reduce CoQ.

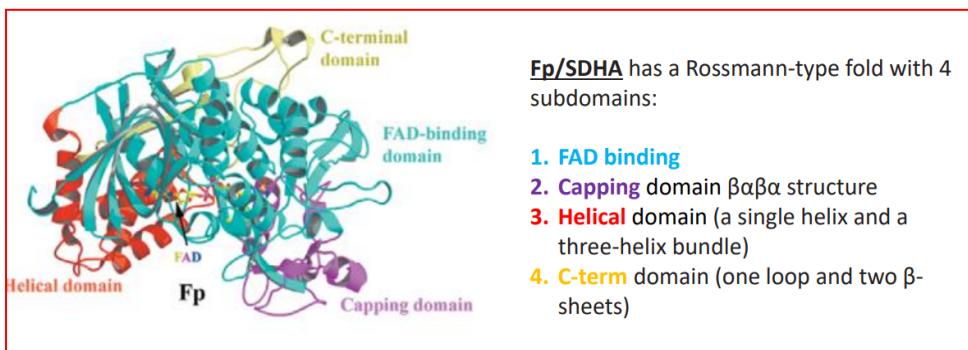


Structurally speaking this complex is much smaller and it is **composed of four subunits**:

- large **flavoprotein** which covalently binds FAD
- iron-Sulfur protein** which coordinates **3 Fe-S clusters**
- CytochromeBL** and **CytosomeBS** coordinate a **heme b** group.

All these proteins are encoded by the nucleolus, not the mitochondria.

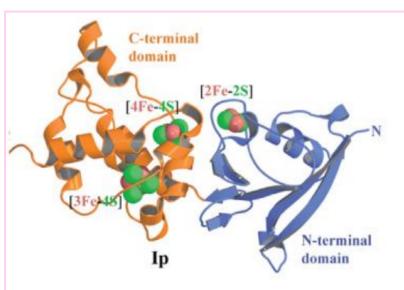
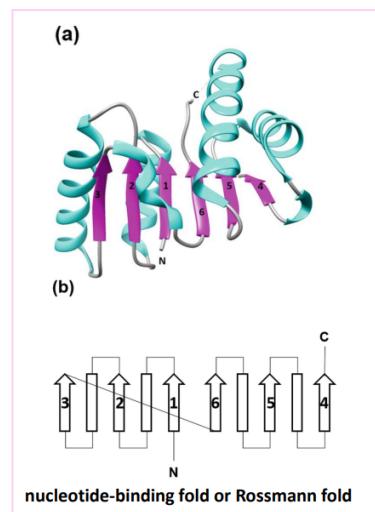
There are several prosthetic groups. SDAH protein is **multidomain**, with portions having different folding and functions.



→ FLAVOPROTEIN (SDHA)

FAD is bound thanks to a structural domain recurrent in nature: **ROSSMAN FOLD DOMAIN**, which is an alpha\beta twist domain. It's so recurrent that also complex I binds NADH with a similar domain.

- its classical structure has 6 parallel **beta strands** and 5 **helices**.
- The domain is made **alternatively** of alpha and beta sheets with a linker that's diagonal.
- the strand order is not consecutive**
- the first strand is placed in the middle of the sheet; additional strands are laid down consecutively outward to one edge.
- the helices are on one side for half of the sheet, and on the opposite side for the other half**
- the helices are mainly amphipathic b/c the sheet is predominantly hydrophobic



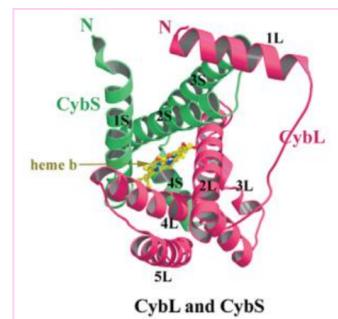
→ IRON-SULFUR SUBUNIT (SDHB)

This subunit has two different subunits:

- N TERM** is an helix/strand mixed domain. One loop coordinates the [2Fe-2S] cluster
- C TERM** is an all helix domain. It has six α-helices and coordinates the [4Fe-4S] and [3Fe4S] clusters

→ CyBs and CyBl (SDHC, SDHD)

- the two cytochromes are called **CyBS** and **CyBL** because cytochrome L is larger. In particular, L has an extra helix that's not present in S.
- the two subunits fold together**
- N-term in the matrix side and C-term in the IMS side for both



A **site for ubiquinone** is needed because it must be reduced. Debatably there are **two sites** which are both able to bind ubiquinone, although one is reduced for each electron passing through the iron sulfur clusters.

In *E. coli* complex II makes a trimeric structure.

The **ELECTRON TRANSFER** starts from **succinate** electrons are passed to **FAD** and then transferred, thanks to FeS, to **ubiquinone**. **ET is fast** as none of the electron carriers are placed at a distance greater than 11Å. **Heme b** here is not in the normal electron path: it's close to the electron path but not directly involved. It seems like it works as n1a FeS, working when there's a possibility of escape of electrons from the normal electron transfer path, acting as a buffer. **This leads to a reduction of ROS production**. Heme B is therefore not present in the normal path and its used only in special cases.

Similarly to complex I, we have iron sulfur clusters electron acceptors:

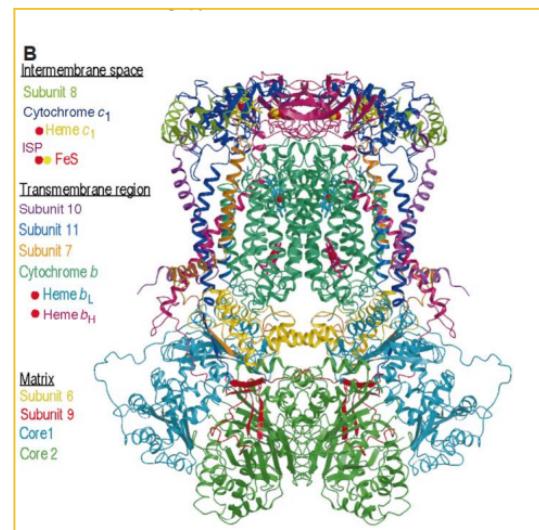
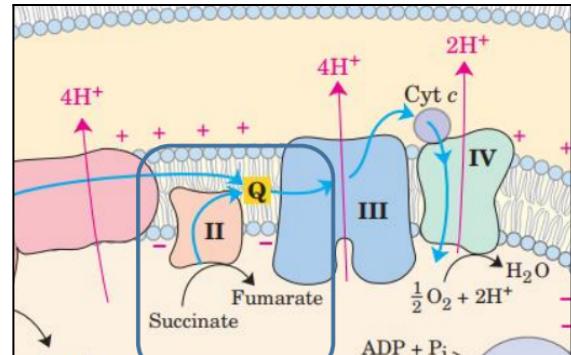
1. Two electrons are released
2. Only one of them is taken up by the sulfur clusters.
3. This means that the speed must be very high and the process must be fast since one electron goes in at the time, and the other cannot stay at the entrance site for too long.
4. Two electrons lead to ubiquinone total reduction.

❖ COMPLEX III

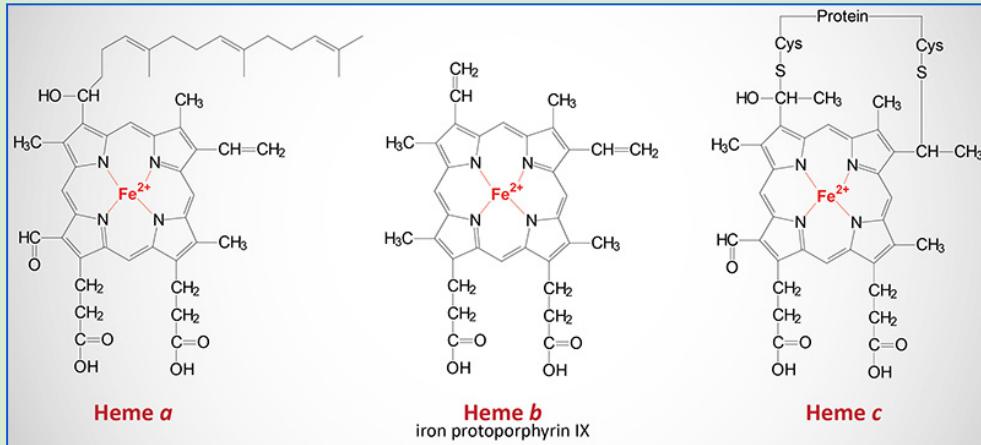
Complex III can also be called **Cytochrome bc₁** complex. It doesn't really matter if ubiquinone has been oxidized by complex I, by complex II, or other enzymes: the electrons are passed onto complex III.

- Complex III is able, for each molecule of ubiquinone, to **pump four protons into the intermembrane space**. The two electrons from ubiquinone are **then transferred to cytochrome C**, which is reduced by complex IV.
- Complex III is an **obligate dimer with an ax of symmetry**. Each monomer is composed of 11 subunits.
 - The portion facing the **intermembrane** space contains many subunits, three of which are the most critical and must be remembered: ***cytc1*, ISP (iron sulfur protein)**
 - The portion of complex III which is placed in the **transmembrane** region is mainly constituted by alpha helices. Also in this case there are many subunits, but the most important one is the catalytic one, **cytochrome B**.
 - In the **matrix side** we have other subunits, among one core 1 and core 2 are the most important, although they do not possess a catalytic function.
- Critical subunits (***cytc1*, *cytb*, ISP**), which possess catalytic functions, have **active redox centers**.
- All subunits are nuclearly encoded except for cytochrome B
- We **cannot eliminate complex III**. If we abolish it, it doesn't matter if complexes I and II are present, we won't have the following steps.

Complex III being also called cytochrome bc₁ complex, it **contains cytochrome b and cytochrome c1**; it also gives electrons to cytochrome c, but they are three different molecules.



In general **CYTOCHROMES** are a category of proteins. Their name derives from the fact that when they were discovered they were colored thanks to their heme group, hence having iron which can be oxidized yielding a reddish color.



Cytochromes always contain a **heme group**.

We can have different types of heme groups: a, b, c.

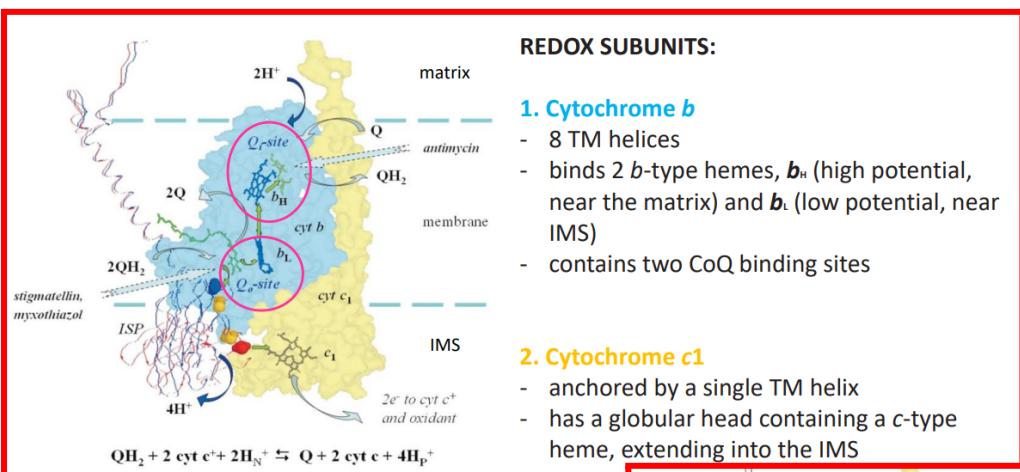
→ **heme B** is the prototype: an iron is coordinated by histidines

Heme a and c have the same bulk structure but with different variations:

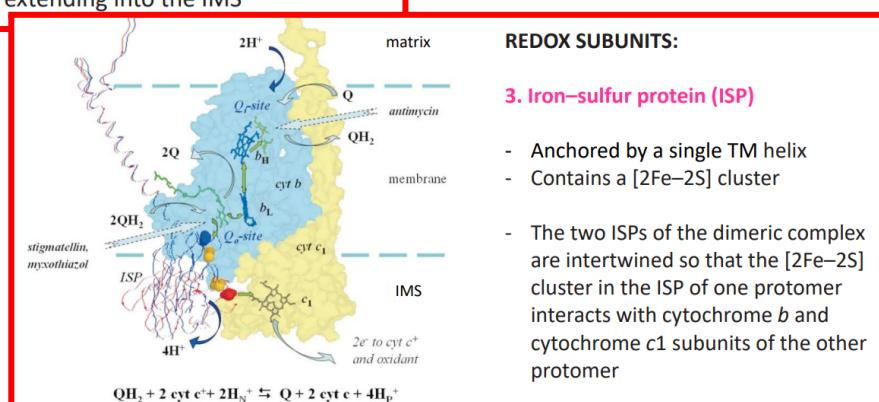
→ **heme A** presents a methyl side chain is oxidized to a formyl group and a hydroxyethylfarnesyl group is attached to the vinyl side chain

→ **heme C** has the 2 vinyl (C=C) side chains, which are covalently bound to **cysteines**.

Now let's look at the three catalytic cores of complex III:



As we can see **cytochromes b** and **c1** both contain transmembrane helices. In cytochrome b there are two *b-heme groups* (l and h), each corresponding to a different CoQ binding site. In cytochrome c1 we have a helix to anchor it, and the more important globular head, which contains a *c-heme group*.



ISP has a single transmembrane helix that anchors it to the membrane and a globular head in the intermembrane space.

The iron sulfur protein from monomer a is entangled with monomer b and vice versa.

This is why complex III is an obligate dimer.

Cytochrome C1 and ISP are actually very similar: while they both have helices that fix them into the matrix, the first contains an heme group, the latter an Fe-S cluster.

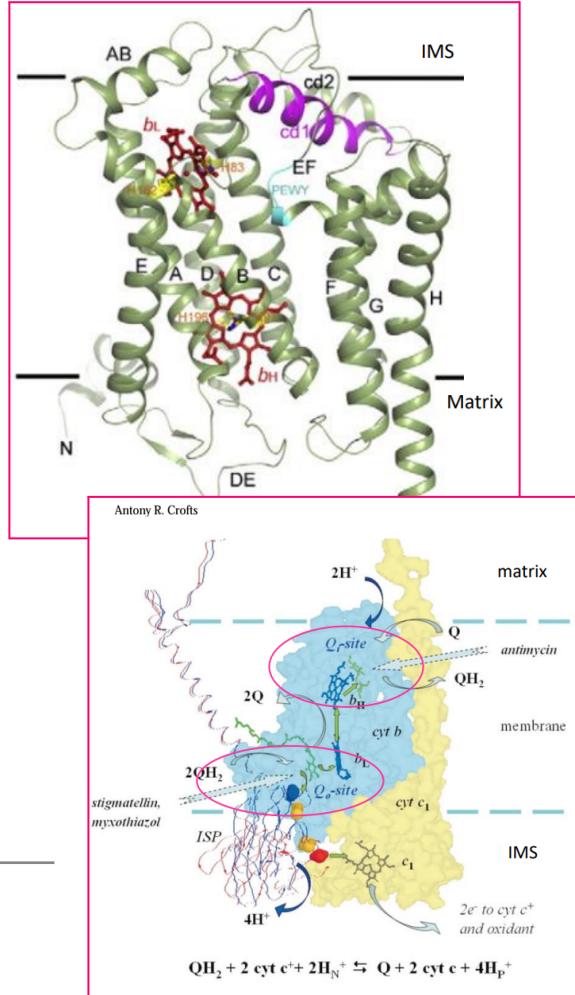
- The **extrinsic domain** (linker between helix and globular head) of the **ISP** subunit can be found in **different conformations** depending on the binding of different inhibitors → different functional states

Instead, **cytochrome B** is an all helical protein.

- **two bundles** of alpha helices: A to E, F to H.
- the **hemes groups are coordinated by histidines**
- the bundles contact each other on the matrix side, but are separated on the IMS creating a gap
- gap that is **bridged by the cd helices** (**cd1** and **cd2**) and by the **EF loop**

CytB has **two binding sites for ubiquinone**:

- **Q oxidation site (Q_o)**: binds ubiquinone coming from complexes I and II. It is formed by *CytB* (helix C, helix **cd1** and ef linker, near heme_L) and **ISP**.
- **Q reduction site (Q_i)**: it is located near heme b_H in proximity to the matrix.



Different **inhibitors** can bind the two sites preventing CIII activity in different ways by **blocking the conformation of ISP in different positions**. This allows us to define how the protein functions.

- Qi site binds antimycin
- Qo site binds stigmatellin and myxothiazol

Different inhibitors bind different Q sites, and lead to different positions of the 2Fe-2S cluster.

- **b state**: obtained with stigmatellin, electron transfer is not allowed because [2Fe-2S] are too close to cytB
- **c1 state**: obtained with antimycin, fast ET. In this case we have [2Fe-2S] close to cyt c1
- There are also several **intermediate** states for the globular head of the iron sulfur protein. They are obtained with myxothiazol

We can see a comparison between c1 and intermediate state: **what changes is the globular head of the iron sulfur protein**.

These positions correspond to different states in the catalysis.

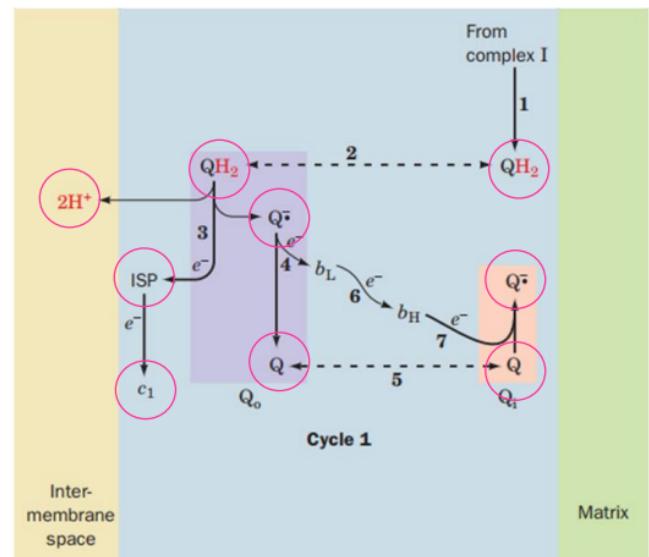
Complex III allows **transmission of electrons from reduced CoQ (which has 2 electron carrier) to cytochrome C**, which is a soluble protein. CC however only **carries one electron at a time**.

This is not the same mechanism as before: we don't have a fast transfer because the transmission doesn't happen in the same protein, but in two.

In this case there is a **bifurcation of the path**: one electron goes into Cytochrome C, one in the Q_i site: this is called the **Q CYCLE**.

It's made up of **two semi cycles**:

1. we start with one reduced CoQ molecule, coming from complexes I, III or other sources.
 - a. CoQ binds at the level of the Q_o-site.
CoQ reduces ISP (and then heme c1 of cytc1) with one of its two electrons. This is associated with the pumping of two protons outside the matrix.
 - b. Now we have a partially oxidized ubiquinone (**semi quinone**) which cannot stay in the original site. Semiquinone gives the **remaining electron** is passed onto heme b_L and then heme b_H. Nevertheless, the electron can't stay at the b_H site because it produces ROS. The electron is accepted by another ubiquinone molecule present in the Q_i site, leading to the production of **semiquinone**.



2. The second semicycle happens in the same manner: one electron is passed onto Cytc1 and another one goes in the b_L, b_H path. However, at the end of this second path, we don't have a completely oxidized quinone anymore, but the semiquinone from before: the **semiquinone** accepts another electron becoming **completely reduced**. Two protons are pumped into the intermembrane space.

At the end we have a molecule of completely reduced CoQ, two molecules of reduced cytochrome C1, and four protons in the intermembrane space.

Cytc1 electrons will go to cytochrome C.

MECHANISM OF ACTION

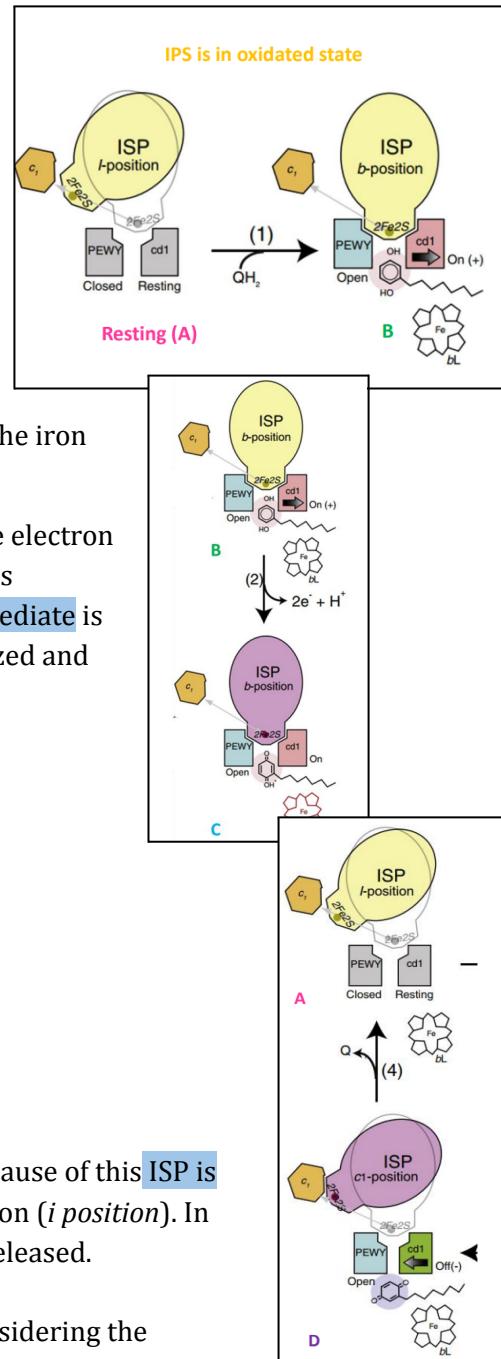
→ Binding affinity modulated ISP extrinsic domain (ED) motion switch hypothesis

The mechanism is made up of **four steps**:

- 1. RESTING PLACE:** ISP is oxidized and its head can assume several positions (*i*-position), sampling the face of cytochrome B trying to understand if ubiquinone is present.

It searches for it, assuming a lot of intermediates. If ubiquinone is not present, the ISP head cannot be blocked.

- 2. TRANSITIONAL STEP:** the molecule of coenzyme Q trapped between PWEY and cd1, is sensed and the **ISP stops** (*b*-position). PWEY takes its open conformation, and the iron sulfur cluster protein is close to the head of coenzyme Q.
- 3. One electron is given to ISP, which undergoes reduction:** the electron will then be passed to cytochrome c1. The second electron is transferred to b₁ (and then b_H), and the **semiquinone intermediate** is generated. Because of this, original CoQ is completely oxidized and ISP isn't in the *b*-position anymore (*c1 position*)



- 4. Now electrons go from cytochrome c1 to cytochrome C: because of this ISP is oxidized again and it now goes back into its original condition (*i* position). In this second cycle CoQ is completely oxidized, hence being released.**

The Q_i site is not considered in this model, because we are only considering the transfer of electrons.

Cytochrome C (soluble carrier like quinone) is the final acceptor of complex III.

- It's another helical protein with an heme group like all cytochromes.
- It can be either oxidized or reduced when complex III is reduced.
- It's linked with complex 4.

COMPLEXES IV AND V

❖ CYTOCHROME C

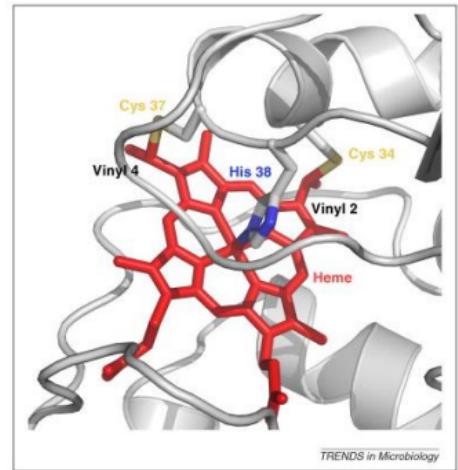
Cytochrome C is a very small soluble protein **between the inner and outer mitochondrial membrane**.

- It shuttles electrons between complex III and IV thanks to the presence of the heme c group, covalently bound to the protein thanks to its bond with two cysteines.
- The iron of the heme group can be found oxidized or reduced depending on the presence of electrons.

The **structure has highly conserved** amino acids, **lysines** in particular.

- lysines allow the interaction of cytochrome C₁ with complex III.
- If the interaction doesn't occur, the electrons in cytochrome c₁ from complex III cannot be passed to cytochrome C.

Electrons are then given to complex 4.



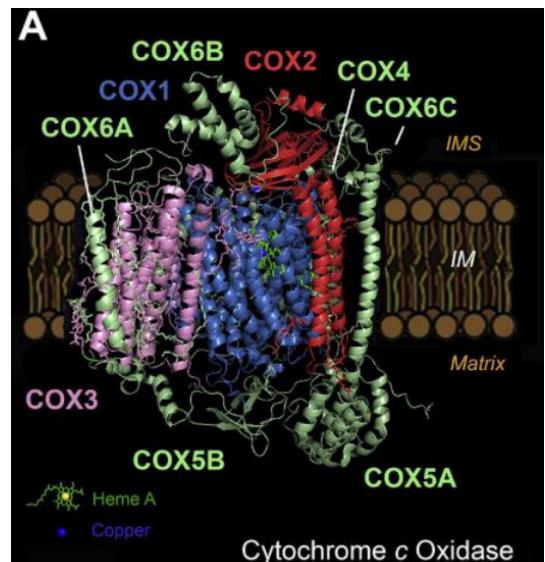
❖ COMPLEX IV

Complex 4 (or **Cytochrome c oxidase**) has:

- three core subunits (*pink, blue, red*): COX1, COX2, COX3.
- other 11 accessory subunits (*green*)
- most subunits are transmembrane ones. In general we can say that CIV is completely membrane bound, with just a few loops out of the membrane.

Cytochrome c oxidase catalyzes the **one-electron oxidations of 4 consecutive reduced cytc molecules** and the concomitant **four-electron reduction of one O₂ to water**.

- The process is coupled with the pumping of 4 hydrogens protons.
- The core subunits are mitochondrially encoded, others are instead nucleic.
- They're mostly intermembrane subunits.



NDUFA4 is a subunit of complex IV, but it's called a complex I subunit because it was initially assigned to it.

- it has one twisted helix
- it's not tightly bound to CIV, which means it's probably not catalytic, and could have a regulatory role.

CORE REDOX SUBUNITS

Core subunits are COX1, COX2, COX3.

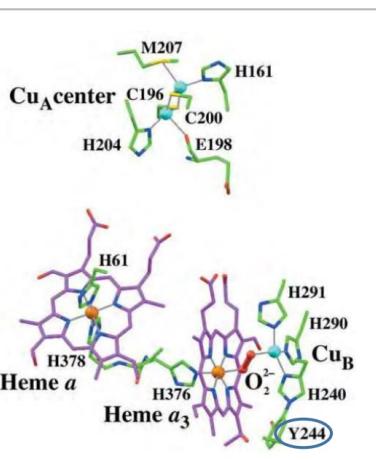
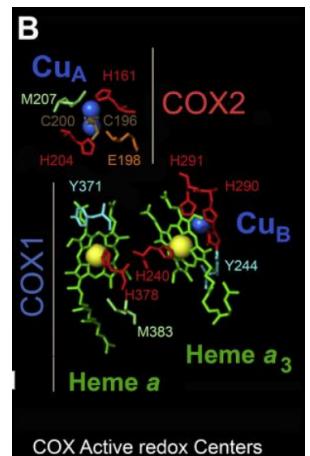
- Extremely conserved
- Highly hydrophobic, integral **membrane proteins**.
- **COX1 and COX3** have no extra membrane domains
- **COX2** has one extra membrane domain that extends into the IMS: a β-barrel anchored by two TM helices that bind soluble Cyt c (at the level of conserved Lys ring)

Core subunits present metal clusters, apart from COX3. However COX3 is still important: if it's not present the complex is not assembled.

COX1 and COX2 contain four redox metal active centers, where electron transfer happens.

1. ET occurs from soluble Cyt c to a **di-copper center (CuA)** located in a loop region at the interface between COX1 and COX2.
2. From CuA, electrons flow linearly to **heme a in COX1**, which transfers electrons to the **heme a3-CuB binuclear center** where O₂ is reduced to water

It is therefore at the level of COX1 that O₂ is reduced to water.



- **CuA** two copper ions are bridged by the sulfur atoms of two **Cys** residues, giving it a geometry similar to that of a [2Fe–2S] cluster.

- The Fe of heme a₃ lies only 4.9 Å from CuB

Tyrosine 244 (Y224) is necessary for the catalytic mechanism because it's necessary for oxygen's reduction to water.

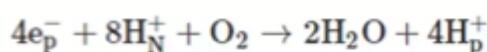
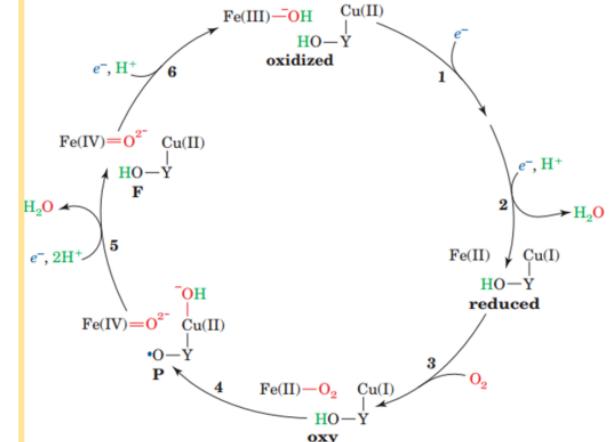
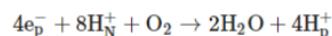
The **reduction of O₂ to 2H₂O** takes place at the ***heme a3-CuB*** binuclear center.

As **substrates** we have: 4 electrons from CytC, 8 protons and one oxygen molecule.

As a **result** we have: 2 water molecules and the pumping of 4 protons.

1. the oxidized binuclear complex is reduced by **2 e⁻** from two reduced CytC. One water molecule is produced.
2. A **H⁺** is taken from the matrix and Tyr 244 is in **phenolic state** (TyrOH)
3. O₂ binds the Fe(II) in heme a3 (entering in the **oxy state**)
4. Tyr 244 donates the H⁺ to an oxyferryl complex, assuming a radical state (**compound P**).
5. **1e⁻** from CytC and **2H⁺** reconverts Tyr 244 to the phenolic state releasing H₂O (b⁺ **compound F**).
6. Last, **1e⁻** and **1H⁺** complete the catalytic cycle by
7. forming the oxidized complex

Mechanism of reaction



As we can see, we have a total of 8 protons at the beginning, but, since 4 are used in the reduction pathway for oxygen, only 4 are pumped to the matrix. In particular we can **distinguish protons in two types**:

- 4H⁺ (**chemical**) are taken up from the matrix during the reduction of O₂ to yield 2H₂O
- the 4e⁻ reduction reaction is coupled to the translocation of 4H⁺ (**pumped**) from the matrix to the IMS

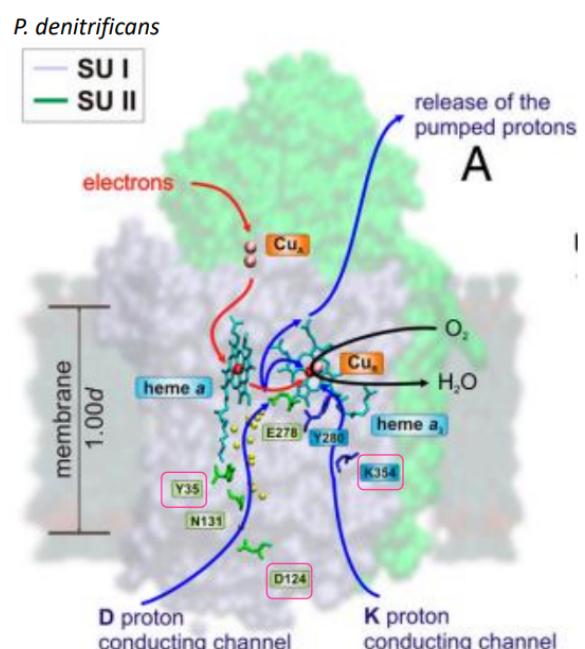
We don't know the actual path of the protons. Maybe there are **two different channels** for protons: **d and k proton conducting channels**.

K-channel (critical Lys):

- from the matrix side of the protein to Tyr 244
- not connected to the IMS
- Supply chemical protons for O₂ reduction

D-channel (critical Asp):

- from the matrix to the heme a₃-Cu_B center
- connected to the **exit channel** towards the IMS
- pump 4H⁺ from the matrix to the IMS
- Contributes to the supplement of chemical H⁺ for O₂ reduction



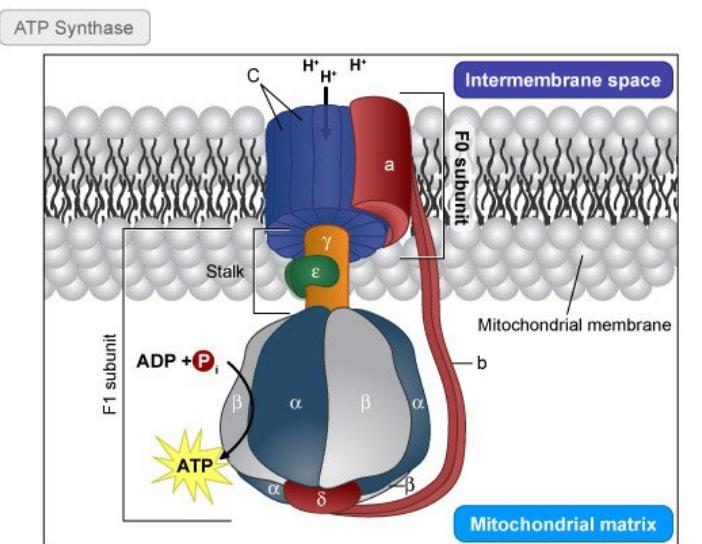
❖ ATP SYNTHASE

The electrochemical gradient created in the previous steps can be used to catalyze an unfavorable reaction, which is the production of ATP, a highly energetic molecule.

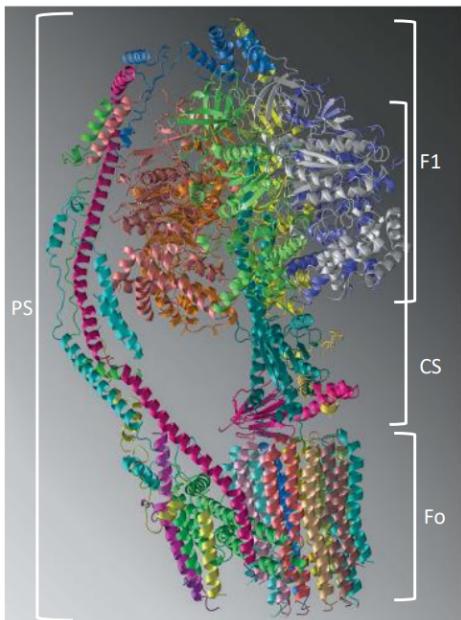
ATP synthase is composed of two functional units.

- **F₀** is a water insoluble **transmembrane protein** (8 types of subunit)
- **F₁** is a water soluble catalytic part in the **mitochondrial matrix** (5 types of subunit)

The two portions are together thanks to the **peripheral** and **central stalk**.



Dept. Biol. Penn State ©2004

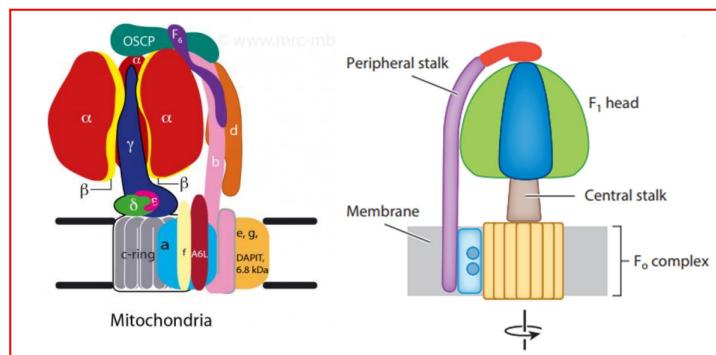


This enzyme is **reversible** so it's able to **catalyze both formation** and **disruption** of ATP.

This choice depends only on the electrochemical gradient. As long as it is present, ATP is produced, but if for some reason it collapses, ATP is degraded (hydrolysis) and so is the mitochondria.

- electrochemical gradient must be kept as stable as possible.

Disrupting the central stalk and/or overall structure of ATP synthase, ATP cannot be produced anymore.



F₁ contains catalytic subunits:

- 6 subunits (3 α and 3 β alternated) that form a ring.
- the protein inside the ring is gamma protein δ .

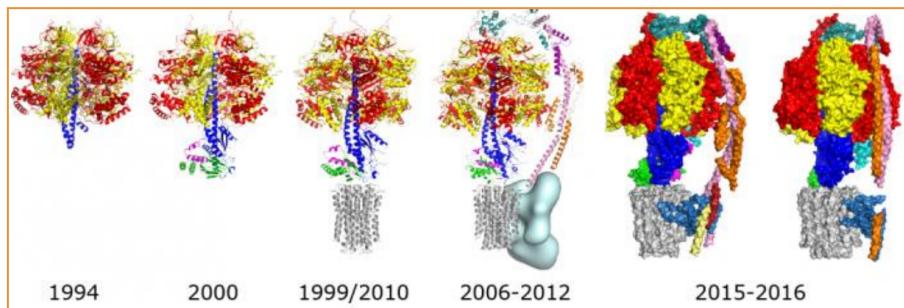
F₀ contains pumping subunits:

- ring of several C proteins
- protein a that is a proton channel
- protein b that forms the peripheral stalk

The structure is similar to a turbine:

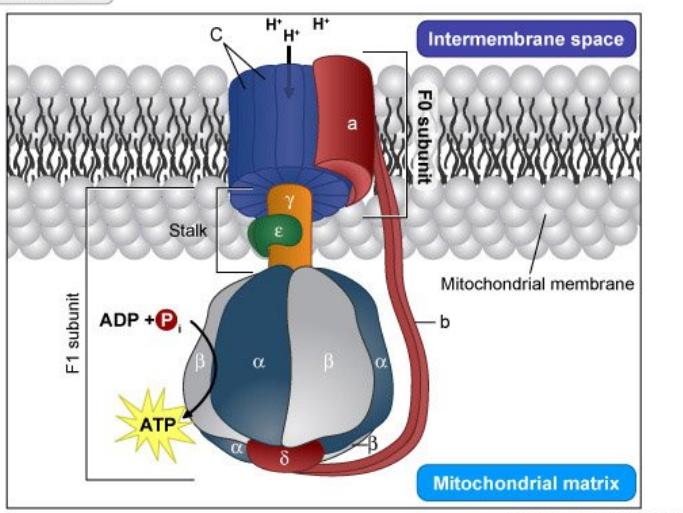
- the **ROTOR** part can rotate. It's made up of the C ring and the gamma stalk
- the **STATOR** part is fixed. It's made up of the remaining subunits.

ATP synthase was crystallized many times and in different species: each time, especially with cryo EM, there were progresses in the discovery of its structure:



- In 2010, the structure of the bovine **F1 C RING** was revealed that it translocates 2.7 H⁺ per ATP synthesized (the most efficient ATP synthase)
- Between different species there are differences in accessory subunits (in the peripheral stalk) and also in the C ring.

ATP Synthase



F1 DOMAIN is the catalytic subunit in the matrix part: from it, ATP exits.

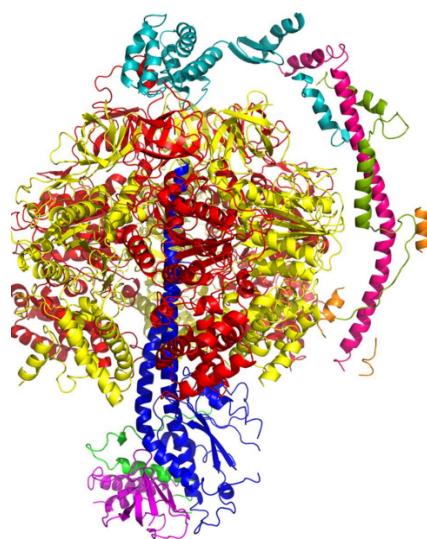
- **globular** assembly of 5 different subunits (α , β , γ , δ , ϵ)
- 3 **α -subunits and 3 β -subunits** (20% identical) present a nucleotide-binding motif (P-loop) and are arranged alternately around a central α -helical coiled-coil in the γ -subunit
- the **γ -subunit protrudes** from the $\alpha\beta\beta$ subcomplex (**central stalk**) creating a sort of "foot", which interacts with the c ring.

thanks to the **central stalk** we have the conversion of energy from the motion of the c ring to the modification in α , β subunits, which leads to ATP production.

Looking at the F1 portion from an **ABOVE VIEW**, we can see the alternance between alpha and beta subunits, considering that the folding of α , β subunits is identical. At the bottom there are all helices domains. The nucleotide binding pocket is at the interphase of the alpha and beta subunits. In particular, pockets are defined mostly by the β -sub, with an essential Arg from the α -sub.

We can also look at the **SIDE VIEW** of F1. The protein is **asymmetric** due to the presence of the γ subunit and because each pair of α and β subunits adopts a different conformation when empty or binding ATP or ADP, each with a different substrate affinity.

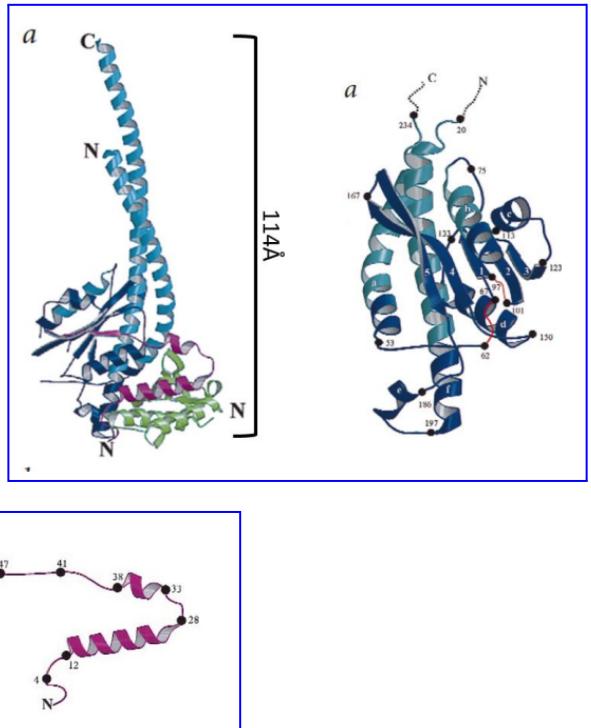
This part is usually fixed (stator)



CENTRAL STALK

The central stalk **interacts with the c-ring of the Fo** TM domain that drives rotation of the stalk during ATP synthesis.

- The **γ -sub** penetrates the catalytic domain with a **coiled-coil helical structure**, where two alpha helices fold one across the other making one bigger helix.
- The remaining part is an $\alpha\beta$ domain made by a **5-stranded β -sheet and 6 α -helices** (a-f), partly forming a **Rossmann fold** (no nucleotide bound)



- The δ -sub and ϵ -sub interact with the Rossmann fold of the γ -sub, forming a foot associated with the c-ring and couples the TM pmf to catalysis in the F1 domain:
 - **δ -sub** has a N-term β -sandwich and C-term α -helical hairpin
 - **ϵ -sub** is in a helix-loop-helix structure

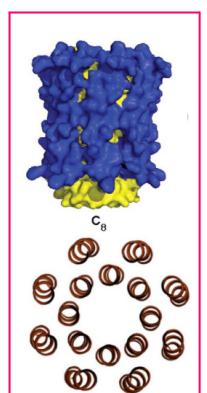
PERIPHERAL STALK

The peripheral stalk is made up of several different subunits, like OSCP (oligomycin sensitivity conferral protein), b, d and F6.

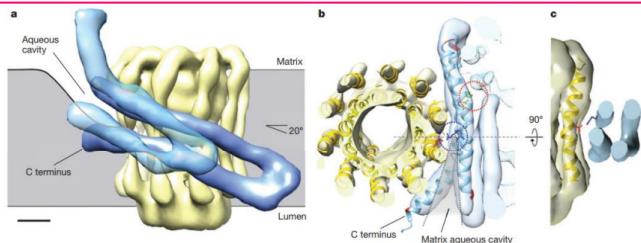
- it's part of the static components of ATP synthase

F0 DOMAIN

- the **c-ring interacts extensively with the foot** of the central stalk via subunits γ , δ , and ϵ
- the c-ring is made up of **two assemblies of alpha helices**, one internal and one external one.
- the **composition of the c-ring** (in particular the number of its subunits) **varies on the species**. The most efficient composition is present in metazoan (a type of sponge), with eight subunits.



PROTON TRANSFER IN THE C RING



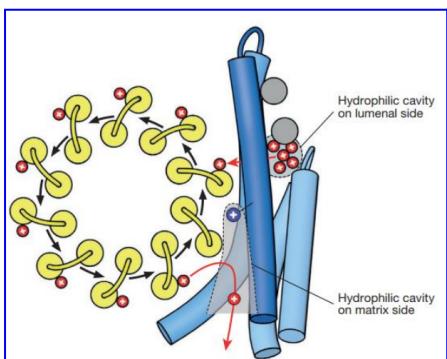
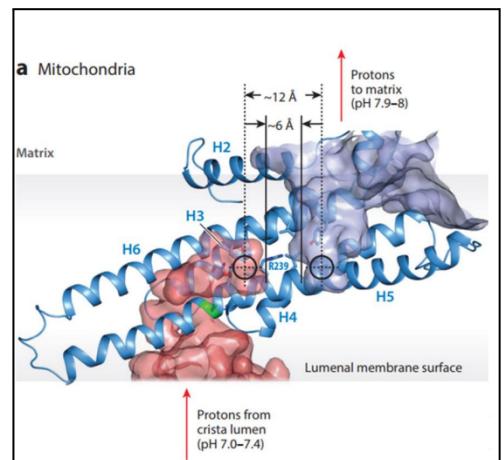
A recent discovery was how the protein actually works. The **interaction between the c ring** (light yellow) and the **a subunit** (light blue) wasn't understood until recent times.

Now we know that **a subunit creates a double transversal hairpin**, which is not so common and it usually cannot be predicted.

- these helices are placed inside the membrane so they are deeply hydrophobic
- there are still some clusters of polar residues, which are **important in order to capture protons and transferring them to the c ring**. Two amino acids are important in particular:
 - an **Arg** couples proton translocation to c-ring rotation
 - Arg lies next to a proton binding **Glu** in the c subunit.

The proton channel is not straightforward, rather, we have two half channels.

- protons must go back to the matrix
- the **hairpins of the a subunit** are in a mainly negatively charged area (as we can see it's red) and are able to **attract protons** and fix them into place.
- the proton entrance channel contains the Glu, while the proton exit channel contains an Arg.



Protons reach the c ring thanks to their attraction to Glu.

1. when protons reach the c-ring, they have **compete with a strictly conserved Arginine for interaction with c-ring Glu**, which carry the proton around the c-ring.
2. when the c-subunit **approaches** the hydrophilic **half-channel on the matrix side**, the **Glu releases** the proton into the matrix

Another function of ATPase has been found, which is totally unrelated to its catalytic function. ATPases present in eukaryotes can form dimers of different shapes. These **dimers can modulate the structure of the inner mitochondrial membrane** by deposing the membrane. These dimers are givinging **different shapes to the mitochondrial cristae invagination**. Hence dimers are not randomly placed.

Without the dimers, cristae are not properly structured. In these cases ATP won't be produced properly. We can say that biogenesis of cristae depends on subunit present only in ATPase.

CELL CYCLE

Living **organisms** reproduce their **cells** by repeated rounds of cell growth and division:

- **CELL CYCLE:** ordered sequence of events in which the cell duplicates its contents and then divides in two. It is a general mechanism, from bacteria to mammals.

In **unicellular organisms** it is connected to the formation of new unicellular organisms.

In **multicellular organisms** it is connected with the regeneration of cell tissues through the formation of new cells.

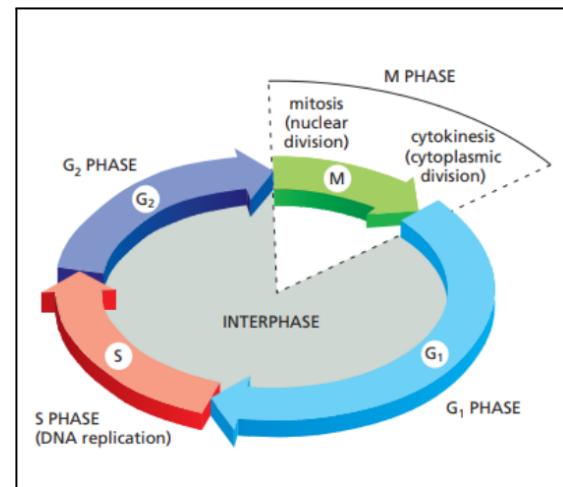
Dysregulation of the cell cycle leads to either lack of tissue (→ atrophies), or uncontrolled proliferation (→ cancer).

Details change from organisms to organisms. The **general mechanism** is however conserved:

1. **TRANSMISSION OF GENETIC INFORMATION:** DNA replication & Chromosome segregation
2. **DUPLICATION:** duplication of organelles and macromolecules
3. **GROWTH** and **DIVISION** must be coordinated.

The cell cycle is tightly controlled in two big phases:

- **M PHASE:** it usually requires less than one hour and it comprises two major events.
 - **nuclear division:** copied chromosomes distributed into a pair of daughter nuclei
 - **cytokinesis:** cytoplasmic division
- **INTERPHASE** is the phase between two M phases, where the cell does not divide, rather it grows. DNA replication also happens in this phase. It can last from a few hours to a day. It comprises three different phases: **G1 → S → G2.**



- **G1 PHASE:** in the G1 phase cells grow before chromosome duplication. There is no discernible change in chromosome shape or size, but the cell grows and accumulates the components. At the end of the G1 phase, if **physiological conditions are not met**, cells either delay the process until they are not ready or they go into the G0 resting state, which can last days or years. Cells in G0 can re-enter in the cycle upon proper stimuli.
- **S PHASE:** DNA synthesis occurs. Two identical sister chromatids are created.
- **G2 PHASE:** A gap phase allowing growth and preparation for separation of the sister chromatids
- **M PHASE:** During mitosis sister chromatids separate to opposite sides of the cell thanks to the mitotic spindle and, with cytokinesis, the cell divides to create the two daughter cells.

❖ CHECKPOINTS

Usually, there are **CHECKPOINTS** before the entrance in the next phases of cell cycle.

→ **checkpoints**: biochemical switches initiating specific cell-cycle events: they are like gates, and only if they allow it, the cell can start the cell cycle.

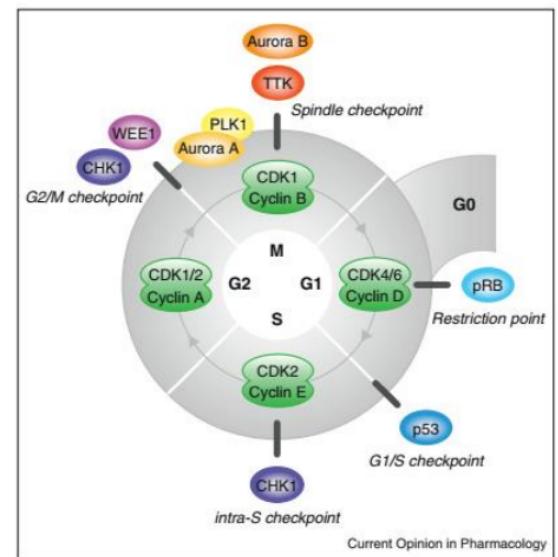
- they must be very accurate
- they are **ON/OFF triggers**, with no in-between: they manipulate events in a complete, irreversible fashion
- they need to be **reliable**: because of this, there are back up mechanisms
- they must be highly **adaptable**: they need to be general in order to act as extracellular signals, but they also must be cell specific.

In the picture we can see how each of the four phases are controlled by different proteins:

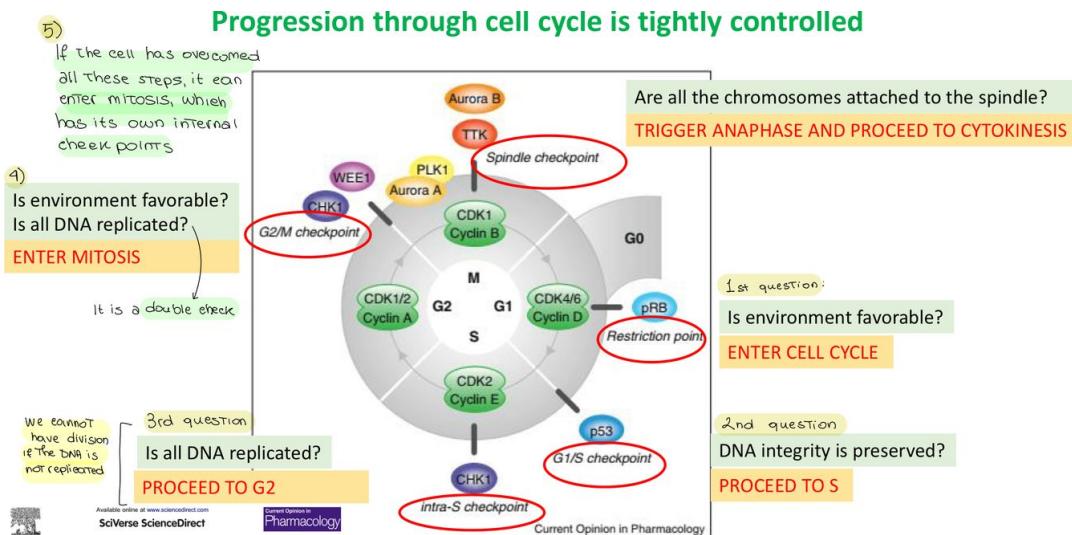
- green proteins are directly controlling the cell cycle (**cyclins** and **cyclin-dependent kinases**)
- other proteins are able to control cyclin/CDK dimers: examples of these are WEE1, p53, pRB.

All the checkpoints we find through the cell cycle **answer different questions**.

- **CDK4/6 + Cyclin D checkpoint** asks:
 - Is the environment favorable? If not, the cell is blocked and does not pass the **restriction point**. This is the most important checkpoint as, once the cell enters the S phase, it is committed to the cell cycle. (**pRB**)
 - Is DNA integrity preserved? If not, the cell undergoes apoptosis. (**p53**)
- **CDK2 + Cyclin E checkpoint** asks:
 - Is all DNA replicated correctly? (**CHK1**)
 - If yes, the cell goes from S to G2
- **CDK1/2 + Cyclin A checkpoint** asks:
 - Is the environment favorable?
 - Is all DNA replicated? (**CHK1**)
 - If yes, the cell goes from G2 to M



Lastly there is, during M phase, a checkpoint to see if microtubules are attached.



❖ CDK/CYCLIN DIMER

The cell cycle is controlled by **obligate dimers** formed by cyclins and CDKs.

The dimer is hence formed by two subunits:

- **cyclin: regulatory subunit**
- **CDK: catalytic subunit.**¹

The activity of the dimers rises and falls during different phases of the cell cycle: cdk levels are constant, **cyclin levels change during the cycle** depending on what needs to be transcribed.

There are different types of **CDKs**.

- **not all of them** are actually involved in the **cell cycle**. Only CDK1, CDK2, CKD4, CKD6 are. Other Cdk's are either involved indirectly, or belong to other pathways.
- they can have different kinds of structures although the **kinase core is always the same**
- they have **different target sequences**: S/T-P-X-K/R or S/T-P-X: as we can see there is always a serine and a threonine in the target sequence, and also a proline (proline directed kinases)
- in yeasts there are fewer CDKs, compared to mammals in which there is a **family of 20 members**.
- Cyclins control CDKs activation and substrate specificity
- **Different combos** of CDK-Cyclin complexes controls **different phases** of cell cycle
- The CDK/Cyclin complexes are the effectors of multiple upstream signals and control several downstream targets

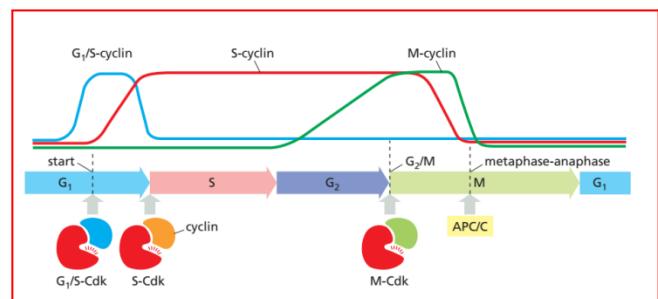
We can divide Cdk's on the basis of how many types of cyclins they can bind:

- **MULTIPLE CYCLIN CDK**: they present **only** the kinase domain
- **SINGLE CYCLIN CDK**: present **different, supplemental domains** with different functions, including also transcriptional regulation.

Mutations in every Cdk have been involved in cancer.

We can see here a picture which represents the cell cycle: as we can see the levels of cyclins go up and down while the levels of Cdk's stay the same.

- **G1/S-Cyc (CycE)** activates **CDK2** in late G1 → progression through START → commitment to cell-cycle
- **S-Cyc (CycA)** binds **CDK2** and causes chromosome duplication. Its levels remain elevated until mitosis and controls some early mitotic events
- **M-Cyc (Cyc B)** activates **CDK1** that stimulates entry into mitosis at the G2/M transition. Its levels fall in mid-mitosis
- **G1-Cyc (CycD)** with **CDK4/6** regulates G1/S-cyclins activity.



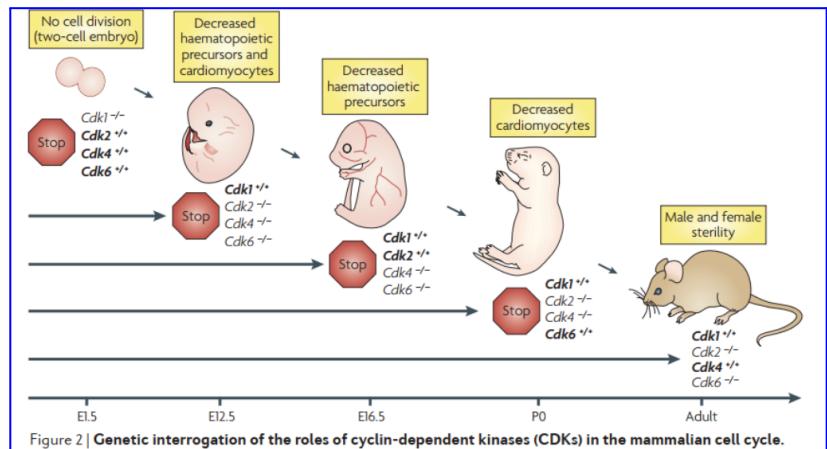
¹ Cdk's are protein **kinases**, which means they are able to phosphorylate the OH group of a protein target in specific residues, such as threonine, serine, tyrosine. The opposite of a kinase is a **phosphatase**, that instead removes the phosphate group from a protein.

The signal coming from all the molecules involved in the cell cycle converts in the transcription of certain genes which have determinate effects (some of them even transcribe for cyclins).

Recently a new model has been proposed, which has been obtained starting from the study of some mutants. By doing specific knockouts of Cdk1, or 2 or 4 or 6 or in combination, researchers have found that actually **the only CDK that is absolutely necessary for the division is CDK1**.

Without it, there is no division at all when the egg cell is fecundated, and the cell dies from the start. Instead, if we keep Cdk1, there is an initial division, although the models are often not viable because of hematopoiesis.

- Cdk1 is necessary for division
- Cdk2, Cdk6 and Cdk4 are not necessary for division in every cell type, just in some



❖ CDKs STRUCTURE

Cdk2 is a prototype for kinase structure: most Cdks are similar (not identical) to it because of the conserved elements. First of all, Cdk2 is divided into two lobes:

★ N TERMINAL LOBE:

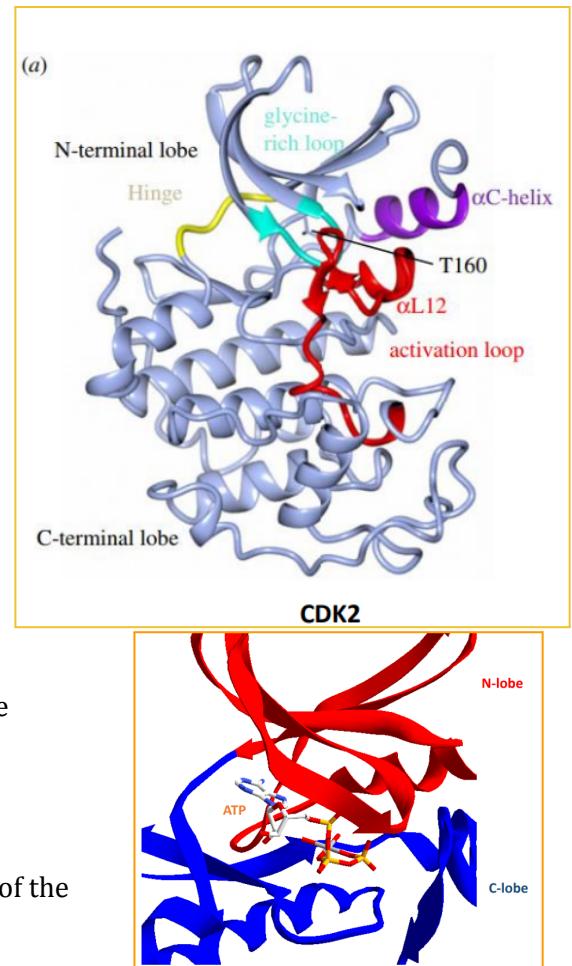
- made up of straight antiparallel *beta sheets*.
- **glycine rich inhibitory group** (g-loop) is necessary for the regulation of the kinase.
- we have a **C-helix**
- there is also a **flexible helix**, important for structural reasons and for regulatory purposes since it can bind to inhibitors.

★ C TERMINAL LOBE:

- only made up of *alpha helices*.
- it has an activation segment (or **t-loop**) which is essential for the function of every kinase.
- the most important amino acids in the loop are 145-172. In particular, **threonine 160** is very important because it can be phosphorylated. Its phosphorylation causes the complete activation of the kinase.

C and N are connected through the **flexible loop**.

Being a kinase, CDK needs ATP. **ATP** is placed at the interface of the two lobes. Hence, in this area is present the active site.



Different Cdks:

- are similar in primary sequence and folding but they are not identical;
- have different substrate preferences
- have different modes of regulation.
- difference is evident in the g loop and the t loop.

❖ CYCLIN STRUCTURE

Cyclins (in green, Cdk2 is in light blue) belong to a **family of 29 different proteins** of different molecular weight, hence having different sizes.

These proteins **are defined by their cyclin box fold:** a series of **100 amino acids** packed in **three alpha helices** with two others packed against this bundle.

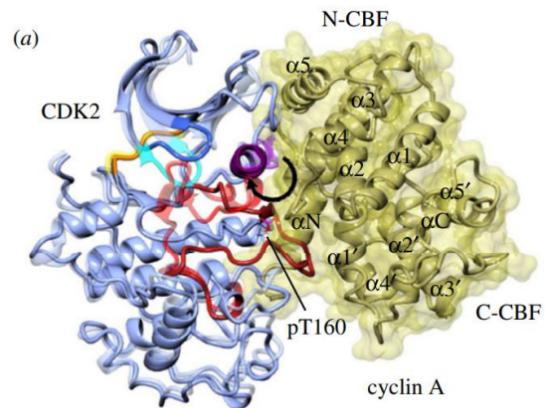
- there can be more than one cyclin box folds.

Most cyclins, for example **cyclin a**, has **two cyclin box folds**, a n terminal and a c terminal one.

- their **structure is almost identical**
- however the **primary sequence identity is very low.**

The **N terminal is mandatory**, while the **C terminal isn't**. Moreover, the position of the C terminal can be different depending on the type of Cdk.

- The C terminal of the cyclin isn't really in touch with the Cdk, whereas the **N terminal has a large interface area with Cdks.**



❖ DIMERIZATION

When a Cdk binds to a cyclin, a **series of conformational changes occurs**. Because of this, the regulatory elements of Cdk2 acquire different positions.

- the N terminal of the cyclin, assumes another position with respect to the C term, changing completely and twisting itself.
- this changes the position of the regulatory elements of the Cdk: in particular, the **g-loop and the helix rotate towards the interface** (the one between the N and C lobes) that causes the other subunits to move. Also the **t-loop changes position**.

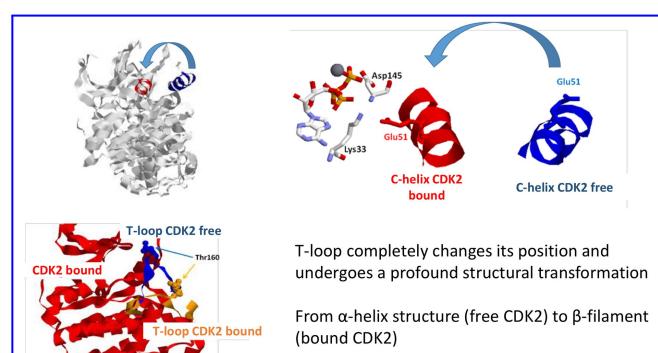
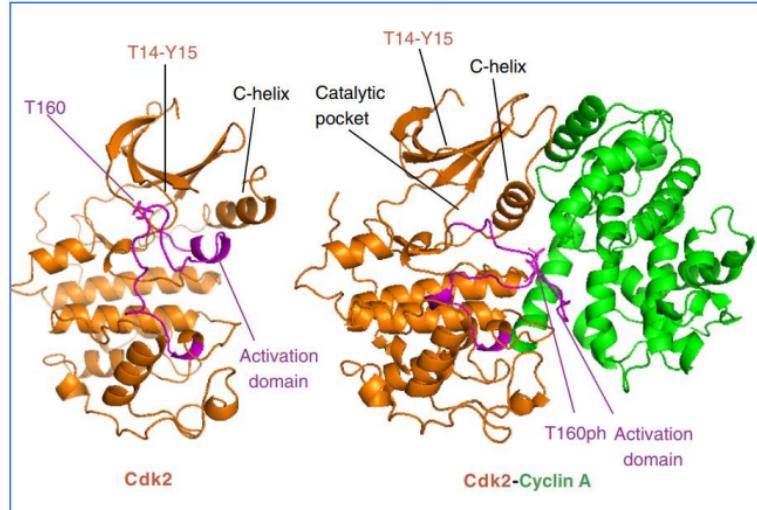
INACTIVE CDK (left)

- Cdk is inactive when it is not bound to cyclin
- C-helix is *outwards* far from the catalytic domain
- **Catalytic site is occupied by t-loop**, with Threonine160 right in the site
- Activation segment (t-loop) is disordered
- The position of the g loop is different.

ACTIVE CDK (right)

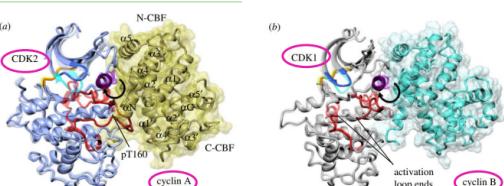
- **C-helix** packs against a specific hydrophobic helix of N-CBF and it **rotates towards the catalytic pocket**. Without the rotation of the C helix the activation doesn't work.
- the **catalytic site is now accessible** because the t-loop is gone
- the **kinase** has enough space to place in the catalytic pocket the substrate for **phosphorylation** thanks to **T160**

In the bound form, thanks to the movement of the C helix, **Glutamate 51** forms a catalytically triad (with Asp145 and Lys33) needed for phosphorylation.



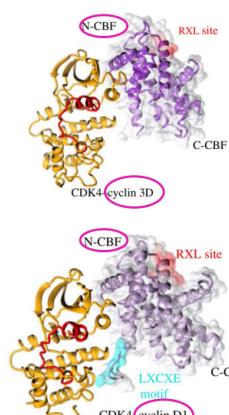
Structural insights into the functional diversity of the CDK–cyclin family

Daniel J. Wood and Jane A. Endicott



- CDK1 activation similar but with less contacts with CycB
- T-loop is more flexible
- CDK4 binds only N-CBF
- Small changes in CDK4
- Solvent filled cleft → substrate-assisted catalysis

CDKs: cyclin-dependent activation



All conformational changes:

- **increase affinity for substrate** (reduce K_M)
- **enhance the catalytic rate** of the reaction (increase k_{cat})

There are **differences in activation between different dimers**.

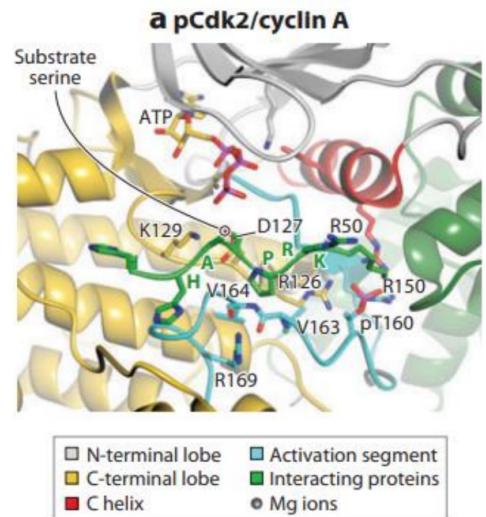
CELL CYCLE

There are several mechanisms that allow the cell to control dimerization between cyclins and Cdks.

- the closest ring of regulation is given by Cdk 1,2,4,6 (and the cyclin they use as dimer) which are directly involved in the activity of the cell cycle
- the second ring of regulation contains several kinases and other Cdks (for example no.7)

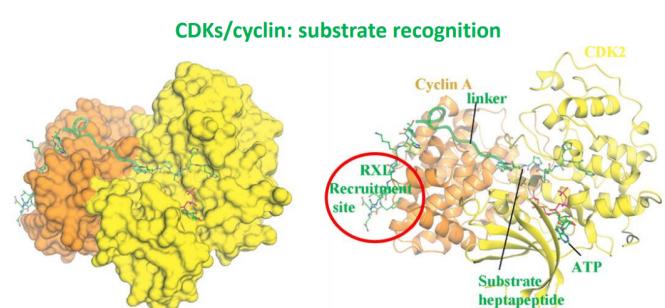
Cdks use as **target sequences**: S/T/P:

- the **threonine/serine** is the one that will be phosphorylated
- the **proline (in position +1)** is also important as it contributes to the overall charge in the C lobe: this will make the C lobe positively charged while the rest of the protein will be only lightly charged.
- in the sequence we also have **positively charged amino acids (in position +3)**: it can be a lysine or an arginine. The positive charge is important because it pairs, electrostatically speaking, with **T160**.
 - ◆ When the dimer is active, the Threonine is **phosphorylated**, bringing a negative charge that is recognised by lysine or arginine.

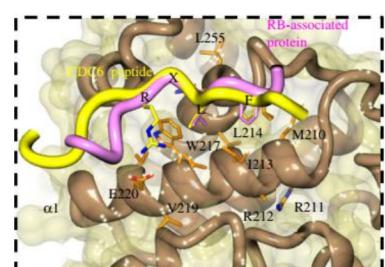


Another important region for substrate recognition is placed in another part of the dimer (Cdk+Cyclin), far away from the active site.

- at the level of the **active site** we have the target sequence, which performs the basic, necessary, recognition.
- the **recruitment site (RXL)** instead is on the cyclin and is also in contact with the target peptide, increasing the specificity of the kinase. Moreover, in this site some inhibitors are bound: hence the site is important not only for catalysis, but also for regulation.



The recruitment site in the **cyclin** favors substrate recognition and confers more flexibility in the substrate target sequence: in particular it **allows binding even when the positively charged amino acid in position +3 is absent**.



In order to activate the dimer:

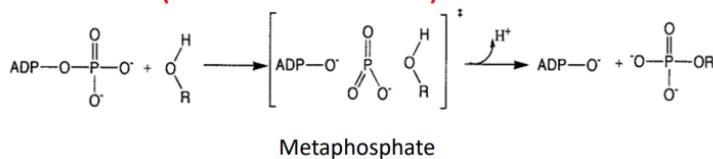
1. the movement of the **c-helix** must be rotated towards the catalytic site
 2. displacement of the t loop from an occlusion of the catalytic site to the site being available: in this case the T is outside and available for phosphorylation.
- T-loop completely changes its position: from an alpha-helix to a beta-filament when bound.

Now let's take a better look at the **CATALYSIS OF PHOSPHORYLATION**: at the beginning we have a non phosphorylated protein with a molecule of ATP accompanied by magnesium. In presence of an active kinase we have the phosphorylated protein and the release of an hydrogen.



There are two mechanisms explaining what happens:

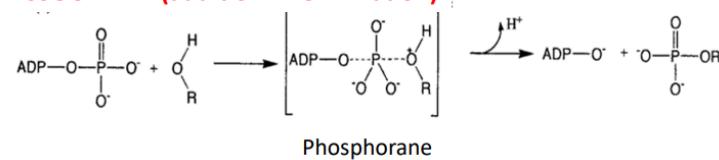
DISSOCIATIVE (elimination → addition)



DISSOCIATIVE MECHANISM: it has a

metaphosphate intermediate. No bonds to the incoming and outgoing groups when the protein encounters ATP: **first the P leaves the group, then the binding happens.** Charge accumulated on phosphoryl oxygen.

ASSOCIATIVE (addition → elimination)



ASSOCIATIVE MECHANISM: it has a

phosphorane intermediate. The incoming group has a substantial **bonding formation before the leaving -P group has left.** The charge is accumulated on the peripheral oxygens of the departing γ-P

Which of the two mechanisms is more probable? We still do not know.

1. scientist proposed that the most probable process of reaction is the **dissociative one**, as the formation of an intermediate pentaphosphate is more probable, on the base of the phosphate group and dislocation of the charges
2. recently instead, due to continuous refinement of the structures and improved calculations, the most probable theory is considered to be the **associative**.
→ so the scheme **is addition and then elimination.**

❖ CONTROL OF CDK/CYCLIN DIMER

- ★ The **FORMATION OF THE DIMER** is necessary so it can be considered as the first actual control.

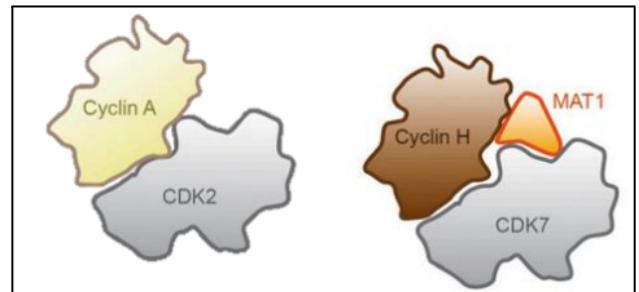
But of course there are other ways, as we need an integration of signals coming from outside and cellular response.

- ★ **CYCLIN BINDING:** there must be a production, at the transcriptional level, of the appropriate cyclins for each of the Cdks, in order for them to work. At different points of the cell cycle, we have the transcription of different cyclins.
- ★ **INHIBITORY PHOSPHORYLATION:** it inhibits either Cdks alone or the already formed dimer.
- ★ **P21/P27/P57 BINDING**
- ★ **ACTIVATING PHOSPHORYLATION**

ACTIVATING PHOSPHORYLATION

We know that **threonine160** (of CDK1-2) **must be phosphorylated**, to be able to recognise the positive charges of the target peptides and **allow dimerization**.

- the kinase that leads to T160 phosphorylation is **CDK7**, which is part of a bigger complex.
 - ◆ The structure of cCDK7 is very similar to the other Cdks: in the t-loop we have T170, which corresponds to T160 of Cdk2. The superimposition between Cdk2 and Cdk7 is very good, meaning that they are indeed very similar. Of course there are some small differences, which mainly regard the t loop:
 - T170 is always phosphorylated, even when Cdk7 is not bound to cyclins
 - the t loop is placed differently
 - the phosphorylation of another residue, S164, is necessary for full activation.
 - ◆ Its full complex is a trimer: near Cdk7 we also have **CycH** and **MAT1**.
 - MAT1 extends the interaction interface CDK7/CycH and works as an assembly factor

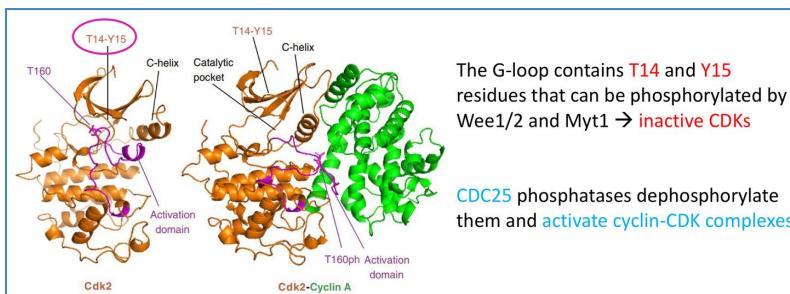


... to summarize		<ul style="list-style-type: none"> • In free CDK2, the T-loop is close to the active site and T160 is placed inward • Under these conditions Thr160 cannot be phosphorylated • Cyclin A binding induces conformational changes in CDK2 and the T-loop is displaced from the active site → complex catalytically competent • T160 is exposed and can be phosphorylated by CAK → active enzyme stable + increased activity
------------------	--	--

INHIBITORY PHOSPHORYLATION

Inhibitory phosphorylation happens at the level of **threonine14** (T14) or **tyrosine15** (Y15), which are contained by the ***g-loop***.

When these residues are **phosphorylated** by either WEE1 or MYT1, they lead to the **inactivation** of Cdks.



→ this is a *very fast* way of controlling dimer formation and it is also *reversible*. It is much faster than, say, controlling synthesis, degradation and transcription of each protein.

→ a **phosphatase CDC25** can technically dephosphorylate the residues and **activate the dimers**.

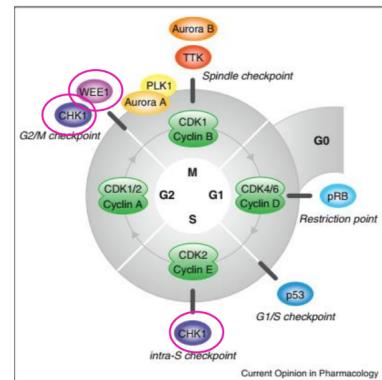
Wee1, Wee2, Myt1 all share the same structure and share the same mechanism for the inactivation of cyclin/Cdk complexes.

- Cdk2/CyclinA
- Cdk1/CyclinB

However, **if replication has not properly finished, CHK1 phosphorylates CDC25 phosphatases**, which, being phosphorylated, cannot free Cdks from inhibitory phosphorylation → inactivation.

In particular:

- CDC25A is phosphorylated for CDK2/CyclinA inactivation.
- CDC25C is phosphorylated for CDK1/CyclinB inactivation



What we are saying also corresponds to the inner/outer circle of protein regulation we have seen in the last lecture, especially in regards to the belonging of molecules like WEE1 to the outer circle.

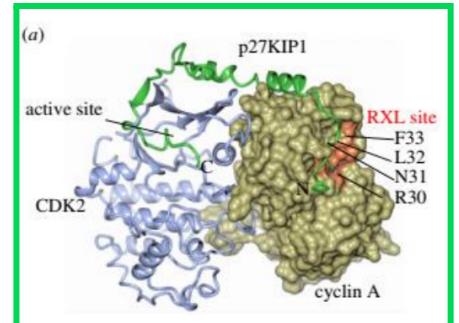
CKIs (CDK INHIBITORS)

CKIs are able to inhibit the activity of CDK/cyclin complexes by interfering with the interaction of the two subunits.

- **Cip/Kip** (p21, p27 and p57) inhibit **complexes** that have already been formed: CDK2/CycE, CDK2/CycA and CDK1/CycB, regulating all phases of the cell cycle

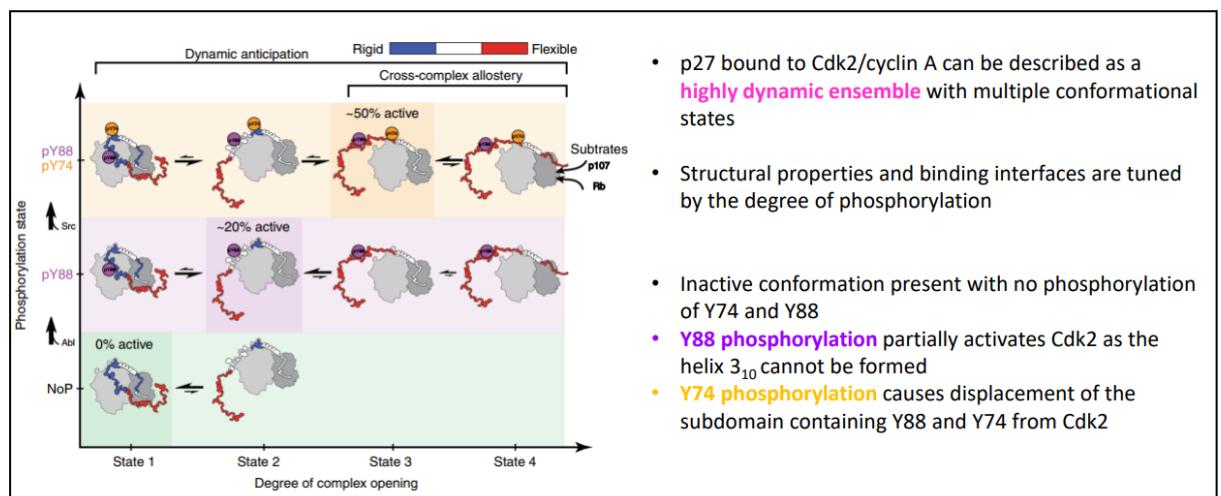
As we can see **these proteins bind both monomers**. They are **intrinsically disordered**, which means in solution they don't have a precise folding., but they do have it when interacting with the target.

→ their helix is present in the catalytic site, which means the target sequence can't be at the interface between C and N lobe. The **RXL motif** present in the helix is able to recognise the site, having a regulatory function.



- **p21** transcriptional target of p53, mediates cell-cycle arrest in G1 and G2 upon genotoxic stress
- **p27** responds to proliferation signals
- **p57** regulates cell cycle during embryonic development

→ the 3_{10} helix formation is regulated by phosphorylation of threonine 88: it cannot be formed if there is phosphorylated

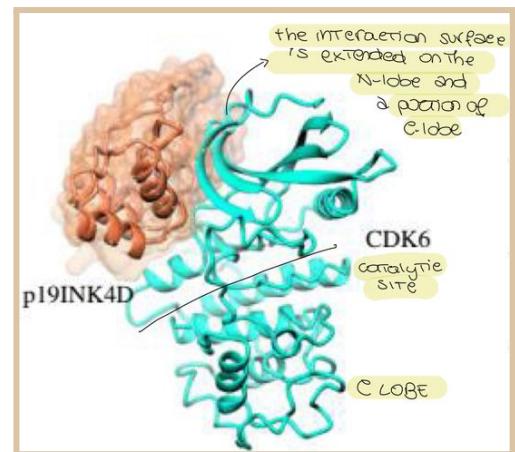


- **INK4** (p16, p15, p18 and p19) bind Cdks to **prevent dimer formation**: CDK4 and CDK6 and prevent activation by cyclin D.

They inhibit CDK4/6 and participate in the regulation of the transition G1 → S, which is the passing of the restriction point.

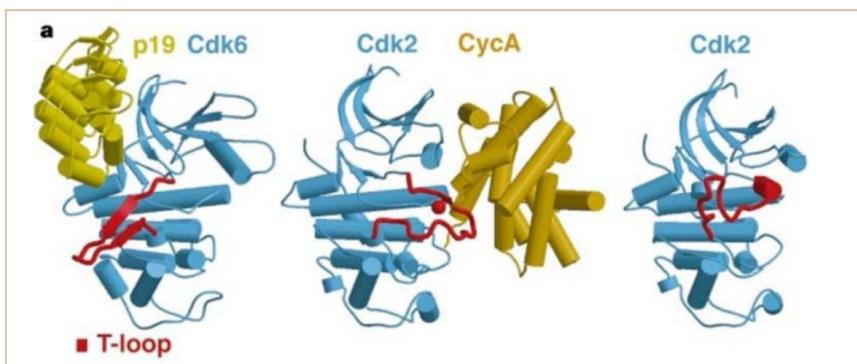
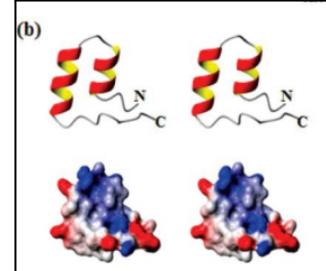
As we can see, INK4 proteins bind far away from the catalytic site.

- by binding, a **structural modification** is initiated (we say this is an **ALLOSTERIC MECHANISM**)
- because of this, the N lobe is displaced: 15° distortion of the N-lobe vs the C-lobe
- the catalytic site is modified, which causes a misalignment of catalytic residues



Structurally speaking, these proteins are characterized by the presence of four or more **ANKYRIN REPEATS**.

- tertiary structures of 30-34 aa that are found repeated in different proteins with multiple functions, for example **protein/protein interactions**
- very **conserved** motifs
- two beta turns, two alpha helices in the middle of which we have another beta turn
- strings of AR are packed:
 - ◆ few of them make up a linear structure
 - ◆ increasing the number of repeats, we have a bending
- AR do not recognize specific sequences



All these mechanisms are studied to prevent and cure cancers. Also to know structures precisely is important in order to build specific drugs.

- starting from the structures we have studied, several different molecules were proposed to be used as inhibitors to prevent cancer.

SIGNAL TRANSDUCTION

In a cell, a signal can be transduced in many different ways. Signaling **transduction** is present in both prokaryotes and eukaryotes, being very **conserved**.

- for example, the regulation of the cell cycle is correlated to both the inside and the outside of the cell.

Although the **mechanisms for signaling are diverse**, there are some things that are always in **common**:

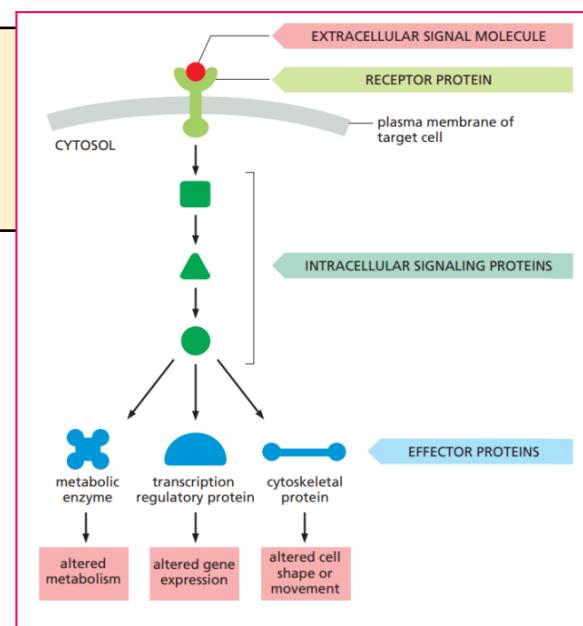
1. we need a molecule to start the signal, which must be sensed
2. the sensing of the molecules must be communicated to the cell
3. in the end, the change of transcription is achieved

❖ CELL SIGNALING

We define cell signaling as the **cascade of reactions in a cell** when the cell is receiving an **extracellular molecular signal** from a signaling cell through a **protein receptor** (of the PM or intracellular) and acts, once the signal arrives to **effector proteins**, to activate or deactivate processes.

- In **bacteria** the signaling is necessary for **sensitivity or mobility**: the signals tell the organism where to move
- In **yeast** it is essential for **mating**
- In **eukaryotes** signaling regulates every aspect of the cell: **communication between cells in multicellular organisms are mediated by extracellular signaling molecules.**

The singling can be of a small cellular distance (**local**), but it also can involve large distances (**at an organism level**). An example of the latter can be hormone regulation.



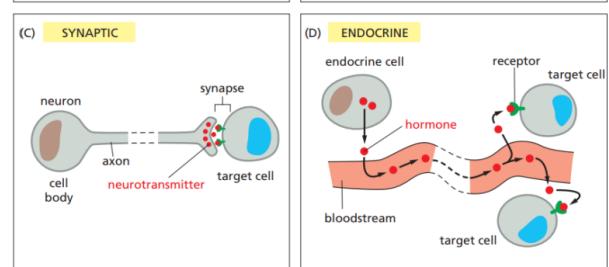
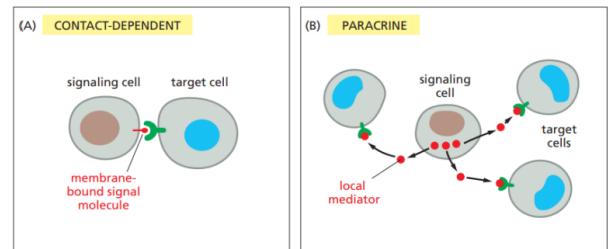
Signaling processes can be **subdivided into four different types**.

1. **CONTACT DEPENDENT:** it's the only signaling process that involves direct contact.

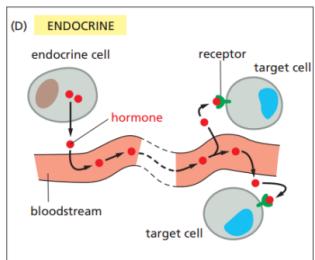
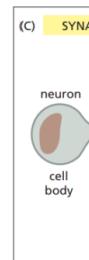
In this case we have two cells: the signaling cell and the target one. The signaling cell exposes on its membrane a protein which the target cell comes in contact with.

→ Examples are **immune system** cells.

2. **PARACRINE SIGNALING:** a signaling cell releases a molecule (local mediator) which the target cell picks up thanks to appropriate receptors. It's **short range signaling**, in which the two cells must be really close. If the cell produces a mediator for it to be picked up by itself, we define the process as **AUTOCRINE**.



3. **SYNAPTIC SIGNALING**. In this case we have **long range distance signaling**. Generally, the local mediator in this case is a **neurotransmitter** which is very far away from the target cell. The neurotransmitter is sent through the axon. At the end of the axon, at the synapse, the neurotransmitter reaches target cell



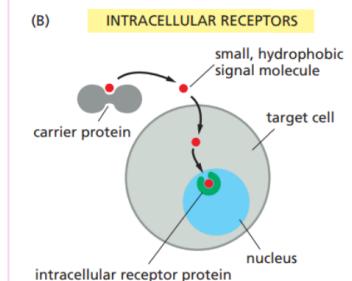
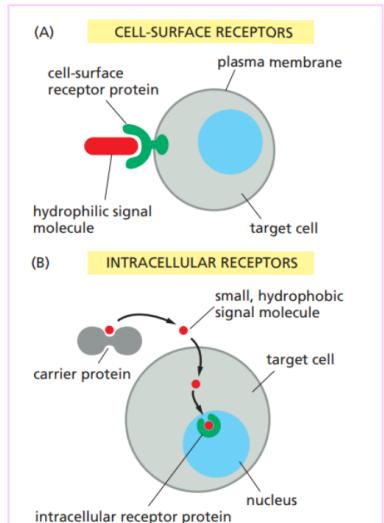
4. **ENDOCRINE SIGNALING**. Cells secrete their signal molecules (**hormones**) into the bloodstream.

There can be **different types of signaling molecules**.

They can be single amino acids, nucleic acids, small peptides, fatty acids or derivatives, steroids or proteins.

- in contact dependent cases, receptors are placed on the **plasma membranes** of the signaling and target cell. In this case the signaling protein is hydrophilic.
- in most cases, signaling molecules are released into the **extracellular space** and then reach the target. Usually hydrophobic because it must pass across the membrane.

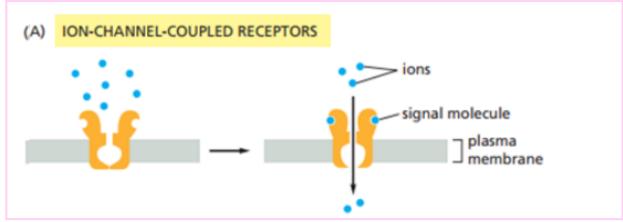
When the signaling molecule comes in touch with the target molecule, the signaling cascade begins.



- **Different combinations** of molecules lead to **different effects**.
 - For example, different combinations of molecules can lead to growth, differentiation, or cell death.
- The **same molecule** can induce different effects on **different cell types**.
 - An example of this is **acetylcholine**, which can cause, on different tissues, secretion of salivary glands or the contraction of skeletal muscle.

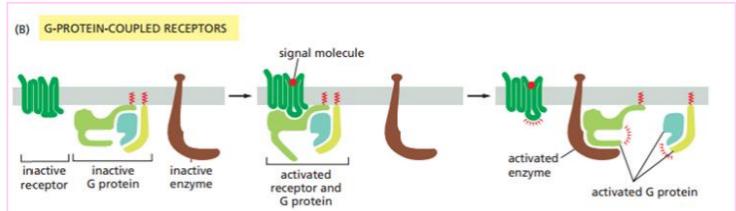
There are **four main types of receptors.**

- ★ **CHANNEL RECEPTOR:** formed by ion channels that can be closed or open depending on when they bind the signaling molecules (*like skeletal muscle*)



★ G PROTEIN COUPLED

RECEPTORS (GPCR): these receptors are made of seven transmembrane helices. The interaction between the receptor and the target protein is mediated by a third protein: heterotrimeric G protein.



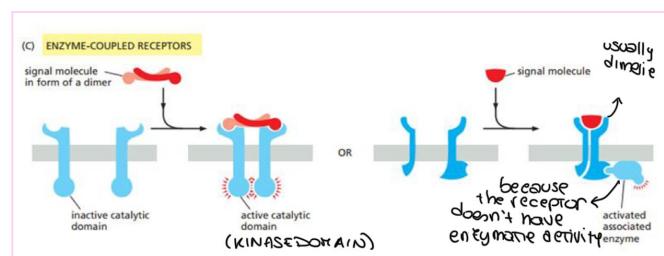
When the receptor is unbound, it's inactive. Instead when the signaling protein arrives, the receptor has some conformational changes that recruit the heterotrimeric G protein. Together, the receptor and the G protein activate the target enzyme. This type of signaling transduction requires what we call **second messengers**.

★ ENZYME COUPLED RECEPTORS.

These type of receptors can be subdivided into two categories:

- receptors with intrinsic enzymatic activity (RTK)
- receptors that directly interact with enzymes that are activated upon stimuli

(an example of these are *growth factors*)

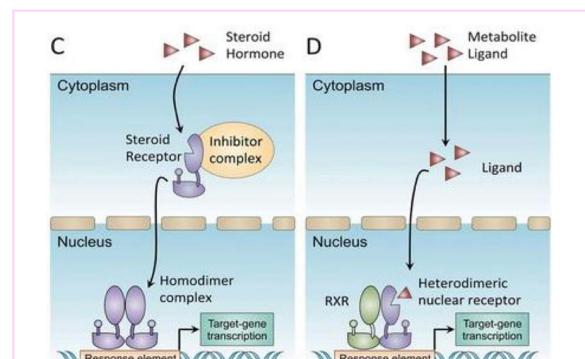


★ NUCLEAR RECEPTORS.

They are receptors that work as transcription factors and are intracellular receptors going from cytosol to nucleolus. When activated they dimerize in the nucleolus and start acting as transcription factors. The gene that is transcribed depends on the sequence. Some sequences are recognized by only specific receptors.

They recognize lipophilic ligands able to cross the PM and that migrate into the nucleus where they act as transcription factors

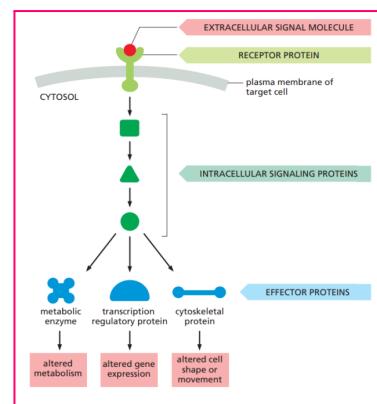
An example is estrogen.



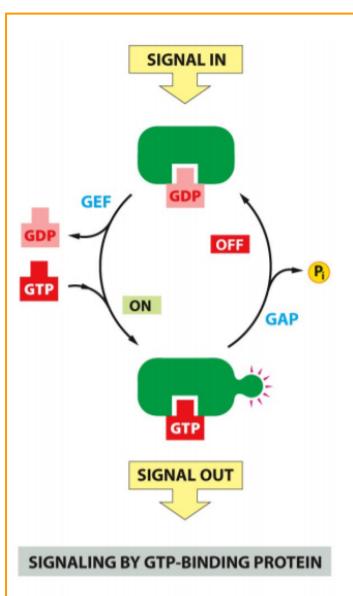
❖ INTRACELLULAR SIGNALING PROTEINS

Up until now we have seen what type of extracellular signaling methods exist and what types of receptors there are.

What we want to see now is the **intracellular signaling proteins**, which are responsible for the transduction of the signal from the receptor to the effector protein.



There are two main ways in which intracellular signaling proteins can work:



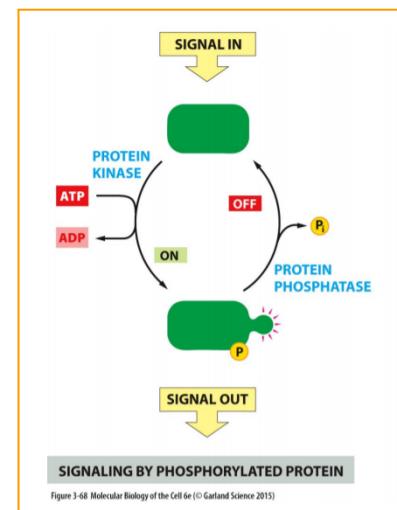
1. the intracellular signaling proteins are placed in **communication** one with the other, thanks to small molecules which are soluble and act as **second messengers**. They can be gasses, or hydrophobic, or hydrophilic molecules which are able to diffuse through the cell.

An example of usage of second messengers is **GTP**. GTP is a highly energetic molecule which is mostly used for signaling.

- The protein binding GDP is in its inactive conformation and cannot transduce a signal. When we have a **signal**, the signaling molecule induces the **exchange** of GDP into GTP (there is no phosphorylation!). This exchange can be helped by **GEF**, which in some cases is the receptor itself, and in others is a distinct protein.
- Then the molecule is in its active conformation.
- Later, when we need to **turn off** the signal, a **GAP** (*GTPase activating protein*, with intrinsic GTPase¹ activity) **hydrolyzes** the **GTP** bound to the signaling protein, which loses a phosphate group, creating GDP again.

2. another way of interaction for intracellular signaling proteins is **direct activation**, without the use of other proteins. A way in which this can happen is through **phosphorylation** (*kinases/phosphatases*), which is able to create a domino effect. All these proteins can be either on or off. Of course, the process involves the usage of kinases and phosphatases.

The return of the initial condition is necessary in both cases.



¹ **GTPases** (which can be either *small* or *heterotrimeric*) are able to hydrolyze GTP.

❖ SIGNAL SPECIFICITY

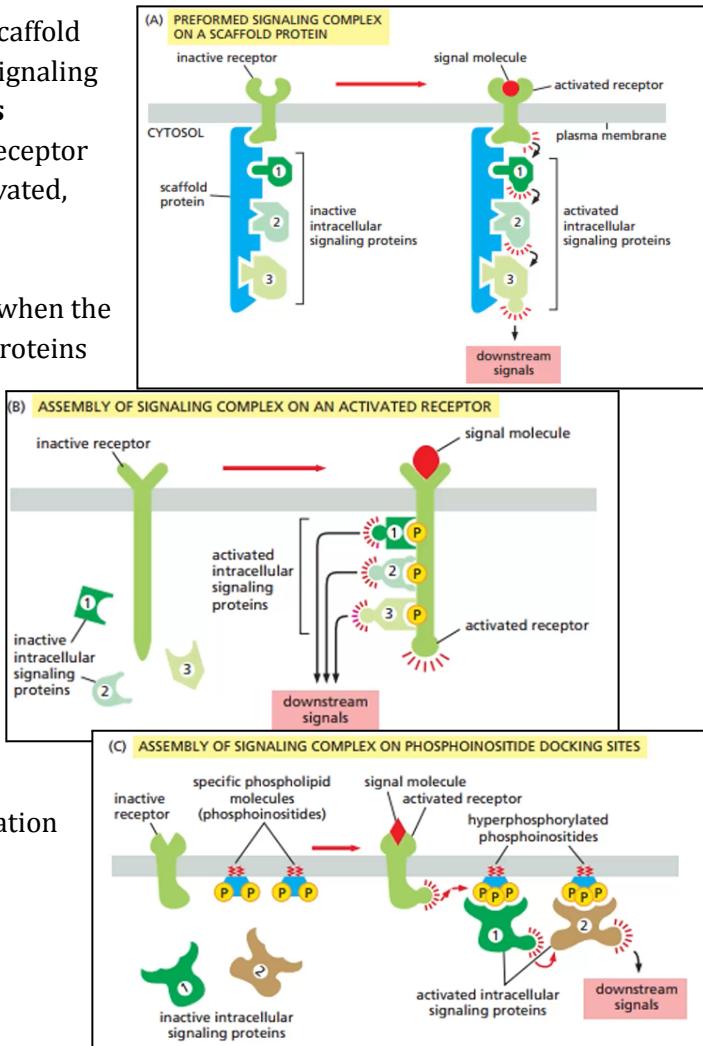
How can we improve a signal's specificity? Meaning, how can a signal be specific for one kinase, so it does not activate all the other kinases?

- **COMPARTMENTALIZATION:** allows localization of proteins interacting with each other in the cell inducing **formation of large complexes** that transduce the signal. These complexes ensure that the molecules interact only with each other and not with inappropriate partners. This is done through the help of:
 - **DOCKING SITES:** domain that binds specifically a target that is the modified (not part of the catalytic site)
 - **SCAFFOLD PROTEINS:** proteins that recruit and bring together groups of interacting signaling proteins into signaling complexes

EXAMPLE A: The inactive receptor contains the scaffold protein, which itself contains other intracellular signaling proteins. **Everything is placed correctly but it is inactive.** Once the signaling protein arrives, the receptor is activated and the intracellular proteins are activated, forming the downstream signal.

EXAMPLE B: Signaling complexes assemble only when the receptor is activated. The intracellular signaling proteins are present but they are not placed in contact with the receptor, hence the **machinery is not placed correctly before activation.** When the receptor is activated by the signaling molecule, certain modifications happen which lead to the recruitment of the intracellular signaling proteins.

EXAMPLE C: Once again the **recruitment of intracellular signaling proteins does not happen previous to activation.** Only when activation happens the proteins are recruited. However, recruitment also involves specific phospholipid molecules.

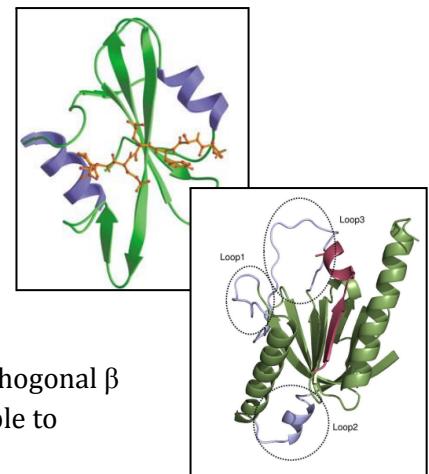


❖ SIGNAL RECOGNITION

Phosphorylation must be recognised by specific protein domains. These **domains must be able to tell when the threonine or the tyrosine are phosphorylated.**

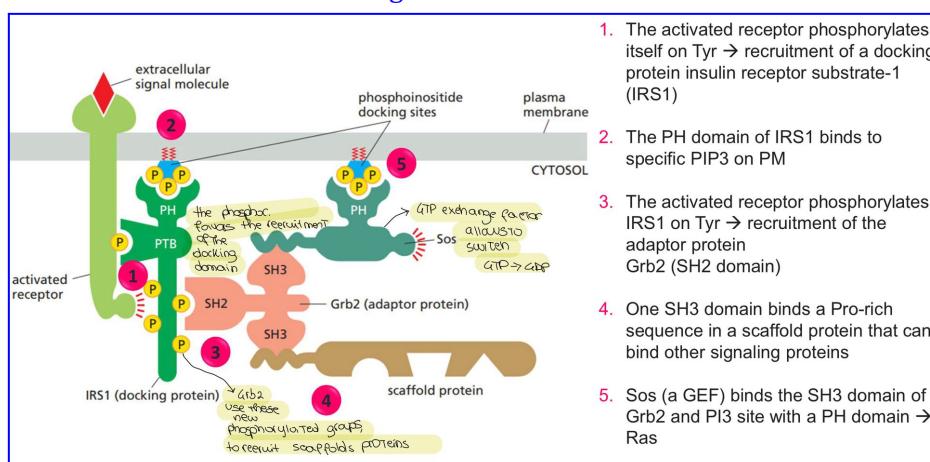
Examples of such domains are

- **Src homology 2 (SH2)**: it is made up of 100 aa folded into a large beta sheet flanked by two alpha-helices. In the middle of the domain recognition occurs.
- **Phosphotyrosine binding domains (PTB)**: 100-170 aa with a structural core made by 7 anti-parallel β -strands forming two orthogonal β sheets, capped at the C-terminus by α helix. This domain is also able to recognise phosphotyrosines.
- **Src homology 3 (SH3)**: this domain recognises proteins that are rich in proline. It mediates substrate recognition and regulation of kinase activity and it is also formed by 5-8 beta sheets.
- **Pleckstrin homology (PH)**: this domain, made up of one beta structure and one alpha helix, binds to the charged head groups of specific phosphoinositides produced in the PM in response to extracellular signals. It allows docking of the plasma membrane



These domains can **fold independently** from one another. We can have **different combinations** of domains.

INSULIN can be an example of extracellular signaling. When the receptor is activated it phosphorylates itself. The phosphorylation recruits docking protein IRS1, which is then phosphorylated by the receptor protein. Once IRS1 is phosphorylated, it triggers the recruitment of adaptor protein Grb2. In particular Grb2 recognises IRS1 phosphorylation thanks to its SH2 domain. Then, Grb2 uses its two SH3 domains to recruit SOS and a scaffold protein, which allows other proteins to interact thanks to its large area.



Timing is very important in signaling: how long is the signal maintained (persistence) and how fast is it transduced (response timing)?

- **PERSISTENCE:** duration of the response. Can range from less than seconds (*synaptic response*) to prolonged or permanent responses (*development, differentiation*)
- **RESPONSE TIMING:** time that occurs to induce a response. Varies in different signaling systems; from milliseconds (synaptic response) to hours or days (development)

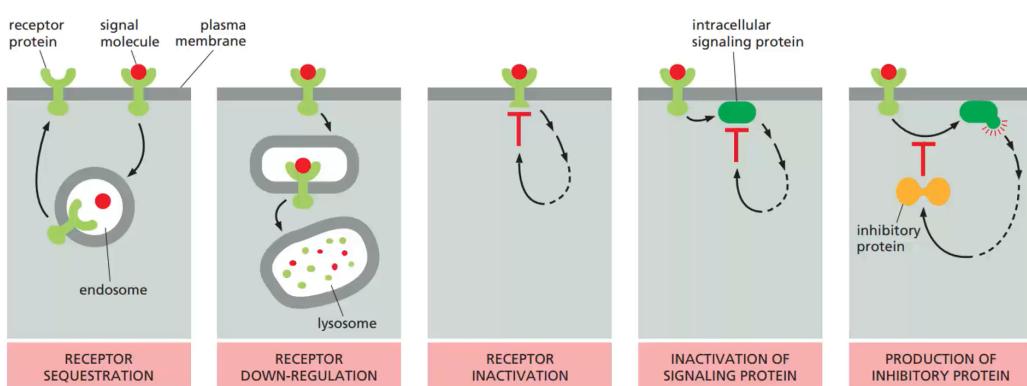
SENSITIVITY is also very important and it can be very variable. Some receptors have high sensitivity others have a low one. Low sensitivity doesn't mean the reception doesn't work, it means that, in this particular condition, it is needed.

1. If the receptor is **far away** there should be a **high sensitivity** (hormones)
2. If the receptor is really **near**, or if the signaling molecules are very common, there should be **low sensitivity**.

Sensitivity depends on:

- **affinity** between ligand and receptor, expressed by the constant of dissociation of the complex (K_d), usually around 10^{-10} M
- **cooperativity**, sometimes we need more than one molecule connected with the receptor. ligand-receptor interactions determine variation in the receptor activation state in response to a small change in the concentration of ligand.
- **amplification**: a small number of activated cell-surface receptors evoke a large intracellular response by producing large amounts of a second messenger or by activating many copies of a downstream signaling protein
- **dynamic range**: certain receptors can respond to wide ranges of signaling molecule concentrations, others cannot.

Sensitivity of a receptor can be modulated in many ways, for example by productions of inhibitors or by receptor sequestration.



Other features:

- **Processing:** we can convert a simple signal in a complex response
- **Integration:** different types of signaling cascades that have the same target in the end
- **Coordination:** coordination of multiple responses in one cell can be achieved by a single extracellular signal

G PROTEIN COUPLED RECEPTORS

❖ GPCR SIGNALING

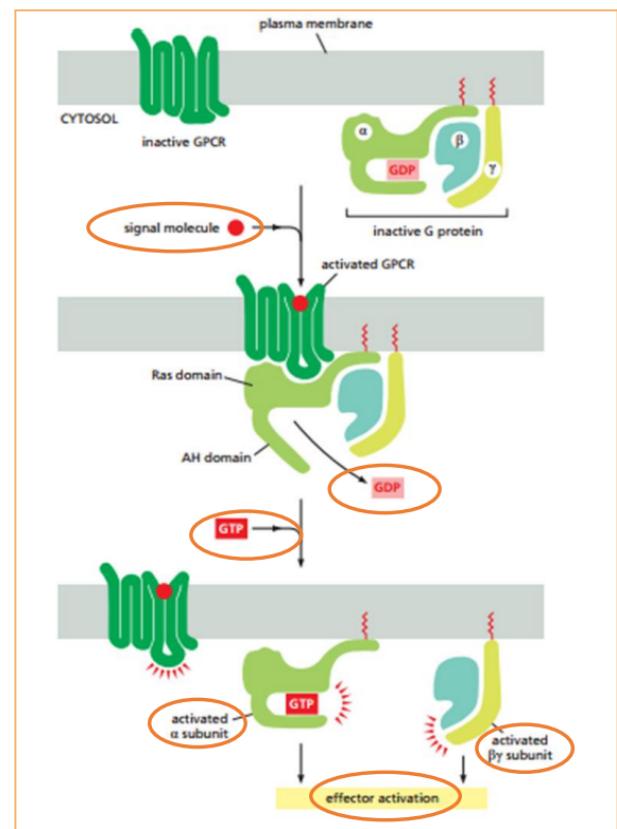
Previously we have seen that there are four different types of receptors:

- **ion channels**: can be closed or open depending on when they bind the signaling molecules
- **G protein coupled receptors**: the receptor uses an heterotrimeric G protein to activate the target enzyme
- **enzyme coupled receptors**: receptors that work with enzymes or have enzymatic abilities
- **nuclear receptors**: when they dimerise they start working as transcription factors themselves

In g protein coupled receptor, receptor proteins are coupled with another protein: **protein G**, which is an heterotrimeric protein, made up of three subunits.

1. when a **signal molecule** is recognised by the receptor, the latter undergoes conformational change favoring recruitment of the g protein
2. once the **G protein** has been recruited, the **GDP** bound to it is **exchanged with a GTP**, and the protein is activated
3. activation of **multiple downstream effectors**
4. when the **signal has to be switched off**, the alpha subunit **hydrolyzes GTP**, allowing to go back to the original conformation

G proteins are called like this because they can bind the guanine of GTP or GDP.



❖ GPC RECEPТОRS

G protein coupled receptors are the **widest family of cell surface receptors**, with more than 800 members. The responses that are mediated by these receptors:

- start from the most **different ligands**: they can be proteins, small peptides, amino acids, fatty acids, ions and photons.
- the **same signal molecule** can activate **different GPCRs**, allowing different kinds of responses. For example adrenaline can activate 9 distinct GPCRs (which are only structurally related)
- **different GPCRs for the same signal are expressed in different cells** and elicit different responses

All GPCRs have **similar structures**, even if they have **low similarity in sequence**.

- their 3d structure is always characterized by **7 transmembrane helices**, which form a single cylindrical structure, almost like a channel (however no molecules pass through)
- the **binding site** for the ligand is more **in depth**, in the membrane.
- there are still some **differences in the 3d structures**. Such variability is:
 - in the **extracellular loop 2**, which can have different conformations
 - in the **7 transmembrane helices**: although all GPCRs have them, the helices can assume structures that are not the same.
- **highest variability is in the ligand binding site**, as it ensures specificity for that particular signaling molecule. **Chemical properties** of the binding sites can also be different, for example we may have different *hydrophobicity*, and also *charge* can be different.

In the table we have the total number of GPCRs that have been found in humans and mice, each with their identified types of ligands.

The total number of GPCRs in the repertoires of mouse and human

Group	Number in mouse	Number in human	Peptide ligand	Biogenic amine ligand	Lipid ligand	Purin ligand	Other ligand	Orphan
<i>Glutamate</i>	79	22	0	10	0	0	4	65 (8)
<i>Rhodopsin (α)</i>	105	101	8	49 (41)	20	4	6 (8)	18 (20)
<i>Rhodopsin (β)</i>	46	43	36 (37)	0	0	0	0	10 (6)
<i>Rhodopsin (γ)</i>	67	64	49 (51)	0	4	0	2	12 (7)
<i>Rhodopsin (δ)</i>	82	63	43 (22)	0	11	10 (12)	2 (1)	16 (17)
<i>Adhesion</i>	31	33	0	0	0	0	1 (2)	30 (31)
<i>Frizzled</i>	11	11	0	0	0	0	11	0
<i>Taste type 2</i>	34	25	0	0	0	0	1 (4)	33 (23)
<i>Secretin</i>	15	15	15	0	0	0	0	0
<i>V1R</i>	165	3	—	—	—	—	—	—
<i>Olfactory</i>	1037	388	—	—	—	—	—	—
<i>Others</i>	25 ^a	23 ^a	—	—	—	—	—	—
Total	1697	791	151 (133)	59 (51)	35	14 (16)	27 (32)	184 (112)

The GPCRs are divided into families according to the GRAFS classification system (*Glutamate*, *Rhodopsin (α)*, *Rhodopsin (β)*, *Rhodopsin (γ)*, *Rhodopsin (δ)*, *Adhesion*, *Frizzled*, *Taste2*, *Secretin*). Additionally the most recent published numbers for pheromone receptors type 1 (*V1R*) [29,30] and *Olfactory* receptors [18,19] are given. A ligand preference for the different families is shown in columns 3–8. *Others* (all receptors are found in both species unless otherwise noted): DARC (Duffy) GPR137 (C11ORF4), GPR23, GPR88, GPR120, GPR135, GPR139, GPR141 (and Gpr141b only in mouse), GPR142, GPR146, GPR152, GPR160, GPR151 (GPCR-2037), HGPCR19, GPR149 (IEDA), GPR143 (OA1), GPR172A (PERVAR1) (found only in human), GPR172B (PERVAR2) (found only in human), TRHR (and Trhr2 only in mouse), GPR137B (TM7SF1), GPR137C (TM7SF1L2), TM7SF3, GPR175 (TPRA40).

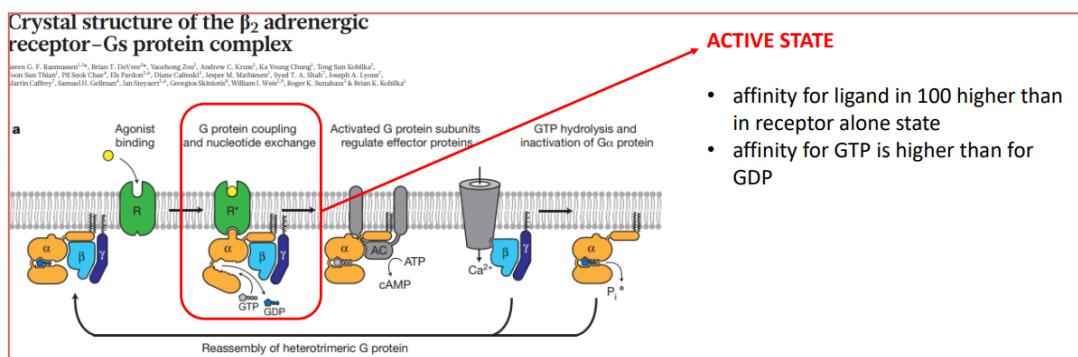
❖ GPC RECEPTEORS: EXAMPLES

The prototypic GPCR is **RHODOPSIN**, present at the level of the **retina** and activated by the **presence of light**. When light is present it will recruit the heterotrimeric G protein.

- Bundle of **7TM α -helices** connected by 6 loops of varying lengths. Of course there are other elements, such as other helices and beta sheets.

Another prototypical GPCR is **B2 ADRENERGIC RECEPTOR**, which has many more structures available and a more detailed mechanism.

- responds to adrenaline and noradrenaline (stress hormones) which are produced by **adrenal glands** upon certain stimuli, for example stress. The hormones are released in the bloodstream and are then recognised by receptors.
- the b2 adrenergic can be expressed in different cell types: depending on the cell type, different downstream signals can be activated.
- 2-adrenergic receptor (β 2AR) the first non rhodopsin GPCR cloned and one of the most extensively studied. It was obtained by adding soluble domains that favor crystallization, as this protein was originally quite hydrophobic.
- it has the classic GPCR structure: 7TM associated together to make a cylindrical structure spanned across the membrane
- they were also able to find **dimers of this protein**, so it is debatable if these receptors work as dimers or not. In the presence of cholesterol, dimerization usually occurs.

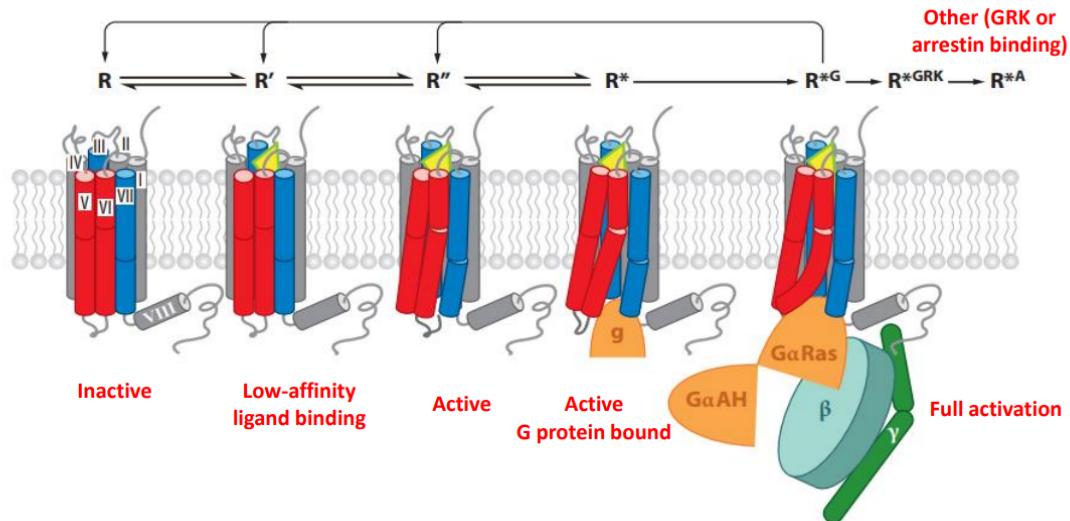


1. the **receptor** is bound by the **agonist**¹, which in this case is adrenaline
2. this promotes interaction between receptor and **heterotrimeric G protein** in its inactive conformation (which is binding GDP)
3. this leads to an **exchange** at the level of the alpha subunit of the heterotrimeric G protein: from GDP to GTP. → the receptor works as a GEF, making it so that there is a higher affinity for GTP with respect to GDP because of small conformational changes
4. when GTP binds, the **β and the γ subunits detach themselves** from the **α subunit**. This is because if α binds GTP, it cannot bind the other subunits.
5. then the α subunit activates **downstream effectors**, and so do β and the γ : however they activate different effectors, so that from one agonist there can be different responses.

The intrinsic GTPase activity of $G\alpha$ leads to hydrolysis of GTP to GDP and the re-association of $G\alpha$ -GDP and $G\beta\gamma$ subunits (**termination of signaling**)

¹ an agonist is a chemical that activates a receptor to produce a biological response.

ACTIVATION AT THE RECEPTOR LEVEL (R to R')



Until we don't have full interaction between the alpha subunit and the heterotrimeric G protein, the process can be reversed.

We have **different conformations for our 7PM receptor**:

1. R is the **inactive** (ground) state, which can be stabilized by binding of inverse agonists or antagonists
2. R' is the inactive **low-affinity agonist-bound state**: the agonist is already bound but there haven't been many conformational changes, apart from the ligand binding site
3. if the binding between the agonists is strong, we obtain an **activated state (R")**. R" differs from R state by only small local changes in the alpha helices
4. R* is the **activated G protein bound state** with a substantial rearrangement of helices and side-chain microswitches on the intracellular side that exposes the alpha subunit of G protein.
5. Finally we have the **fully activated state (R^{G*})**. The transition is accompanied by the release of GDP and proceeds unidirectionally. Moreover, with the exchange to GTP, subunits $\gamma\beta$ are discarded.

❖ HETERO trimeric G PROTEINS

Heterotrimeric G protein is made up of:

→ an **α subunit**

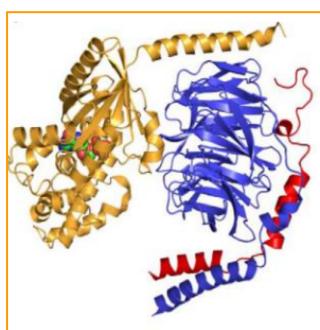
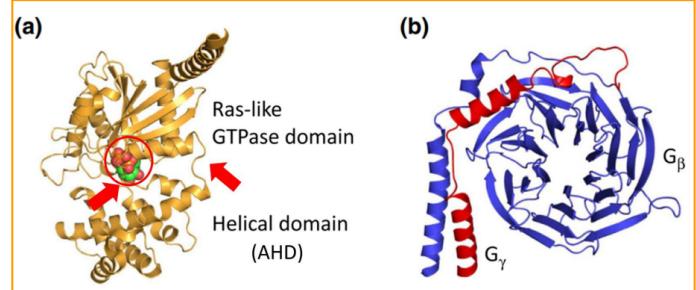
→ a **$\beta\gamma$ dimer**

which can be separated or interacting (although $\beta\gamma$ dimer always works together).

The **alpha subunit** is instead made up of two different domains:

- **helical** domain (AHD)
- **ras-like** GTPase domain

in the middle (red circle) we have GTP.



The N term of the α subunit and the C term of the γ subunit are very close to each other, although they do not interact. The two terminals are inserted in the plasma membrane thanks to lipid modifications.

- the α subunit undergoes palmitoylation or myristoylation
- the γ subunit can be bound to farnesyl or geranyl-geranyl

the localization of GPCRs in the membrane is ensured by the presence of lipid modifications.

There are **many types of GTPases** (=enzymes able to hydrolyze GTP), which have different names according to their functions.

Small GTPases (small G-proteins): soluble in the cytosol

- **Ras** (cell growth)
- **Rab** (vesicular transport)/**Ran** (nuclear transport)
- **Arf** (vesicular transport)
- **Rho** (cytoskeleton organization)

- GTPases are molecular switches regulating a plethora of cellular events
- GTP-binding is the active form of signaling
- GTP hydrolysis switches the signal off (ON/OFF signaling)

GTPases belonging to the RAS superfamily² can be divided into:

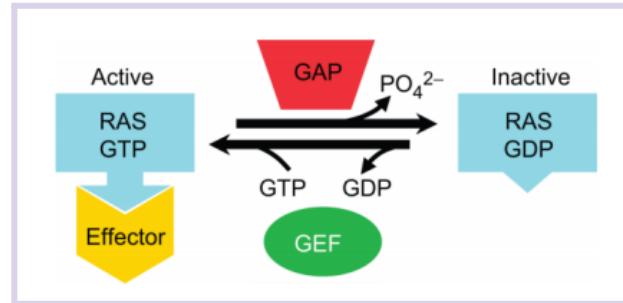
- **small GTPases**, which are cytosolic and only contain one catalytic (RAS) domain.
- alpha subunit of **heterotrimeric G protein**: it's not soluble, so it's not present in the cytosol, only in the plasma membrane. It doesn't contain only the catalytic (RAS) domain, also a helical domain.

² The Ras superfamily, derived from "Rat sarcoma virus", is a protein superfamily of small GTPases

HOW DO THE EXCHANGES AND THE HYDROLYSIS WORK?

In other words, how does the cell know it's time to exchange GDP with GTP, or hydrolyze GTP into GDP? Also for this, there are proteins that aid the process.

- **GEF**, which is an *GTP exchange factor*.
 - in the case of small GTPases, GEF is a small protein
 - in the case of heterotrimeric G proteins, GEF is represented by the GCP receptor itself.
- **GAP**, which is a *GTPase activating protein*



GTP half-life at 37°C

In water	60 days
Ras	25 min
Ras+GAP	36 ms (at 25°C)
Gα alone	few sec
Gα+RGS	30ms
RGS: regulators of G-protein signaling	

When taking a GTP molecule in solution, unlike ATP, which is unstable, it is very stable. It has a half life of 60 days. However if we add a small Ras protein, its half time decreases to 25 minutes. This means that, **without a doubt, Ras proteins have a GTPase activity.**

However 25 minutes for hydrolysis is still way too long, which means something isn't exactly quite right and this is not exactly an enzymatic mechanism.

If we also place a GAP in solution, the half life becomes 36 milliseconds. This means that **to fully activate the GTPase activity of RAS we always need a GAP.**

This is not only valid for small proteins, but also for heterotrimeric G proteins: the alpha subunit (which contains RAS domain) itself takes a few seconds to activate the protein. Instead, with RGS (similar to GAP), it only takes 30 milliseconds.

→ **RGS is needed for hydrolysis of the alpha subunit. However the alpha subunit itself is much more efficient than the RAS domain alone.**

There is not only one type of alpha subunit, rather, **we have a family of alpha subunits** which share some similarities.

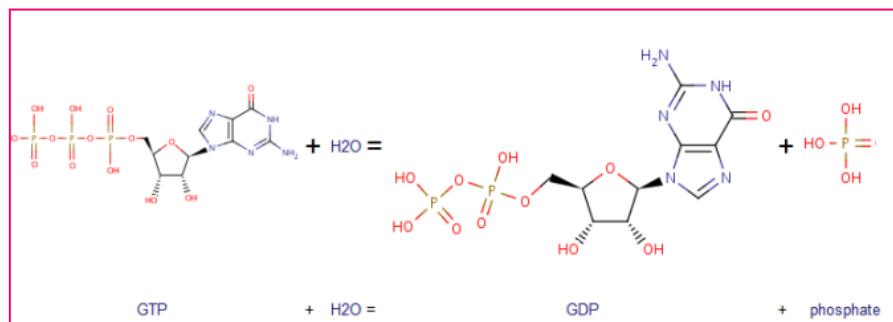
In humans we have 18 different members of the alpha subunit family. They can be subcategorized in different groups.

Gα_s → activation of adenylyl cyclases (AC)
Gα_i → inhibition of certain AC
Gα_{11/12} → PLC-beta activation
Gα_{12/13} → activation and PM recruitment of Rho GEFs

→ the most important groups are α_i and α_s, which have an impact on cyclases

❖ ALPHA SUBUNIT

The reaction mediated by the alpha subunit is the following one.

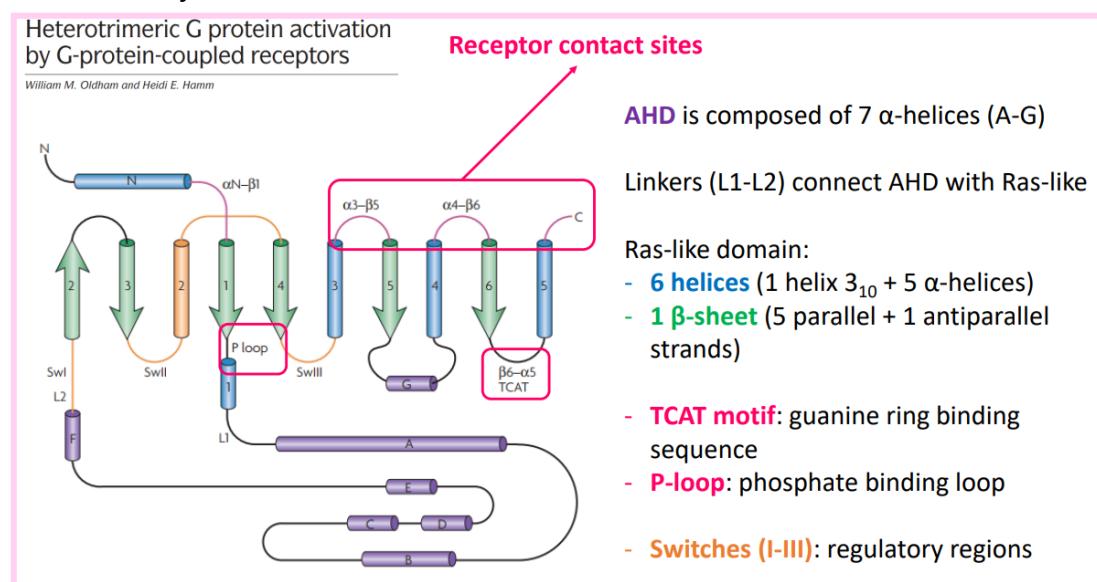


The functions of the alpha subunit are three:

1. **mediates the interaction between $\beta\gamma$ dimer and the receptor**, as it is placed in the middle. We can say it provides an interface.
2. **binds GTP/GDP** and hydrolyses GTP
3. **provides downstream signaling**

Now that we have talked about the function of the alpha subunit, we can go more in depth about its structure.

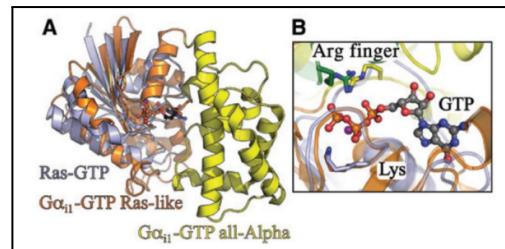
- as we can see AHD (*all helix domain*) and the RAS-like domain are not one after the other, rather they are combined.



- **TCAT** binds GTP/GDP
- **P-loop** understands if you have three or two phosphates, meaning they understand if GTP or GDP are bound.

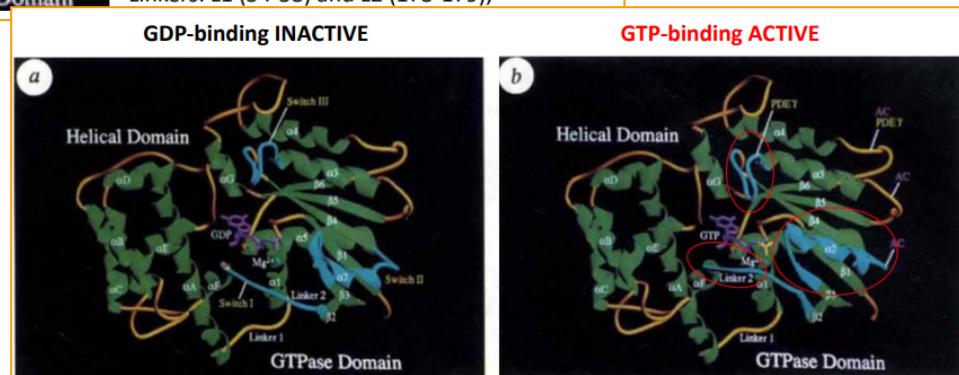
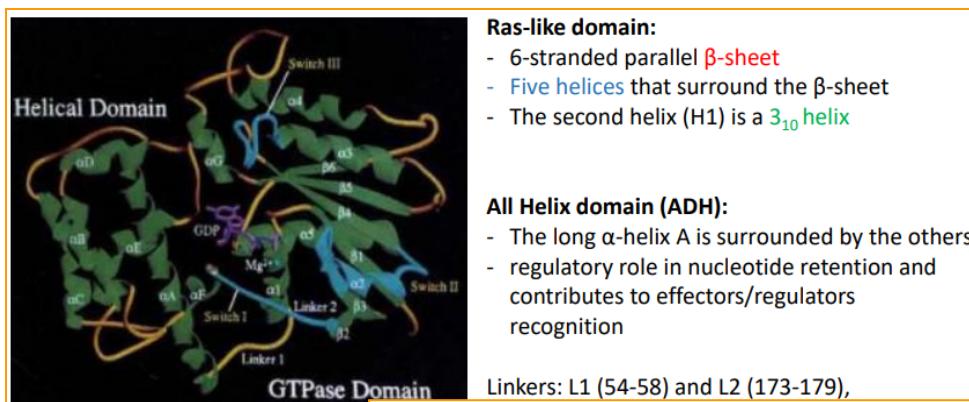
The *RAS like* domain is very similar to *RAS* domains, being easily superimposable. Of course by superimposing RAS with an alpha subunit, the AHD isn't covered at all.

The **binding of GTP happens in a pocket of the RAS like domain**, at an interphase between RAS like domain and AHD domain. Still, binding of GTP depends only on the RAS like domain.



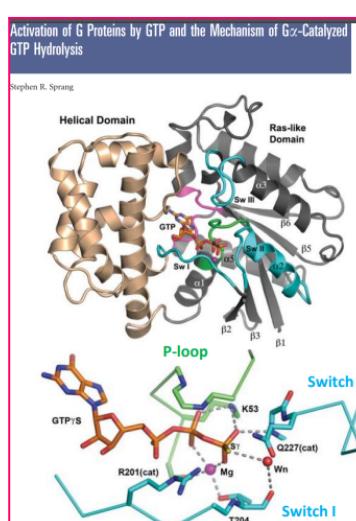
The active sites share some highly **conserved structural elements**, including:

- A conserved **Lys** in the P-loop
- The **Arg finger** (intrinsic from the AHD in $\text{G}\alpha$, in small GTPases it must be provided by the GAP protein). This arginine is **necessary for the catalytic mechanism**, and it can be put into place only when RGS and alpha subunit cooperate.



Scientists were able not only to determine the structure of a GDP binding heterotrimeric protein, but also of an active, GTP binding one.

Differences between active and inactive states were highlighted: they are what we call **SWITCHES**. They contain catalytic residues and undergo conformational changes on conversion of GTP to GDP.



Nucleotide binding site

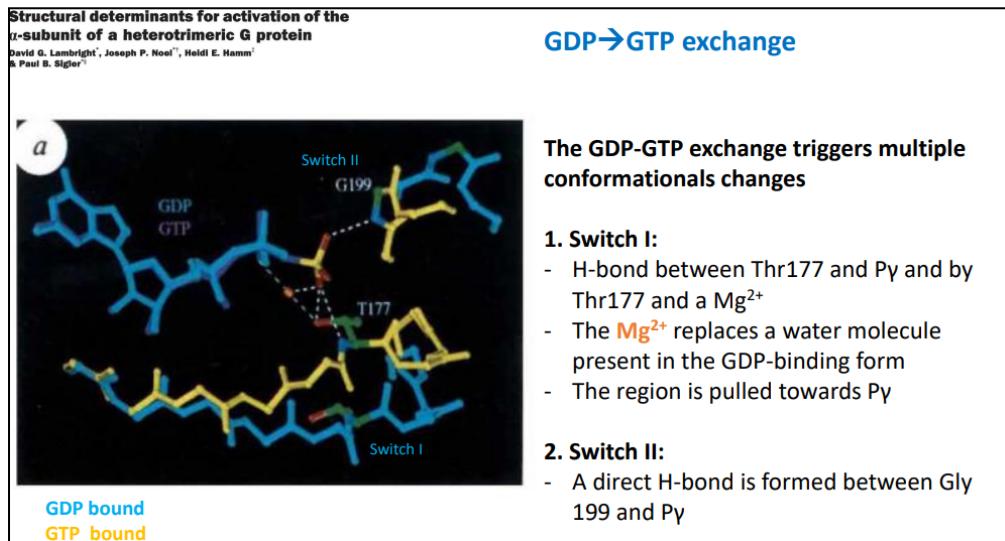
- Similar in structure and sequence to that of small GTPases
- Guanine recognition motif
- **P-loop** that envelops $\text{P}\alpha$ and $\text{P}\beta$
- Two dynamic structural elements (**switch I and II**) that respond to the presence or absence of Py. They contain catalytic residues and undergo conformational changes on conversion of GTP to GDP
- **Switch III** participates in effector/regulator binding

Switch I and **Switch II** can sense the presence/absence of the gamma phosphate, changing conformation accordingly. They sense the status of the phosphate:

- **directly**
- **indirectly**, thanks to a magnesium ion and a water molecule

Switch III instead has a regulatory function.

Now we can understand **what happens during the EXCHANGE**:



In yellow we have the active conformation, in blue the inactive one.

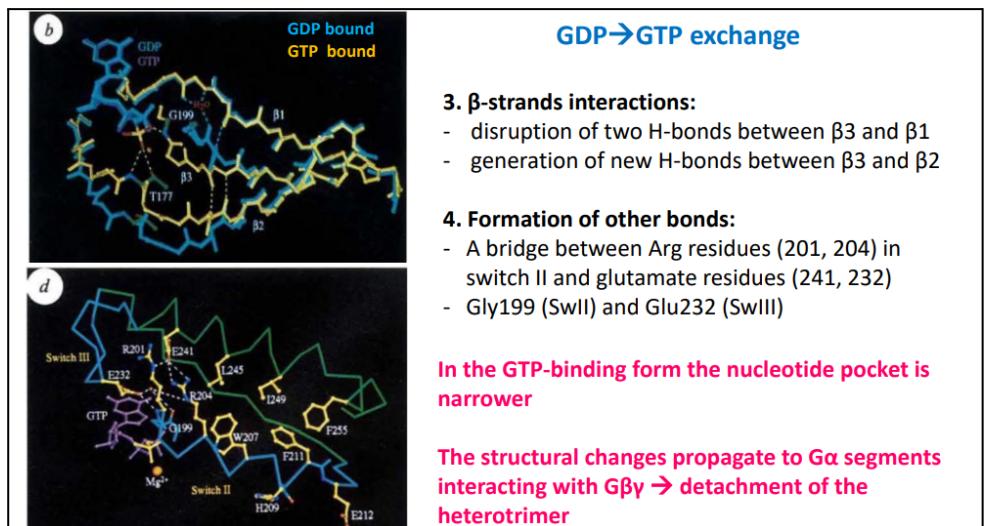
When the **Py is added**:

- SWITCH I:** an hydrogen bond is formed between Thr177 and Py.
SWITCH II: an hydrogen bond is formed between Gly199 and Py.

This means that with GTP, the **molecule grows closer to it and the binding site is tighter**.

Other conformational changes happen near the area:

For example we have **disruption of hydrogen bonds** between $\beta 3$ and $\beta 1$ strands, **and the formation** of hydrogen bonds between $\beta 3$ and $\beta 1$. Once again, this makes the structure tighter. Lastly, **bridges are formed** between arginines, and between glutamate residues.



These **small structural changes** propagate in the alpha subunit, leading to the **displacement of the RAS like domain** with respect to AHD and the **detachment of the beta-gamma dimer**.

Also the **HYDROLYSIS** of GTP happens at the same site. The energy of GTP is used as a barrier to prevent the association of the beta-gamma dimer.

- two residues are necessary for GTPase activity: the first is Arg_{cat} , which is the *arginine finger* from earlier) and Gln_{cat}

Our **GROUND STATE** in this case is the one of **GTP binding**.

Stephen R. Sprang

1. GROUND STATE – GTP binding

1. GROUND STATE – GTP binding

- Catalytic site pre-organization presents a significant activation energy barrier to $\text{G}\alpha$ GTPase activity
- Py forms contacts with the conserved **Lys53** (P-loop), **Gln227** (SwI), **Thr204** and **Arg201** (SwI)
- Py also interacts with Mg^{2+} and a molecule of water (Wn)

In the ground state our Py from GTP is making contact with several other residues, a water molecule and magnesium ions.

As we cannot represent a transition, in the next step we already see GDP.

The **water molecule and the magnesium positions are different**, highlighting how the bonds with Py are changing, since it's about to be released.

The **hydrogen bonds with the residues** (like the one with $\text{Arg}_{\text{catalytic}}$ and Gln_{cat}) **also change**.

Stephen R. Sprang

1. GROUND STATE – GTP binding

2. TRANSITION STATE

- Gln_{cat} has an allosteric role, aiding in the pre-organization of the enzyme-substrate complex and interacting with the catalytic Wn
- the **catalytic Wn** occupy a position expected for a nucleophilic engage → in-line attack of the Py
- The **hydrogen bond network** involving Arg_{cat} , Gln_{cat} , W_{nuc} , and the Py orients W_{nuc} for nucleophilic attack and stabilize developing charge at the β - γ bridge leaving group oxygen
- **Dissociative mechanism**

Stephen R. Sprang

C

TRANSITION STATE – GDP and AlF_4^- binding

D

3. TERNARY COMPLEX – GDP and Pi binding

3. TERNARY COMPLEX

- Pi is within hydrogen bonding distance of the $\text{P}\beta$, stabilized by **Lys46** (P-loop) and **Arg_{cat}** (switch II)
- To accommodate Pi, **switch II adopts a more regular helical structure** affording its movement away from the catalytic site

GTP hydrolysis induces modifications in switch regions → return to GDP conformation and recruitment of G β γ

Iommari - Genomics 2022

GTP binding site coming from RAS and the GTP binding site coming from the alpha subunit are slightly different:

- magnesium ions, water, charged residues in the catalytic site are the same
- the **arginine finger** presents a different:
 - ◆ in the heterotrimeric G protein, it belongs to the alpha subunit. RGSs do not have the arginine finger.
 - ◆ instead in RAS domains, the arginine finger is absent so it must come from the GAP.

This means that **RGS** *increases* the activity, while the association between **RAS and GAP** is *necessary*.

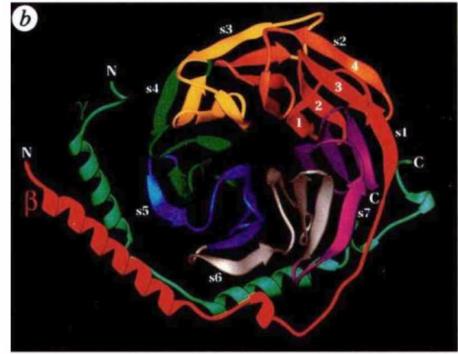
❖ BETA GAMMA DIMER

There are different versions both of the beta and of the gamma subunits, depending on the tissue they are expressed in.

An important structure in the beta gamma dimer is the **beta propeller**, present in the beta subunit. It's formed of 7 beta sheets arranged like blades. Each sheet is made up of four antiparallel strands that are wrapped around themselves. The beta propeller mediates protein-protein interaction.

- The beta propeller also contains WD repeats, short structural motifs (called like this for the repetitions of tryptophan and aspartic acid).

At the N term we have an alpha helix of the beta propeller interacting, through a coiled coil, with the gamma subunit. The gamma subunit also interacts with the blades 5 and 6 of the propeller.



The beta-gamma dimer (GB1-5,Gy1-13) is present in many proteins, and it can be more or less conserved.

- The N-term, C-term and hinge of the **gamma subunit** are **less conserved**, being located on the external face of the dimer. The variability is an effector of specificity and diversity.
- The **beta subunit is instead much more conserved**, especially the part in contact with the alpha subunit.

Different G $\beta\gamma$ dimer combinations are not equal, and may have distinct functions.

The **beta-gamma dimer** was initially defined as the negative regulator of G α activity, preventing spontaneous/improper activation.

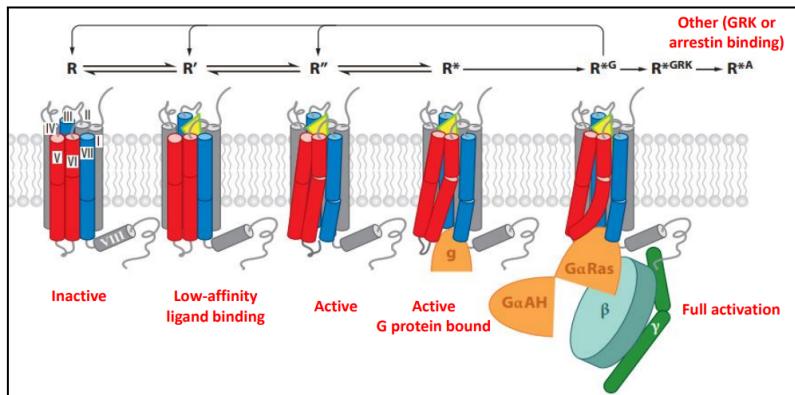
It plays a key role in **modulation** of canonical effectors in **GPCR-dependent signaling**.

Recently, the dimer was identified as **regulator of transcription**, anterograde and retrograde trafficking and modulator of second messenger molecule generation in intracellular organelles.

Examples of transcription regulator functions are the following:

1. Activates the cardiac muscarinic-gated K $^{+}$ channel (Kir 3 family) in cardiac cells together with 4 PIP2 molecules
2. Inhibits the activation of the voltage-dependent Ca $^{2+}$ channels protein family (Cav family)
3. Regulates AC activity (depending on the AC type and the composition of G $\beta\gamma$ dimer)
4. Stimulates PLC
5. Activates PI3K β and PI3K γ
6. Initiates MAPK signaling cascades through different mechanisms

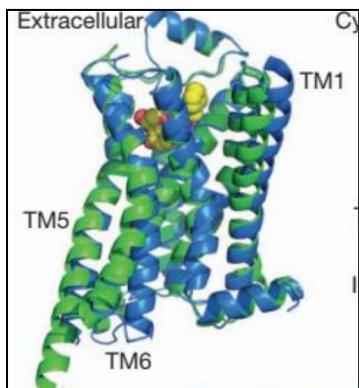
❖ RECRUITMENT OF THE G PROTEIN



Looking at the β_2 adrenergic receptor:

The structure of the heterotrimeric G protein was first crystallized years ago. Scientists found out that:

- $\text{G}\beta\gamma$ dimer does **not interact with the ligand**
- the **ligand** has extensive **interactions with $\text{G}\alpha$** .
 - ◆ Elements involved in the interaction are ICL2, TM5, TM6, hN, h4, h5 and $\beta 3$.
Where **hN and h5 have the biggest interactions**, and h5 is the element which is inserted on the ligand.



Here we see the active and interactive conformations:

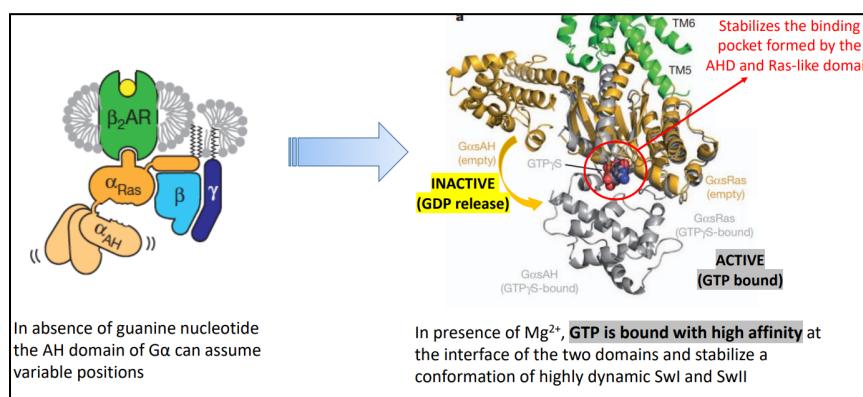
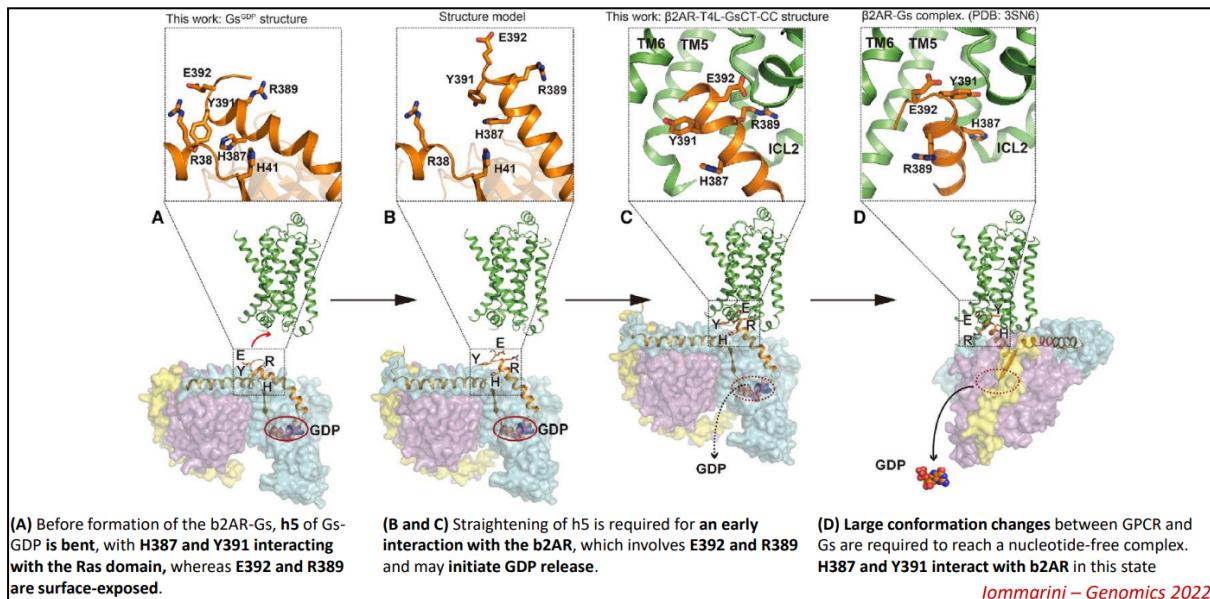
- blue → inactive: the cavity has a close conformation
- green → active: alpha helices are further away from the center of the cavity: this **opening is due to the presence of the ligand**.

By **moving TM5-6**, **ligand binding opens a cavity** where the $\alpha 5$ helix of $\text{G}\alpha$ can be docked.

In the ligand binding site there are different interactions:

- **Receptor core:** *Non-polar interactions*, packing of Y326-R131-Y391 and *electrostatic interactions*
- **Intracellular side:** *polar interactions*
- **other interactions** such as F139 entering into a hydrophobic pocket as to stabilise the overall interaction and polar interactions position ICL2

When the heterotrimeric protein binds GDP it is inactive: helix 5 is bent and not extended, which means that the receptor cannot interact. Activation of the receptor only happens with a big conformational change that promotes the relocation of GDP and the contact of the alpha helices.



When the **α RAS-like domain is bound to the βγ dimer**, we cannot observe activation due to its high affinity towards GDP. Here, the **α AH domain can assume variable positions**.

However, when **GTP is bound**, activation of our domain is present. The switches become more stable and the βγ dimer is released.

When GTP is bound:

1. Movement of the C-term of the α5-helix away from the β6-strand to permit further interactions
2. Rotation of Ras domain relative to the receptor
3. Formation of more extensive interactions between the β2AR ICL2 and the N-term of Gαs

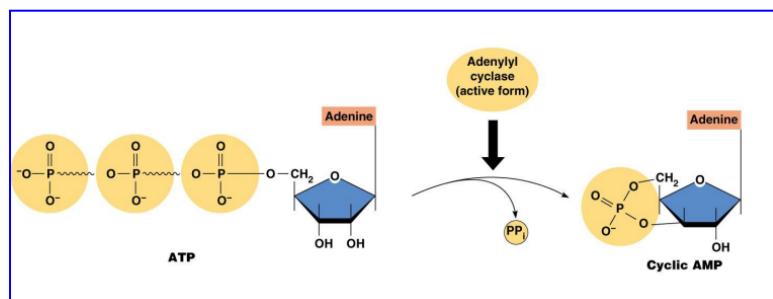
The ACTIVE conformation (GTP-binding) is not compatible with the binding of Gβγ → **G Protein dissociation leads to downstream signaling.**

❖ ADENYLYL CYCLASE (AC)

Adenylyl Cyclases are a wide class of enzymes responsible for cAMP production.

cAMP is a second messenger important in many biological processes. Cyclic AMP is synthesized from ATP by adenylate cyclase located on the inner side of the plasma membrane.

- **Adenylate cyclase** is activated by a range of signaling molecules through the activation of adenylate cyclase **stimulatory G (Gs)-protein-coupled receptors**.
- **Adenylate cyclase** is inhibited by agonists of adenylate cyclase **inhibitory G (Gi)-protein-coupled receptors**.

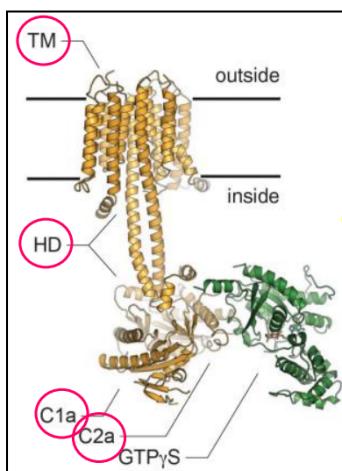
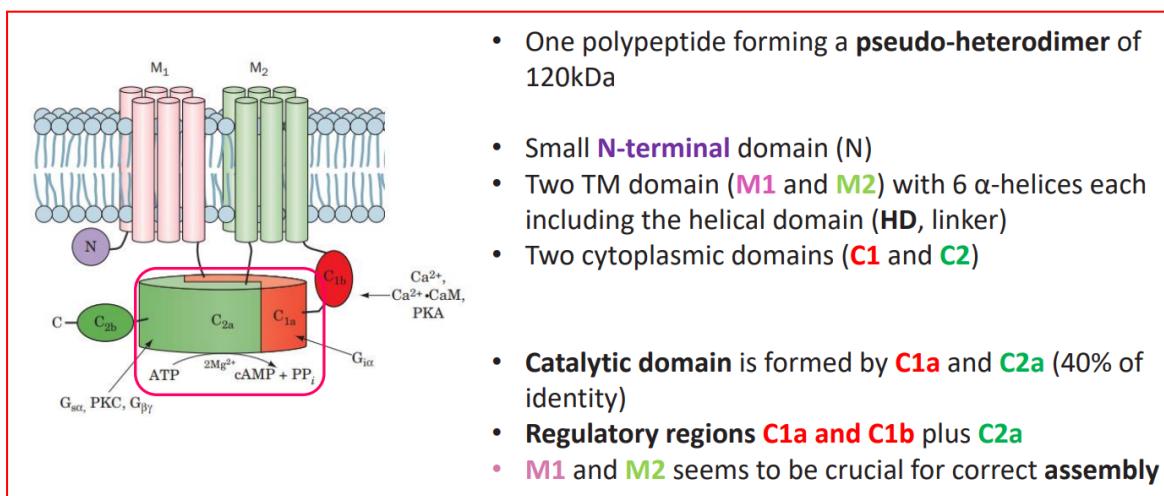


This mechanism is highly conserved. Once cyclic AMP is formed a signaling cascade begins.

There are **six unrelated classes of ACs**. The most important is class III (unique in animals). All the members share the same catalytic function, however regulation might be different.

Mammals have **9 (+1) different isoforms** (tissue-specific) differing in their regulatory properties.

All class III ACs have the same characteristics:



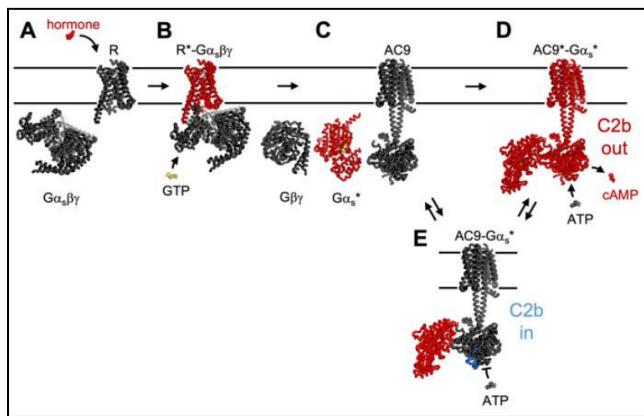
In particular we can consider AC9.

→ The **TM helices** have high similarity and can be superimposed
 → **HD** is made by H6 and H12. It has a coiled-coil structure and is critical for protein function. Mutations on this domain can cause pathogenic diseases.
 → The **catalytic domain** is similar to that of other ACs. It's formed by **C1a** and **C2a** domains organized in a pseudo 2-fold symmetry. The nucleotide binding site at the interface of the two half-sites. **The catalytic site requires Mg2+** (physiological ligand) or Mn2+ **for catalytic activity** (Mn2+ is more effective and has higher affinity). The domain is overall very conserved.

Most important **INTERACTIONS BETWEEN ACS AND GPCRS** are mediated by:

- Insertion of **Switch II** ($G\alpha$) helix into a groove formed by helix $\alpha'2$ and $\alpha'3$ of C2a
- A **loop between the HD1 and C1a** domains is also created.

Now we can see the **AC9 model of activation.**



- A. An **agonist** (hormone) binds and activates GPCR
- B. The activated receptor R^* bound to the heterotrimeric G protein activates the $G\alpha_s$ protein (GDP- GTP exchange)
- C. The **activated $G\alpha_s^*$ subunit interacts with AC9 through SwII**
- D. AC9*- $G\alpha_s^*$ is the **fully activated state**: ATP binding active site is available for the incoming substrate (**C2b out**)

AC9- $G\alpha_s^*$ can also be in the **occluded conformation**: with the active sites occupied by a fragment of the C2b domain (**C2b in**). (E)

This is in equilibrium with active conformation: auto-regulatory mechanism to prevent cAMP overproduction.

❖ PROTEIN KINASE A

PKA is a *cAMP-dependent protein kinase*.

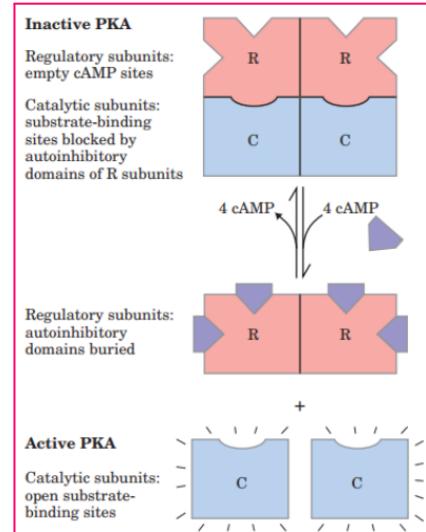
PKA consists of **two types of subunits**:

1. an **autoinhibitory regulatory subunit (R)** of 49kDa with high affinity for cAMP
2. a **catalytic subunit (C)** of 38 kDa

In **absence of cAMP**, R subunits and C subunits form an **enzymatically inactive heterotetramer R_2C_2** .

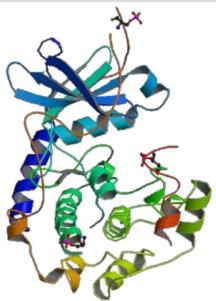
The **binding of cAMP** by R subunits determines the dissociation of R_2C_2 in a R2 dimer and two C monomers.

PKA was actually the first kinase to ever be crystallized. It has a conserved kinase core shared by all members of the protein kinase superfamily.



The **catalytic subunit** of PKA contains the following elements:

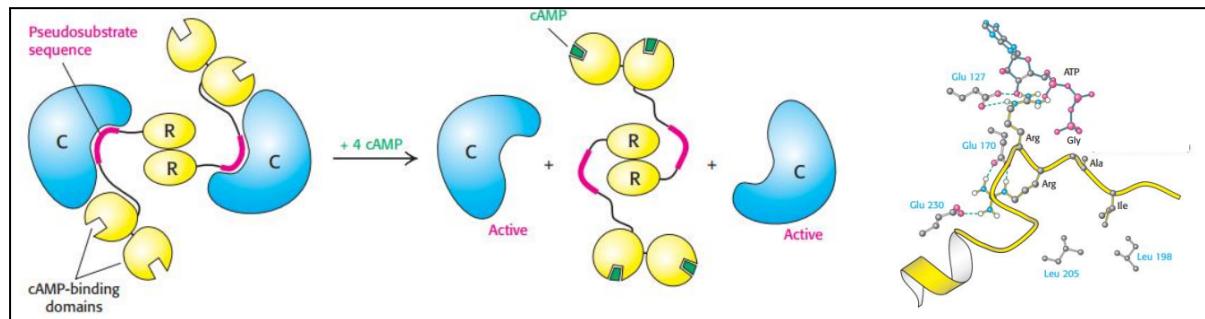
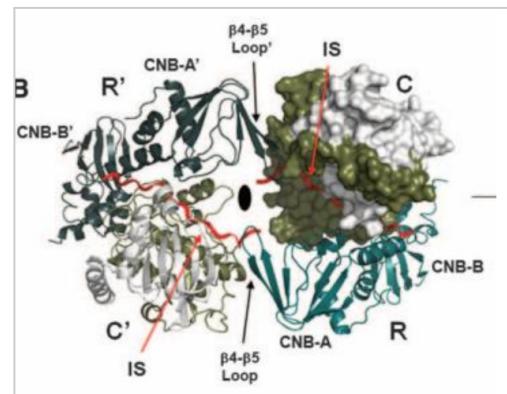
- The large **C-lobe** is composed by α -helices: it facilitates substrate recognition and provides the catalytic machinery for phosphoryl transfer
- The smaller, more dynamic **N-lobe** contains an antiparallel β -sheet
- The **catalytic pocket** is placed at the interface of the two lobe



The **regulatory subunit** instead works as a symmetrical dimer.

There are **two classes of R subunits** (RI and RII). The R subunit contains the following elements:

- Flexible **N-term dimerization/docking (D/D) domain**
- Flexible **linker w/ an inhibitor site (IS)** recognizes the active-site cleft of the C subunit.
 - ◆ RI have Gly or Ala at their P-site \rightarrow inhibitors as pseudo-substrates
 - ◆ RII have a Ser at their P-site \rightarrow inhibitors as substrates. Phosphorylation slows the rate of association with C subunit
- Two highly conserved **cyclic nucleotide binding domains** (CNB-A and CNB-B) \rightarrow bind cAMP



Each RI chain contains the sequence Arg-Arg-Gly-Ala-Ile, identical to the consensus sequence for phosphorylation with the exception of Ala instead of a Ser (pseudo-substrate sequence).

1. In the **heterotetramer** the **pseudosubstrate sequence occupies the catalytic site** \rightarrow substrates not allowed to enter
2. The binding of cAMP to the RI2 subunits induces a **conformational change** that extracts the **pseudosubstrate sequences** from the catalytic sites.
3. The **released C subunits are free to bind and phosphorylate target proteins**

cAMP-binding: Arg381R (CNB-A) and Tyr397R (CNB-B) are bound to the adenine rings of cAMP

summary

1. Adrenaline **binds** the β_2 -AR (**GPCR**)
2. Activated receptor recruits the heterotrimeric G-protein and favors the **GDP-GTP exchange**
3. **Active G_as** subunit triggers the activity of AC and **cAMP is produced**
4. The second messenger binds the R subunits of PKA **releasing the autoinhibition**
5. PKA active **C subunits phosphorylate their targets**

- **FAST RESPONSE** (sec) direct phosphorylation of enzymes (i.e. glycogen mobilization)
- **SLOW RESPONSE** (hours) phosphorylation of the TF CREB (cAMP responsive element BP) → transcription regulation

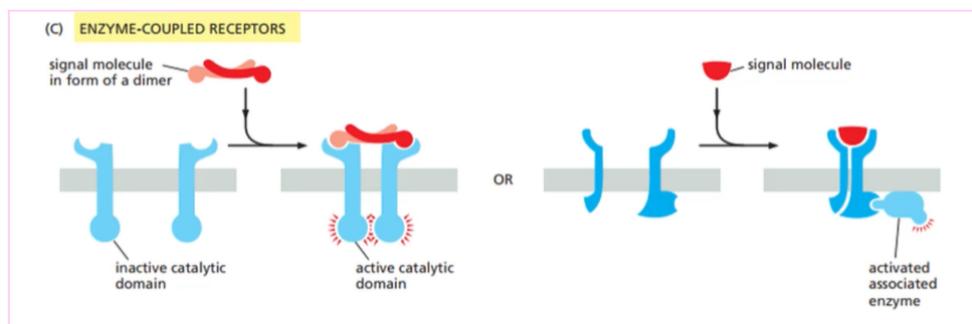
RECEPTOR TYROSINE KINASE

We have seen that there are different types of receptors and mechanisms through which the signal can be transduced in the cell.

Now we focus on **enzyme coupled receptors**. These receptors can be of two types:

- ★ intrinsic
- ★ direct enzyme interaction

Here we study receptors with an intrinsic enzymatic activity, which have at least one domain serving that function. Usually, these receptors work as kinases.



During signaling, usually the **receptor** stands in an **inactive** (monomeric) form. Upon the **binding** of the (dimeric) ligand, **dimerization** of the receptor is induced.

Usually, the **receptor binds the signaling molecule in the extracellular part of the membrane**, while the **catalytic site is intracellular**.

Sometimes dimerization does not occur, but in that case the binding of the ligand in the extracellular side still causes conformational changes which have an effect on the cytosolic side.

- the most studied class of these receptors is the family of **RECEPTOR TYROSINE KINASES** (RTKs). As the name suggests, these receptors are able to phosphorylate tyrosines (not threonine or serines!)
 - ◆ all the ligands for RTKs are small proteins which can be secreted or exposed by the starting cell.
 - ◆ 58 known RTKs in human cells, classified into 20 structural subfamilies, each dedicated to its complementary family of protein ligands.
 - ◆ Phosphorylation of Tyr on the cytosolic part of the receptor creates **P-Tyr docking sites for various intracellular signaling proteins** that relay the signal

Here we have a table of the most important tyrosine kinases: all of them are involved in cell growth or survival.

TABLE 15-4 Some Signal Proteins That Act Via RTKs

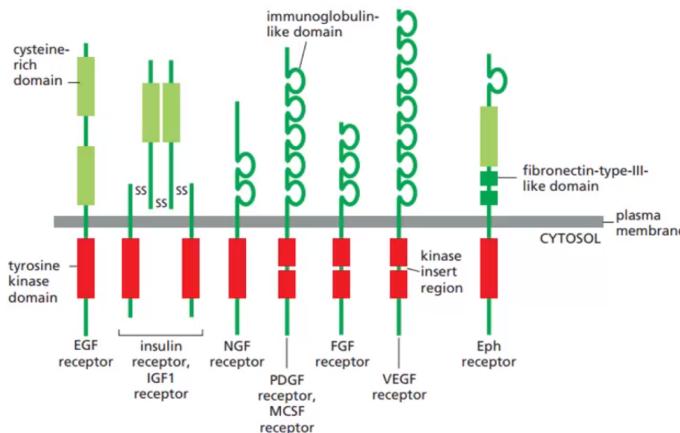
Signal protein family	Receptor family	Some representative responses
Epidermal growth factor (EGF)	EGF receptors	Stimulates cell survival, growth, proliferation, or differentiation of various cell types; acts as inductive signal in development
Insulin	Insulin receptor	Stimulates carbohydrate utilization and protein synthesis
Insulin-like growth factor (IGF1)	IGF receptor-1	Stimulates cell growth and survival in many cell types
Nerve growth factor (NGF)	Trk receptors	Stimulates survival and growth of some neurons
Platelet-derived growth factor (PDGF)	PDGF receptors	Stimulates survival, growth, proliferation, and migration of various cell types
Macrophage-colony-stimulating factor (M-CSF)	M-CSF receptor	Stimulates monocyte/macrophage proliferation and differentiation
Fibroblast growth factor (FGF)	FGF receptors	Stimulates proliferation of various cell types; inhibits differentiation of some precursor cells; acts as inductive signal in development
Vascular endothelial growth factor (VEGF)	VEGF receptors	Stimulates angiogenesis
Ephrin	Eph receptors	Stimulates angiogenesis; guides cell and axon migration

As we have already mentioned, **ligands** for RTKs are just **small proteins**. This is opposed to GPCRs, which have several different molecules as ligands.

→ Frequently (not always), ligands are **dimeric**.

Examples of ligands are *epidermal growth factor* or *insulin*.

❖ GENERAL STRUCTURE OF RTKS



Extracellular domain (ECD): ligand binding

One TM α -helix

Intracellular domain (ICD): Tyr kinase domain

Juxtamembrane (JM) and C-term domains: regulatory role

Structure and mechanism of activation are conserved from nematodes to humans

Mutations → cancer, diabetes, inflammation, bone disorders...

Each of the figures represents a different RTK. Nevertheless, we can see that all of them have similar characteristics. In particular, **shared elements are**:

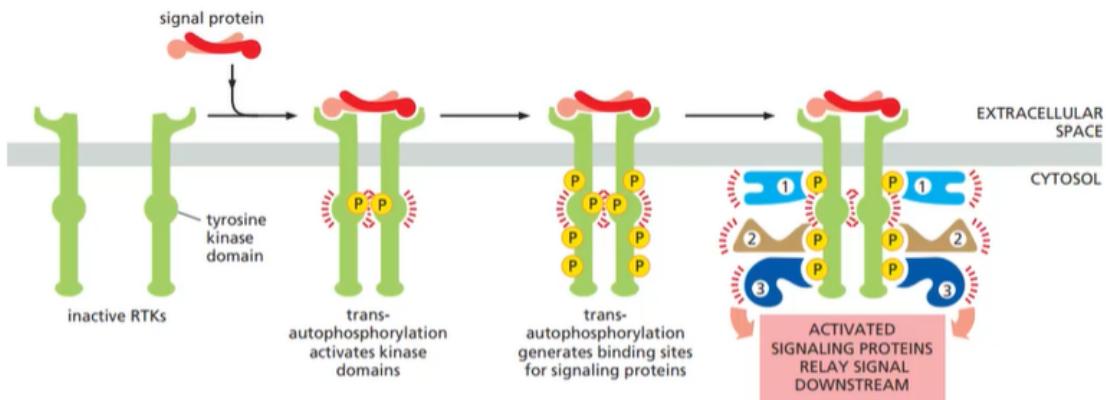
- **extracellular domain** that binds the ligand: this is essential because, being the ligand a protein, it cannot pass the membrane.
- **one transmembrane helix**
- **intracellular domain** which is the **tyrosine kinase domain**, having an enzymatic activity.

There can be some differences in the intracellular domain: it's always a tyrosine kinase domain but it can work either compactly, or with a small region in between the domain. This region has a regulatory function.

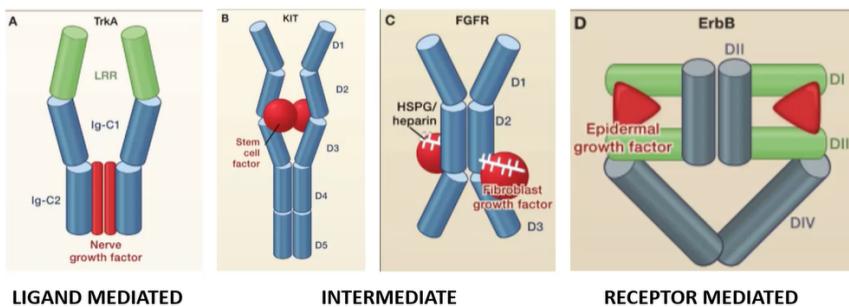
- **juxtamembrane section**, which is close to the transmembrane helix
- **C term**, which can be different in different RTKs but is always present. It has a regulatory role.

While the tyrosine kinase domain is usually identical, or real similar, between different RTKs, **extracellular domains can contain different structural domains** belonging to different subfamilies.

The general **ACTIVATION MECHANISM** is depicted here, but it can be different for different receptors. In general:



1. inactive RTK is **monomeric**
2. dimeric protein induces **dimerization**
3. the proximity of the two TK domains lead to **transactivation** and **transphosphorylation**: this means that chain A is able to phosphorylate chain B and vice versa. This leads to a series of phosphorylations which induce the presence of a series of docking sites for other molecules.



As there are differences between the extracellular domains, these can be activated in different ways. There are two opposite situations:

Activation can be:

- **ligand mediated:** in this case there is no actual contact between the two monomers of the receptor.
- **receptor mediated:** the interaction of the two monomers of the receptor is mediated exclusively by the mediator itself.
- **intermediate solutions:** where the interaction is mediated by the receptor *and* the ligand.

❖ TYR KINASE DOMAIN

Tyr KINASE DOMAIN:

- Similar to other kinases (PKA, CDK...)
- N-lobe with a β -sheet and C-lobe all helices
- α -helix C in N-lobe which movement allows catalytic activation
- Activation loop** in the C-lobe regulates the catalytic function

The Tyr Kinase domain is similar to the kinase domains of other molecules we have already seen.

- N lobe and C lobe are present
- activation loop** is present in different conformations: occluding the catalytic site in the inactive conformation, and protruding outside in the active conformation.
- helix C** can assume different conformations, causing activation of the enzyme

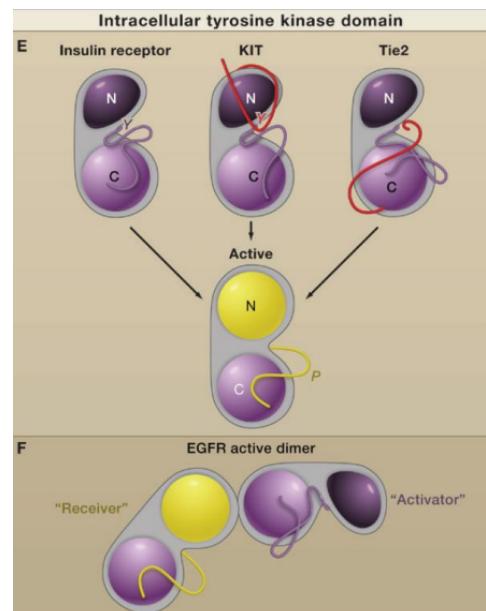
In the case of CDKs, it was the interaction between CDK and cyclin that promoted the conformational change causing activation.

Here:

- the **active** forms are structurally **very similar**
- the **inactive** forms may assume **different configurations** (depending on the RTK). This is because there are different mechanisms of autoinhibition, which must be released for RTK function.

We can consider three main types:

- InsR**: the **activation loop** interacts directly with the active site blocking the access to substrates and/or ATP. **Phosphorylation** of key Tyr disrupts autoinhibitory interactions, leading to the active state.
- KIT**: The **juxtamembrane** region interacts with the active site (i.e. α C helix, activation loop), stabilizing an inactive conformation. **Phosphorylation** of key Tyr destabilizes autoinhibitory interactions, leading to an active conformation.
- Tie2**: the inactivation mechanism is the same as the KIT one. The only difference is that the **C-term has regulatory function**



These activation mechanisms of the Tyr Kinase domains are similar to PKA. However, EGFR is different, and its activation mechanism is similar to cyclin/Cdk dimer.

- EGFR**: Allosteric activation by direct contacts between the C-lobe of one TKD (activator) and the N-lobe of another TKD (receiver). The activator destabilizes autoinhibitory interactions in the activation loop of the receiver. **No activation loop phosphorylation is required.**

❖ EPIDERMAL GROWTH FACTOR: EGFR

EGFR is an *epidermal growth factor*. As it was one of the first RTKs that were found, it was thought to be a prototype. Only later we found out it was an exception.

- it's related to mutations in different tumors, like breast cancer.
- it is a small subfamily in the family of RTKs, made up of four members: *EGFR*, *ErbB2*, *ErbB3*, *ErbB4*.
- **The EGFR family is able to recognise different ligands**, a list of which we have below.

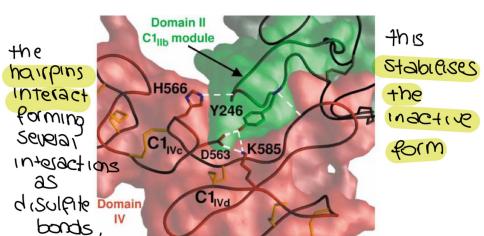
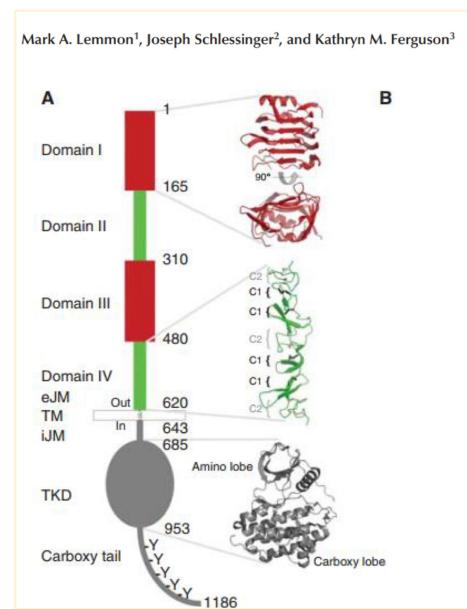
EGF and HB-EGF: epidermal GF and heparin-binding EGF
 TGF- α : Transforming GF a
 ARG: Amphiregulin
 EGN: Epigen
 EPR: Epiregulin
 BTC: Betacellulin
 NRG1-4: Neuregulins

We have:

- a **large extracellular domain**.
- transmembrane helix with the **juxtamembrane segment**
- **TK domain**
- long **C term tail** which presents Tyrs that can be phosphorylated, creating docking sites.

There is a repetition of similar domains:

- **domains I and III** provide binding for the **ligand (EGF)**
- **domains II and IV** are cysteine rich and are able to make **di-sulfite bonds**.

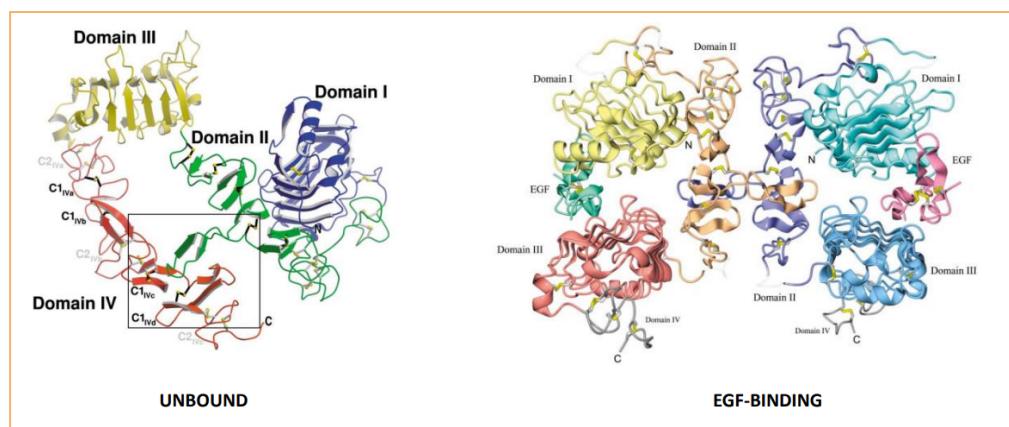


When the receptor is **not active**, the conformation that is present is the following. Of course, we are looking at a **monomer**.

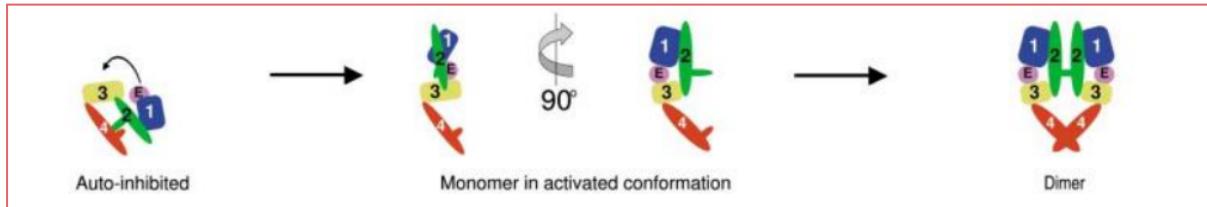
In this case, **domains II and IV** are involved in an **intra-receptor interaction** through **hairpins** that prevent dimerization. They make di-sulfide bonds.

Instead the **active form** is, of course, **dimeric**. The activation, due to the presence of the ligand, instead **causes consistent conformational changes** at the level of the extracellular domain that are then propagated in the intracellular one, favoring dimerization.

One of these conformational changes is the fact that **domains II and IV** are no longer interacting in the same monomer, but with another one, leading to **dimerization** through the same **hairpins**.



Basically:



- ★ the initial conformation is an autoinhibition: this is due to the formation of intramolecular bonds through hairpins
- ★ the final active form has a heart-like shape, in which the hairpins that mediated inhibition provide, in this case, contact.

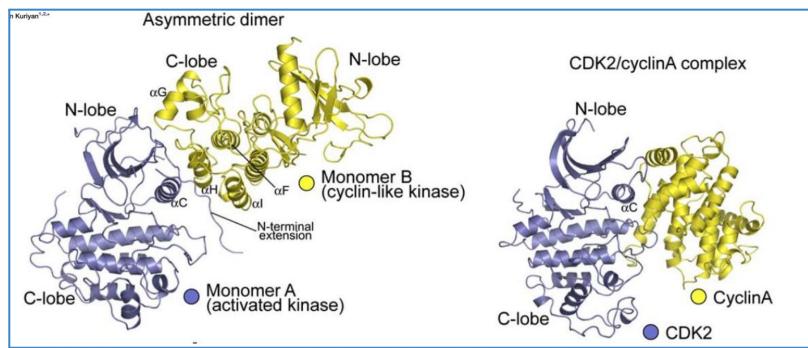
The **binding of the ligands (2 x EGF)** initiates the conformation changes. However it's domain II and IV **hairpins which provide complete activation**.

- as EGF doesn't provide any additional surface for dimerization, we say that **this activation is purely receptor based**.

In the **dimer**:

- The C-lobe of the **activator** TKD interacts with the N-lobe of the **receiver** TKD

We can see that these allosteric conformational changes are similar to that of CDK/Cyc activation: the C lobe of the activator is akin to a cyclin, inducing a conformational change at the level of the receiver domain.



- Helix C moves from a faraway position close to the interface between the two N and C lobes where the active site is present. The activation loop, present in the active site, is displaced far away during activation.

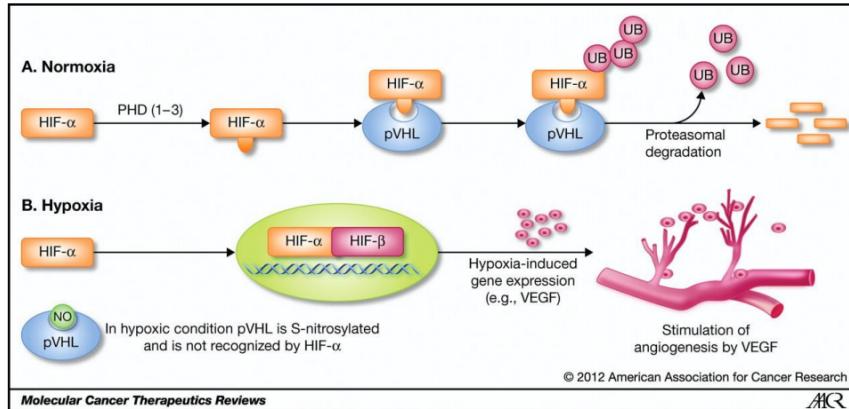
Transphosphorylation between the two monomers is necessary for activation: **the activator stimulates the activity of the receiver, which phosphorylates the activator**.

- In Cdk2, Thr160 was necessary for the final activation of the dimer. In this case we of course don't have the same residue: we have a **Tyrosine that is always phosphorylated**, however it is not enough to activate on its own dimerization (we talk about **basal activity** which is not enough to have a downstream signaling.). This means that **phosphorylation of the activation loop is not required**.

❖ HYPOXIA AND ANGIOGENESIS: VEGFR

HYPOXIA is a condition in which the body or a region of the body is deprived of adequate oxygen supply at the tissue level.

The transcription factor family called **HIF** (*Hypoxia Inducible Factors*) is able to sense when oxygen levels are inadequate. This family is dimeric and it's composed by two subunits: α , which is the oxygen sensing one, and β , which is the constitutive one.

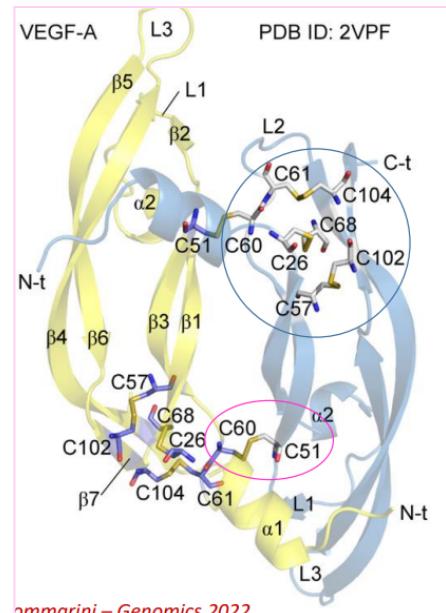


- If the **oxygen levels are normal**, residues of the alpha subunit can be hydroxylated into prolines and recognised. This activates protein degradation. Which means that, if there is enough oxygen, **HIF protein is degraded**.
- If the **oxygen levels are low**, the α subunit cannot hydroxylate its residues, hence remaining stable. HIF goes into the nucleus where it works as a transcription factor, leading to **activation of multiple responses**:
 - increase in glucose consumption
 - stimulation of new vessels: **ANGIOGENESIS**

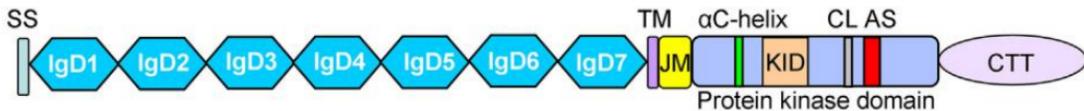
Vascular Endothelial Growth Factor (**VEGF**) signaling tries to stimulate the production of new vessels (for the need of oxygen).

VEGF It is a very small protein with different isoforms (so we need to differentiate the activities but the overall function is the same) and works as a homodimer.

- The protein is formed by two chains (two identical **antiparallel β -sheets** of 4 strands), kept as a dimer by **several disulfide bonds**
- Those isoforms are produced by alternative splicing (ex 6a, 6b, 7). Some even contain heparin binding sites.
- Two receptors binding sites (VEGFR1 and 2 are different).
- The **receptor** for this protein VEGFR1 is a **RKT**



Now we can look at the receptor for this protein, which is called *Vascular Endothelial Growth Factor Receptor (VEGFR)*.



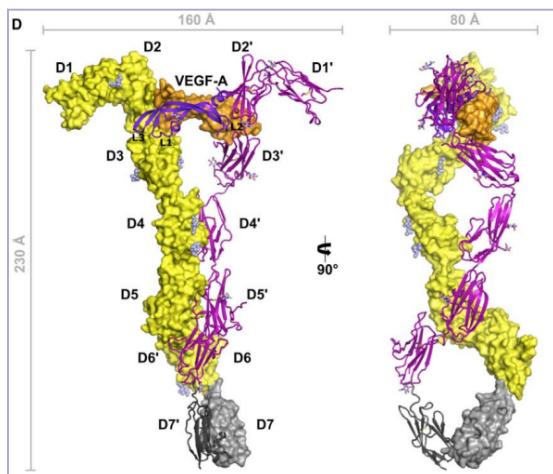
Starting from the N terminal going towards the C terminal:

- **Extracellular segment** with 7 immunoglobulin like domains (IgD1-7)
- **TransMembrane** segment
- **JM** segment
- Intracellular protein-tyrosine kinase domain that contains a **kinase insert domain (KID)**, hence being broken in several domains. KID is of about 70 residues that occurs between the D and E helices.

In this case, **both TK domains are activated** (unlike EGF, where only one monomer is phosphorylates). When active, they **activate in trans downstream** tyrosine residues.

There is diversity in the VEGF, hence we also have **diversity in the receptor**. There are different kinds of receptors: 1, 2, 3.

- there can be accessory subunits connected to the core ones
- all three receptors are able to create blood vessels, only 2 and 3 are able to create lymphatic vessels
- **VEGFR-1** can also be **soluble**, working as a regulatory receptor that competes with the membrane receptors. It binds VEGF, inhibiting its activity.



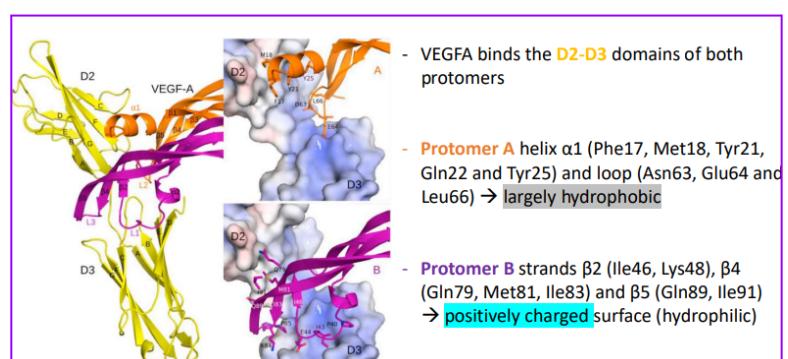
Here we can see the **ligand VEGF-a** binding receptor VEGFR (containing the two **monomers**). Here the receptor and the ligand are in an **intermediate situation**.

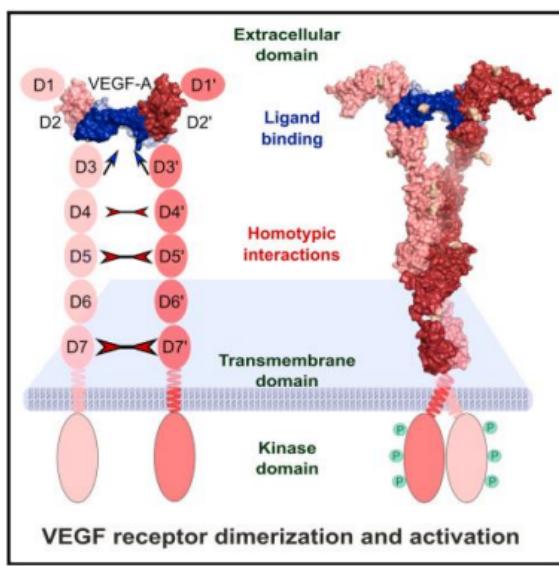
- VEGF is necessary for the interaction between the two monomers
- the monomers are also able to interact between themselves without the ligand

D1 is bent away from D2 and does not interact with VEGF-A. Instead, D4, D5 and D7 form homotypic interactions

It's important to keep in mind that there are **two interactions** that occur between ligand and receptor. **VEGFA binds the D2-D3 domains of both protomers**.

- largely **hydrophobic** interactions
- positively charged surface (**hydrophilic**)





As a mechanism of activation they proposed that:

1. the presence of the ligand induces, mediated by hydrophobic and hydrophilic interactions, a series of homotypic interactions at the level of the extracellular part.
2. homotypic interactions in D4, D5 and D7 drive the dimerization process and receptor activation through the vicinity of the two monomers.

This is a classical example for RTKs.

Currently, there are not many structures of the TK domain of VEGFR (only inactive ones, none active). However:

- Kinase domain is interrupted by an insertion of about 70aa
- Always the same kinase overall structure
- Five critical Tyr residues (Y951, Y1054, Y1059, Y1175 and Y1214)
- pY1054 and pY1059 are necessary for full activation

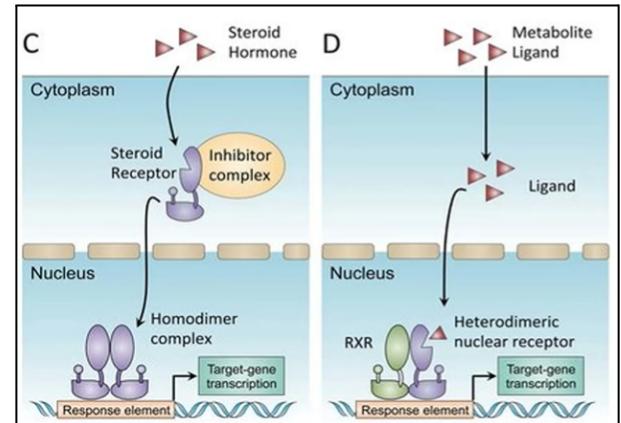
Several **inhibitors have been designed for this receptor**, as it is highly connected with cancers. It could virtually stop the angiogenesis of vessels used by the tumor to grow.

NUCLEAR RECEPTORS

Nuclear Receptors are peculiar, since they act as receptors and transcription factors together.

There are two ways in which nuclear receptor proteins can work:

1. they can be placed in the cytosol and, upon the binding of the ligand, go in the nucleus. Here it will form a homodimeric complex. The ligand here is a **steroid**.
2. the nuclear receptor is already present at the level of the response element in the nucleus, and it becomes activated in site thanks to the ligand, which goes from cytoplasm to nucleus. The ligand here is usually a **metabolite**.



When the receptor is activated, it favors the recruitment of the proteins necessary for active transcription of genes whose promoters contain the responsive element.

This is a **very old way of signaling transduction, where the receptor also transcribes, being able to bind DNA**. This means that, while humans have 48 genes encoding for NRs, in worms have more than 270.

Other characteristics:

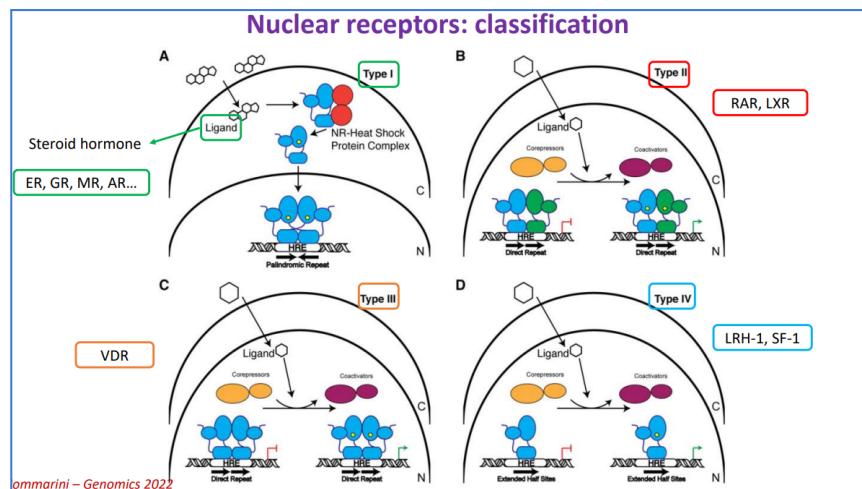
- Important role in a number of biological processes (metabolism, reproduction, inflammation...)
- Regulated endogenously by **small lipophilic** (hydrophobic) **ligands** (steroids, retinoids, phospholipids).
- Some of these receptors are named **orphan receptors**: no ligand yet identified
- Involvement in pathology (cancer, diabetes, chronic inflammation)

There receptors are subclassified in families:

- **FAMILY 0B** includes two members, both orphans and both **exceptions**, as they do not contain DNA binding domains. This means they need to form **heterodimers** with other NRs, otherwise they wouldn't be able to have a role in transcription
- **FAMILY 1** is classified into many different types: they can be thyroid hormone receptors, retinoic receptors and so on. All of these molecules are **hydrophobic and small**.
- **FAMILY 2** contains receptors that can be bound by **fatty acids or retinoids**.
- **FAMILY 3** include estrogen/androgens receptors that correspond to **steroids**

Another way to classify these receptors is according to types:

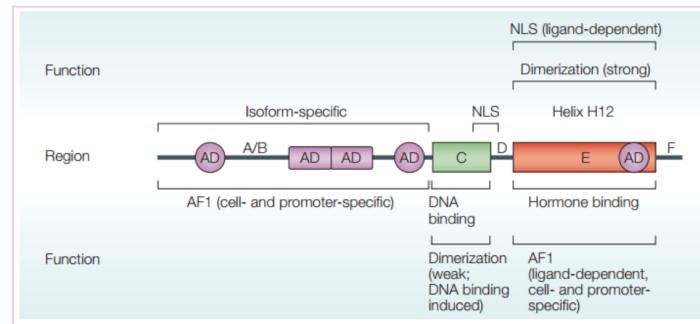
- **type I:** these receptors respond to steroid hormones, the receptor is present in the cytosol in its inactive conformation: upon the binding of the ligand it goes into the nucleus.
Homodimers are created.
- **type II, III, IV** include receptors that are already present in the nucleus.
Type II forms a **heterodimer**, type III forms a **homodimer**, type IV are bound as **monomers**.



1. **Monomers** – just few NRs (LRH-1, NGF1-B, SF-1)
2. **Homodimers** – typical of SRs, with an extensive surface. GR is an exception
3. **Heterodimers** – with RXR, interactions similar to ER

❖ STRUCTURE

Topologically speaking, this is how a nuclear receptor looks: it is a polypeptide with many different functions. Starting from the N term all the way to the C term:

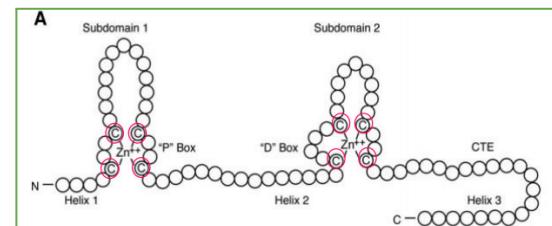


- **activating function 1:** it's **different** among different members. It can be cell specific, tissue specific, isoform specific and so on. (N-term). The diversity is caused by the fact that this **domain has to be specific for different co-regulators.**

We don't have any structures available for the N term: this is due to the high variability. Moreover, AF1 is subject to post translational modifications.

- **DNA binding site** is instead very **conserved**. Since it must bind different responsive elements, the binding site can have differences in the amino acids, which **provide specificity**. This ensures that the receptor can **bind to many different elements**.

DNA binding domain contains 2 subdomains, containing 4Cys coordinating zinc atoms (**canonical Zn-finger motif**) followed an amphipathic α -helix and a loop



- **activating function 2:** it binds the **ligand**. Structural changes that begin here are propagated through the whole receptor. It's **conserved**. AF2 is another side which allows **binding to co-regulators**, which can be either activators or inhibitors.

As we said, the overall structure is similar, although there are some differences for recognition of different ligands.

It's made up of 11 α -helices and 4 β -strands. The helices are placed in three different layers, which generate a **hydrophobic pocket for ligand binding**, which stabilizes the structure. In particular, helix 12 can assume **multiple conformations** upon ligand binding, which allows interaction with **different co-regulators**. This means that, together with helix 12, also the pocket can assume many different conformations.

The structure is mainly hydrophobic. Nevertheless, there are some **hydrophilic residues** to allow **specificity**.

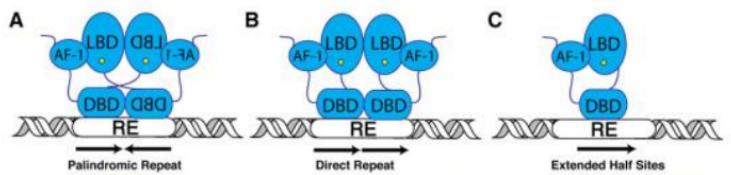
Between the DNA binding domain and the ligand binding domain we have a hinge domain.

- **hinge domains** are conserved, but their length can vary. Hinge domain is important because it contains the **nuclear localization sequence (NLS)**, which is necessary to find the receptor in the nucleus.

❖ TARGET SEQUENCES

Nuclear receptors can recognise certain DNA sequences through their DNA binding domain. Examples are:

- **PALINDROMES**: contain two opposite repeats separated by a spacer region that varies in length (typically 3bp). They are recognised by a homodimeric receptor. → *estrogen receptor*
- **DIRECT REPEATS**: bound by heterodimers separated by a spacer sequence (0- 5bp)
- **EXTENDED MONOMERIC SEQUENCES** have an A/T rich sequence directly upstream and are recognised by a monomer.



❖ ACTIVATION MECHANISM

The activation mechanism is an **allosteric one**.

- **Ligand binding** induces **conformational changes** within the receptor, which in turn **binds specific DNA sequences** throughout the genome (responsive elements).
 - ligand binding domain and DNA binding domain are separate: the first induces the other

Upon DNA binding **they recruit co-regulator proteins**, chromatin remodelers **and the transcription machinery** to activate or repress target gene expression.

1. **Helix 12 controls the entrance of the ligand** binding pocket
2. The orientation of Helix 12 is determined by allosteric **changes induced by the ligand** binding. The fixing of H12 leads to formation of AF2.
3. It's H12 which controls the **recruitment of co-regulators** within its surface.

The first mechanism that was proposed is the **MOUSETRAP MECHANISM**.

- some elements of the receptor do not change position depending on the presence/absence of a ligand, while others do. We have **helices that completely change their position** depending on whether the ligand is bound or not, being either "open" or "close"

This model is not actually the most common one, being true for only certain receptors like the retinoid one.

- In reality, most of the time, H12 isn't present in one conformation or the other, it is rather **highly flexible and can be in many different configurations**. This mechanism is called **DYNAMIC STABILIZATION MODEL**.

Independent of the mechanism, **HELIX 12 IS CENTRAL FOR THE GENERATION OF A SURFACE BETWEEN RECEPTOR AND CO-REGULATORS**. This is done thanks to ligand binding the receptor in AF2 and allowing conformational changes.

❖ CO-REGULATORS

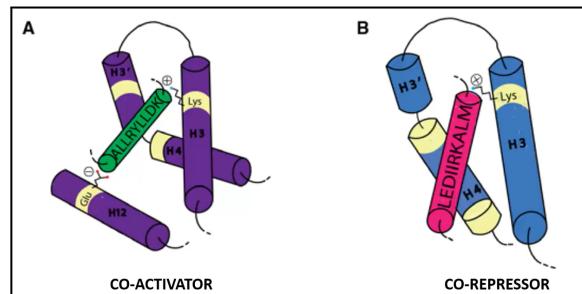
When the **receptor is activated** by the ligand, H12 is placed in such a way that it, and other helices, recognise **co-activators** by hydrophilic residues Glutamates and the Lysins.

Leu residues of the motif lie within the hydrophobic groove of the AF-2 surface and the helix boundaries are bound by charged clamps (Glu and Lys)

Co-activators are recognized on their α -helix containing the **NR motif**.

The take home message is that interaction between nuclear receptors and activators is mediated by both **hydrophobic and hydrophilic** residues.

When the **receptor is not active**, H3 and H4 are making another structure which is missing H2, that can bind **co-repressors**.



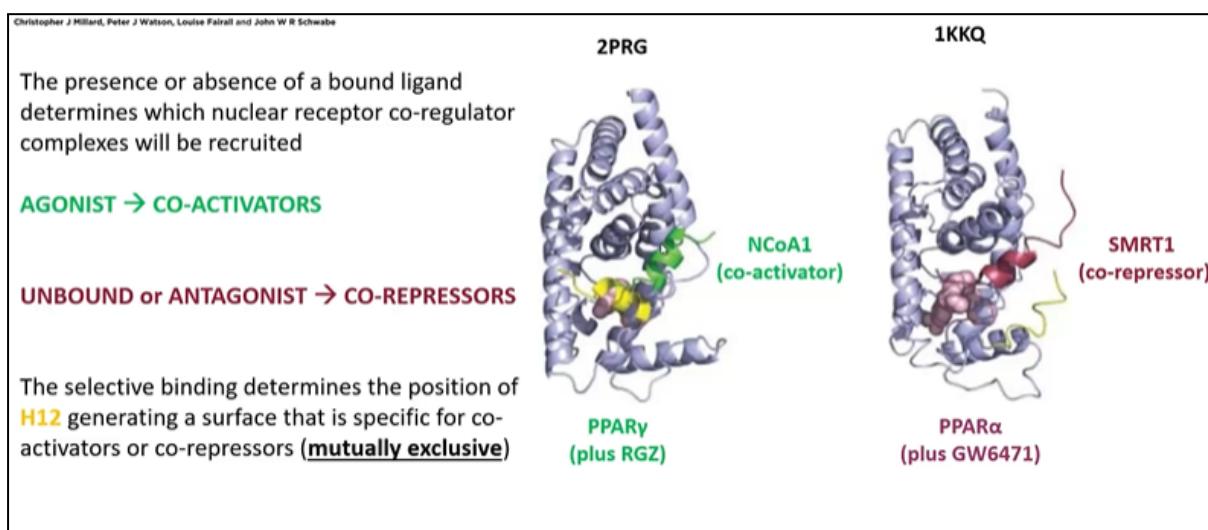
Co-repressors present the conserved motif known as ***CoRNR box***.

In general, **NR are able to recruit co-regulators**. These proteins physically interact with transcription factors and other components of the transcriptional machinery, while not being transcription factors themselves. They can be of course subdivided into co-repressors and co-activators. More than 300 nuclear co-regulators have been identified.

→ **Co-regulators are not directly recruited** to specific loci in the genome, they work associated with transcription factors (like nuclear receptors).

Here we have a comparison between two very similar NRs (PPAR γ e PPAR α), one interacting with a coactivator(RGZ) and one with a co-repressor (GW6471): remember that co-regulators are not ligands, so they don't have to be proteins: they can be fatty acids too, and so on.

- **co-repressors and co-activators are mutually exclusive.**



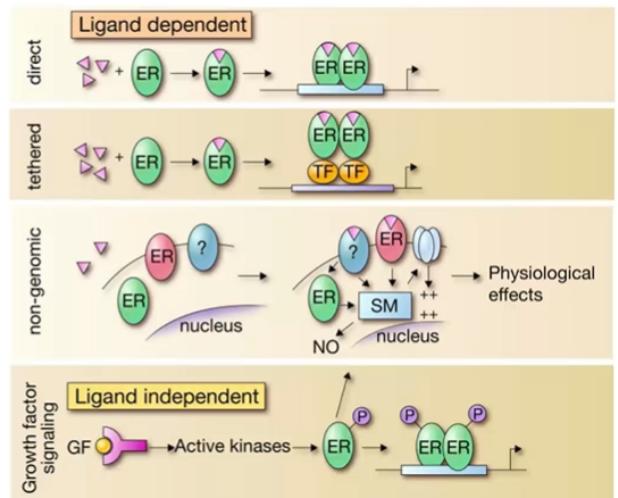
❖ ESTROGEN RECEPTORS

Estrogen receptors are a big family of receptors, able to bind estrogen-derived hormones. There can be different families of estrogen receptors: they can be **GPCRs** but also **NRs**.

- Estrogen **nuclear receptors** are ER α and ER β :
 - ◆ Intracellular or associated to PM
 - ◆ form *homodimers* or *heterodimers* with transcriptional activity
 - ◆ Other mechanism of action
- Estrogen **GPCRs**:
 - ◆ Associated to PM
 - ◆ Provides part of the non-genomic ER response

There are four ways to mediate the signaling given by hormone levels: some are **ligand dependent** and some are **ligand independent**.

- ★ the **estrogen binds to the ligand and is then translocated** into the nucleus
- ★ we have a **tethered** way, where the ligand is still bound to the nucleus. However in this case **the estrogen is not directly bound to the ligand**
- ★ **non genomic regulation** is also ligand dependent: here **the binding of the ligand induces the recruitment of other molecules**
- ★ **growth factor signaling** instead depends on nuclear receptors responding to the presence of estrogen.



Without estrogen signaling, there are a series of disorders that are much more probable. Likewise, the overstimulation of estrogen signaling is also dangerous.

- it is important to create drugs that **fix the amount of estrogen receptor signaling**

Lack of estrogen signaling

- Cardiovascular disorders
- Reduced bone density
- Menopause symptoms

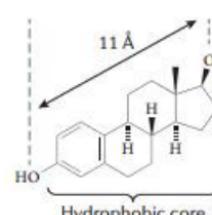
Overstimulation of estrogen signaling

- Cancer (breast, ovarian, endometrial, prostate, lung)
- Increased cell proliferation and invasiveness (EMT)

To control the amount of signaling made by the receptor, we take control of the hydrophobic ligand binding pocket houses two polar regions located at the opposite ends of the cavity.

- **Glu353 and Arg394** are anchoring points for the **OH group in position 3** of E2 (necessary for affinity)
- A water molecule stabilizes the interaction
- **His524** interacts with the **OH group at position 17** (important for specificity)

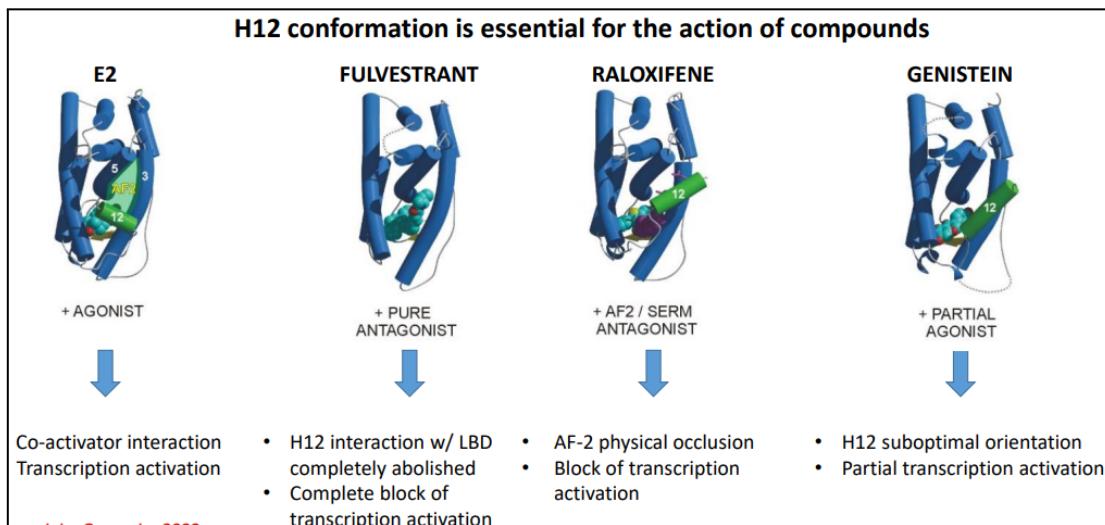
The common core of the estrogen receptor contains two OH groups that we have just mentioned. for specificity and affinity.



- Flat ligands preferentially bind to **ER β** to **ER α** , activating slightly different signaling pathways.

Changing the interactions with OH 17 means having a positive or a negative regulation of transcription, because OH 17 changes the position of helix 12.

- **Bulky side chains** can be added to this part of the ligand to make an antagonistic compound, **blocking the receptor**.



We can see the changes to H12 we described above.