

# **Exploration Analyse de Réseaux de Contacts entre Protéines et d'Arbres Phylogénétiques**

**Institut d'Informatique d'Entreprise - Conservatoire National des Arts et Métiers  
Institut PASTEUR - Unité d'Immunogénétique Cellulaire  
Perrine BARJOU**

# INTRODUCTION

*Contexte Biologique*

*données astronomiques*

*comment les exploiter*

*trouver leurs fonctions*

*comment observer ces fonctions*

*=> Interactions protéines protéines*

# INTRODUCTION

Donc développer une application pour:

**à partir de**

la séquence d'une protéine et le nom de l'organisme



- Proposer une liste de partenaires structuraux ou fonctionnels de cette protéine
- Identifier les résidus aux interfaces
- Visualiser graphiquement les réseaux obtenus

# INTRODUCTION

Un projet à intégrer dans la  
continuité de deux autres

# PLAN

- Les Interactions Protéines Protéines:
- Construction de la base de données proteinDB
  - Problématique
  - Solution
- La visualisation graphique
  - Cytoscape : pour la 2D
  - PyMOL : pour la 3D

*rappel des outils utilisés????*
- Une application : reconnaissance de l'IL2 par son récepteur

# **INTERACTIONS PROTEINES PROTEINES**

**Une interaction entre deux  
protéines partenaires potentiels  
peut s'observer:**

- soit expérimentalement**
- soit de manière prédictive**

# Déterminer un interaction

## Expérimentalement

répertoriées dans différentes banques de données

- DIP
- MINT
- IntAct
- HPID
- BIND

etc...

## Par Prédiction

### liens STRUCTURAUX

- Projections de paires par BLAST
- Fusion de gènes (algorithme de Rosetta Stone)
- Compensation évolutive d'interface
- Domaine Pairs/ Shuffling

### liens FONCTIONNELS

- Profils Phylogénétiques
- Gene Clustering
- Voies Métaboliques
- Les liens bibliographiques

# Les Interactions Expérimentales

Exemples:

- coprécipitation, cocrystalisation**
- étude biochimique**
- double-hybride**

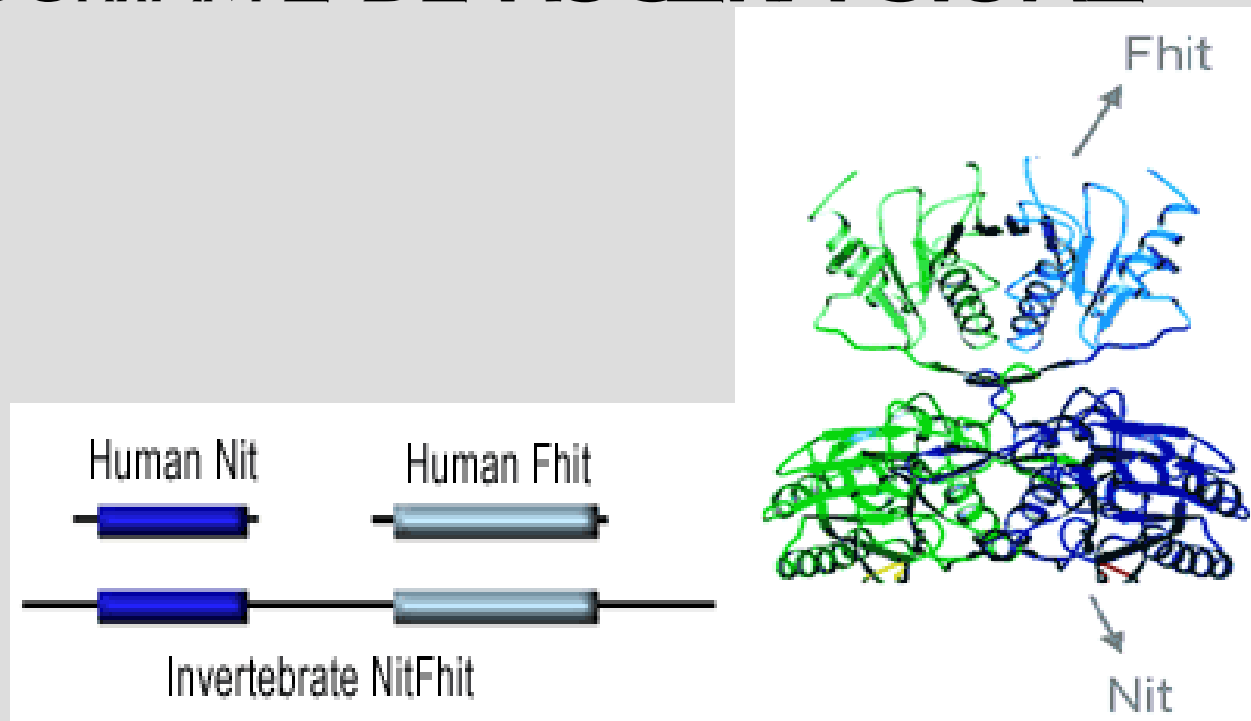
Stockées dans des banques de données accessibles par le WEB:

- DIP** : Database of Interacting Protein
- BIND**: Biomolecular Interaction Network Database
- MINT**: Molecular Interaction database



# Exemple d'interactions structurales prédites (1/2)

L'ALGORITHME DE ROSETTA STONE

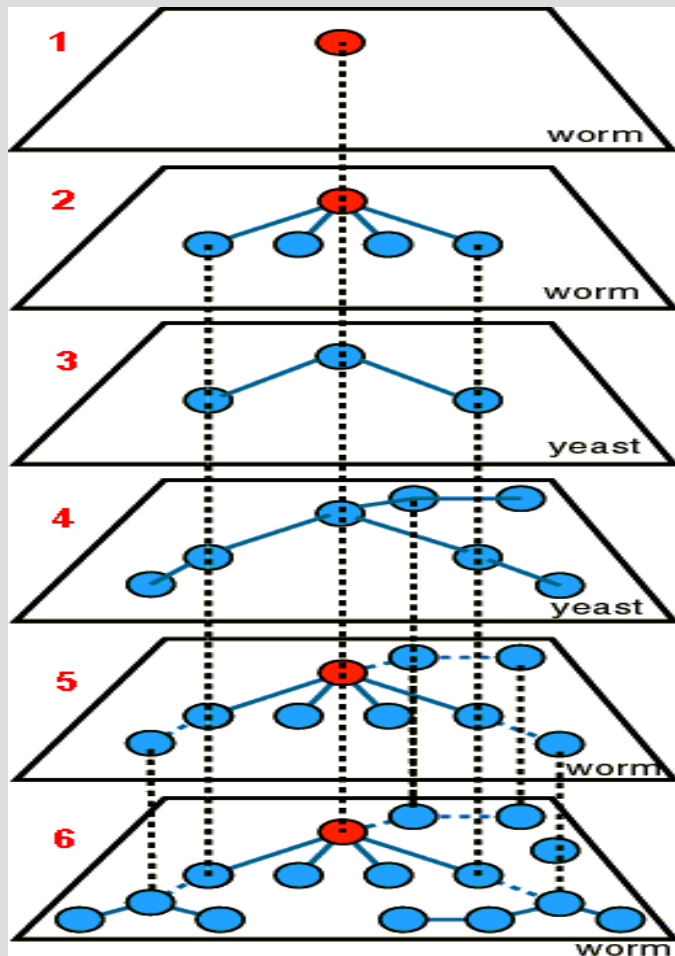


Identifier les domaines fusionnés ou séparés durant l'évolution

# Exemple d'interactions structurales prédites (2/2)

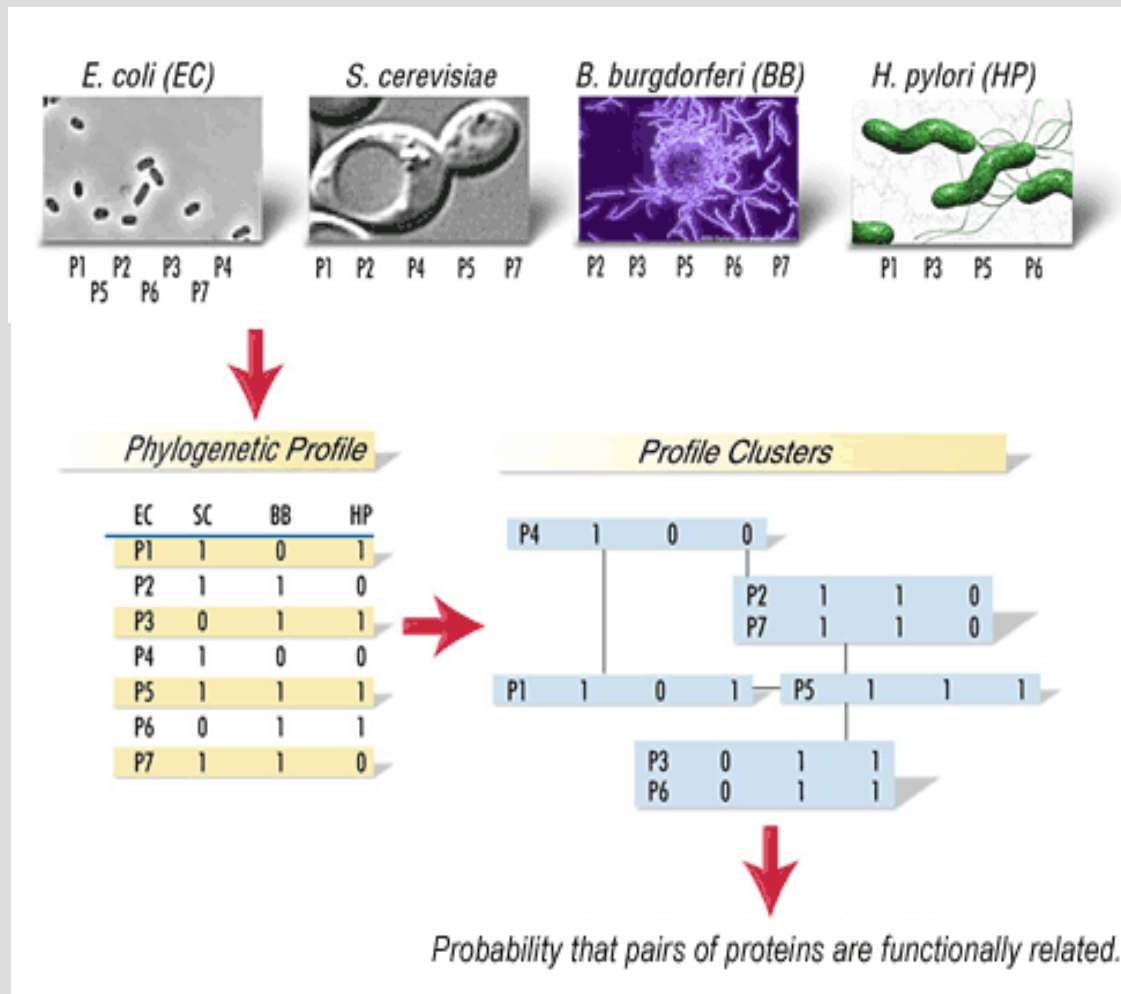
Compensation évolutive d'interface

*expliquer*



# Exemple de liens fonctionnels prédits

## LES PROFILS PHYLOGÉNÉTIQUES



IDENTIFIÉES  
GROUPES  
FONCTIONNELS DE  
PROTÉINES  
ENCODÉES POUR  
DES FONCTIONS  
PHYSIOLOGIQUES  
SPÉCIFIQUES

# Travaux Réalisés

## Ce que l'on veut:

- Pouvoir interroger une source de données sure, non redondante, avec des séquences )  
interactions ) *identifiées de manière homogène*  
graphes )
- Pouvoir visualiser des réseaux protéiques en 2D puis en 3D

# PLAN

## Les Interactions Protéines Protéines:

### Construction de la base de données proteinDB

- Problématique
- Solution

### La visualisation graphique

- Cytoscape : pour la 2D
- PyMOL : pour la 3D
- rappel des outils utilisés????

Une application : reconnaissance de l'IL2 par son récepteur

# LA BASE DE DONNEES

## BUT:

Pouvoir interroger les données sans relancer de nombreux calculs.  
Avoir des données de séquences protéiques identifiées de façon homogène

## PROBLEMATIQUE:

- Banques de Séquences HETEROGENES
  - GenBank                      gi                      (~2 millions de séquences)
  - SwissProt                      sp                      (~150 000 séquences)
  - Autre Banques    id aléatoire
- une meme séquence peut avoir plusieurs **id**, et un **id** peut correspondre à plusieurs protéines

**ET IL N'EXISTE AUCUN LIEN CLAIR  
ENTRE CHACUNE DES BANQUES**

# LA BASE DE DONNEES

## Solution:

CONSTRUIRE UNE BASE DE DONNÉES EN LOCAL

## Dans la base:

LES TABLES PRINCIPALES

- TABLE 1 (PROTEINS):

TOUTES LES SEQUENCES PROTÉIQUES CRÉES UNE SEULE FOIS ET  
POUR LESQUELLES JE RECRÉE UN IDENTIFICATEUR UNIQUE

- TABLE 2 (INTERACTIONS):

TOUTES LES INTERACTIONS PROTÉINE-PROTÉINES

- TABLE 3 (GRAPHS)

TOUS LES IDENTIFICATEURS CONSTITUANTS LES NOEUDS DES GRAPHS  
ET LES CEUX CONSTITUANTS LES LIENS ENTRE CES NOEUDS

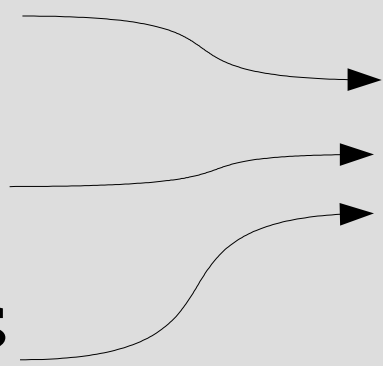
# La BDD : la table proteins

## SOURCES

NRPROT

SWISSPROT

AUTRES BANQUES



## BDD

Table 1:  
les séquences  
protéiques  
avec toutes leurs  
informations

avec une règle:

1 séquence = 1 organisme

# BDD: la table interactions

Listes d'interactions expérimentales  
(DIP, MINT, BIND, IntACT...)

Listes de liens prédits  
(Structuraux: ROSETTA, ect..)  
(Fonctionnels: Profils Phylogénétiques, etc...)

Table 2:  
les listes  
d'interactions  
protéines-protéines

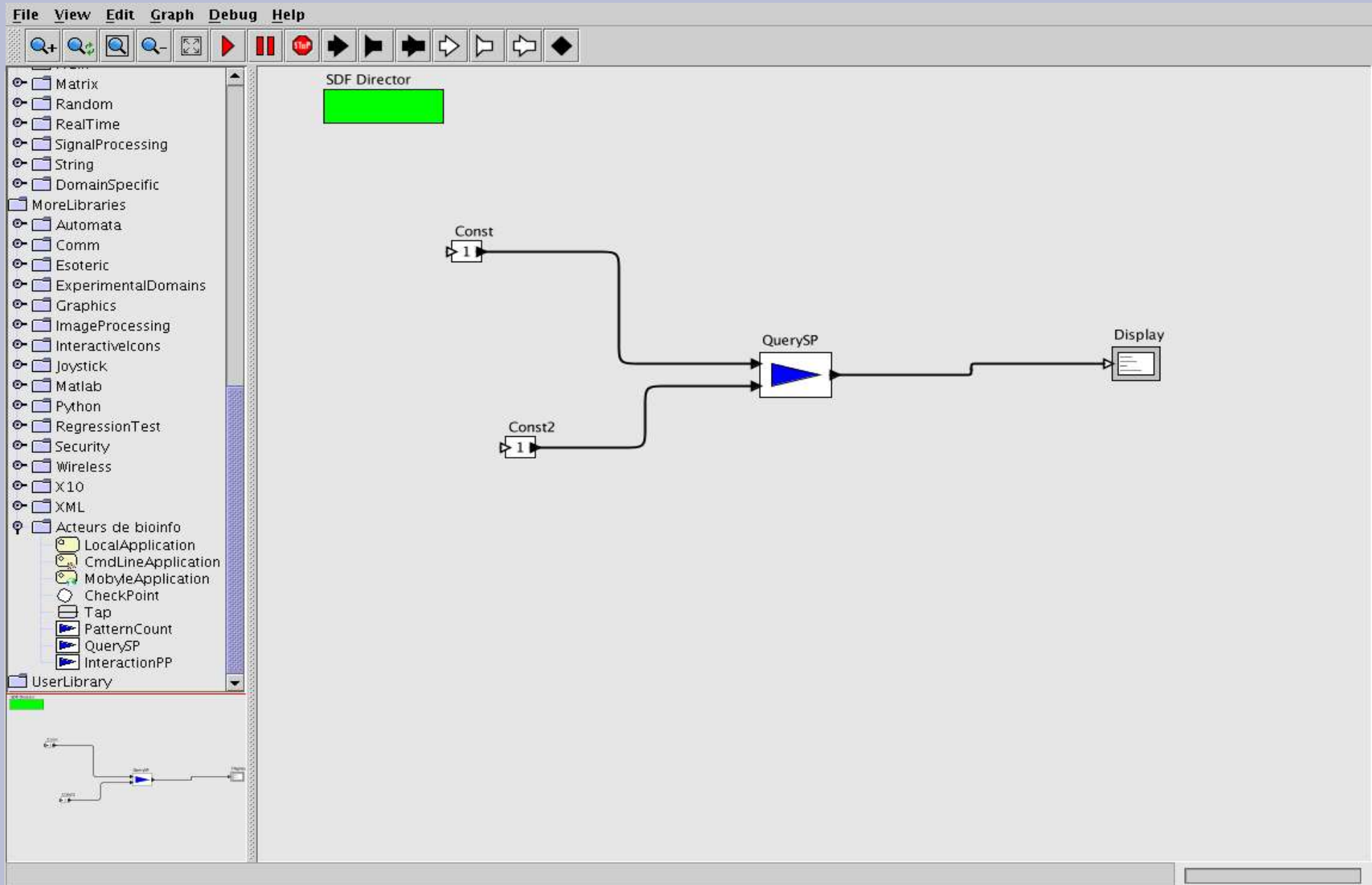
Mettre les règles

# BDD: la table graphs

Table 3:  
les listes de noeuds  
et de liens pour  
dessiner un graphe

Détailler et mettre les règles

# PTOLEMY: outil d'interrogation et maintenance de la base



# PLAN

## Les Interactions Protéines Protéines:

### Construction de la base de données proteinDB

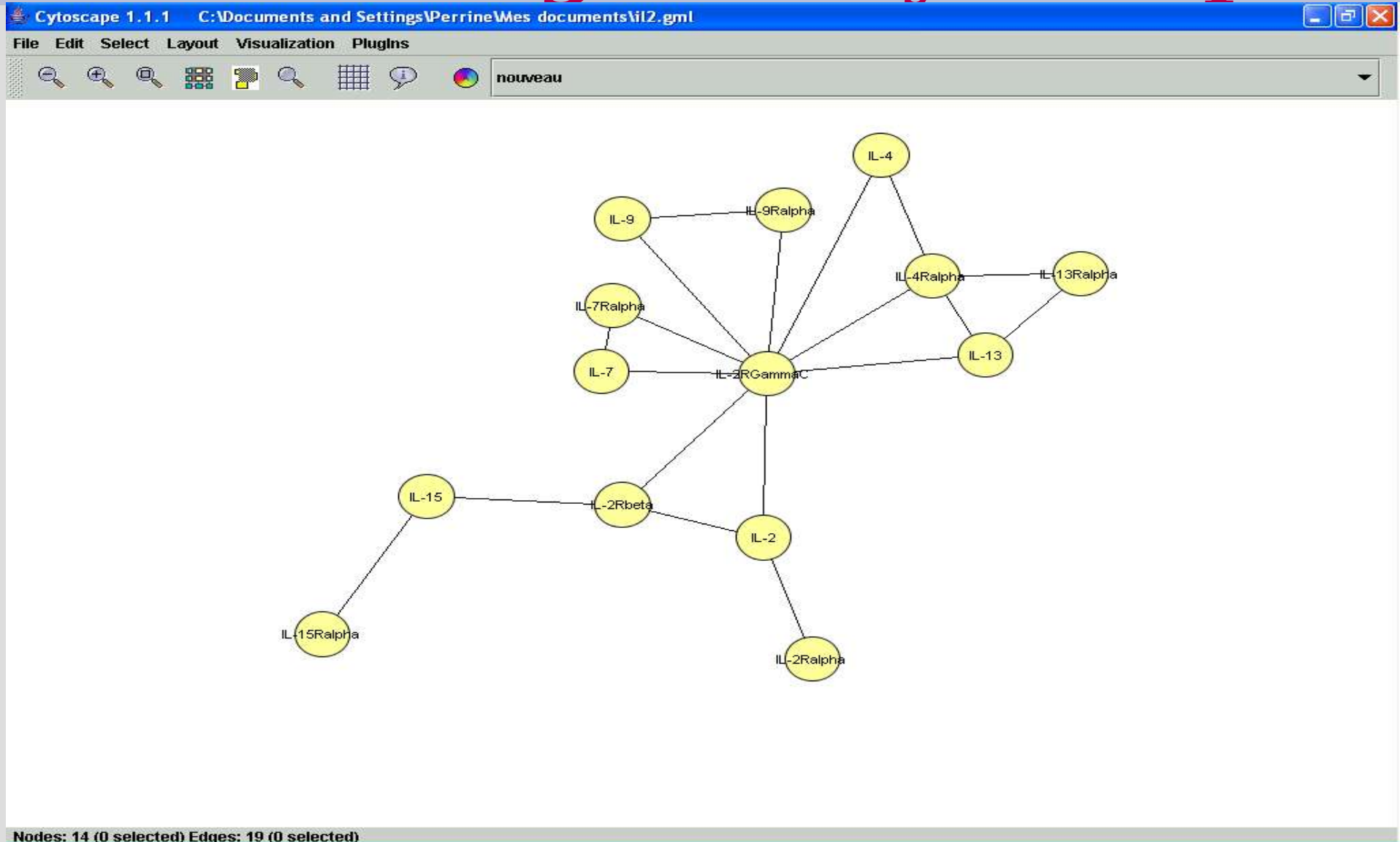
- Problématique
- Solution

### La visualisation graphique

- Cytoscape : pour la 2D
- PyMOL : pour la 3D
- rappel des outils utilisés????

Une application : reconnaissance de l'IL2 par son récepteur

# Choix d'un visualisateur 2D : le logiciel Cytoscape



# Visualisateur pour la 3D: PyMol

Une capture Pymol

# PLAN

**Les Interactions Protéines Protéines:**

**Construction de la base de données  
proteinDB**

- Problématique
- Solution

**La visualisation graphique**

- Cytoscape : pour la 2D
- PyMOL : pour la 3D
- rappel des outils utilisés????

**Une application : reconnaissance de l'IL2 par  
son récepteur**

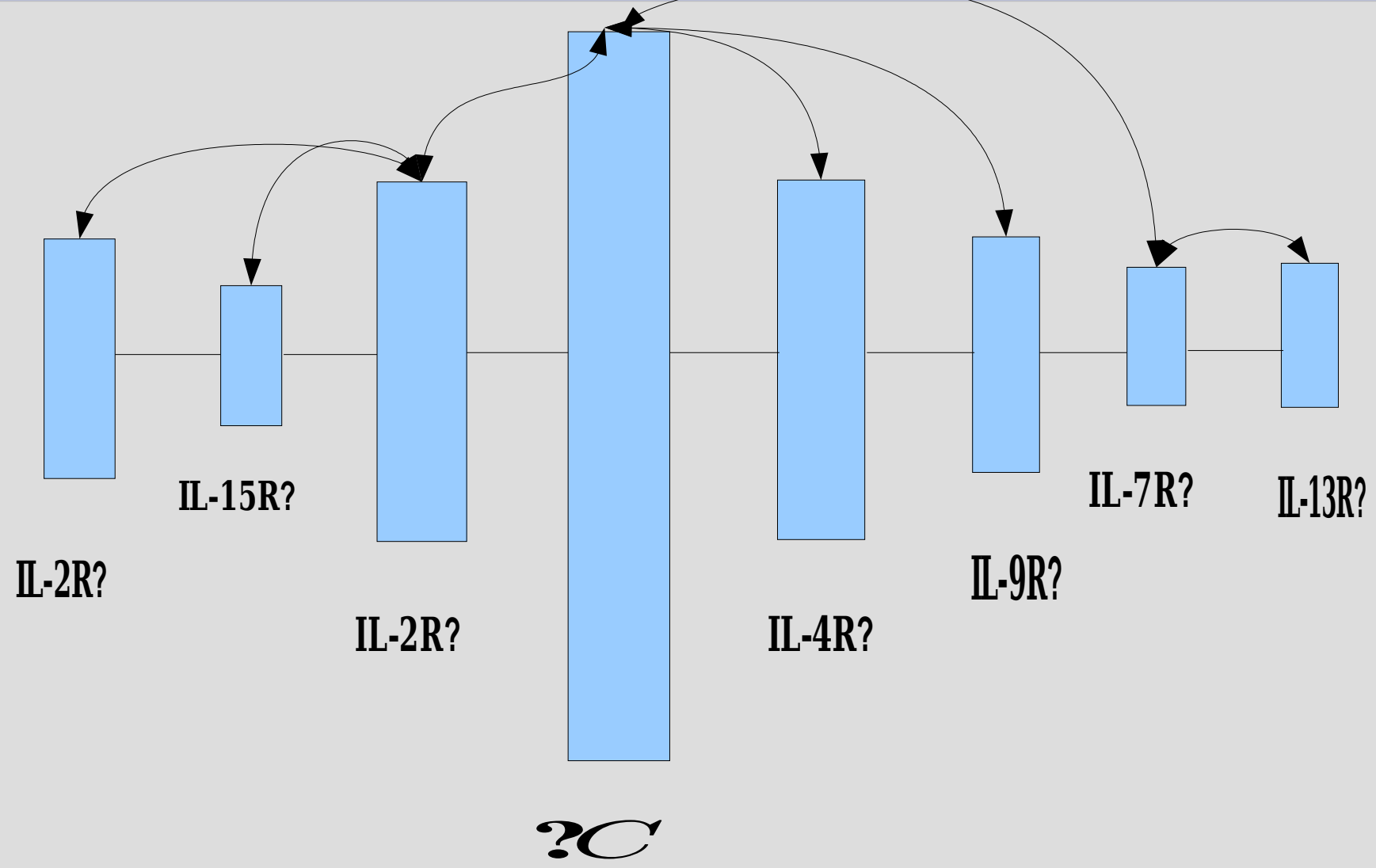
# Application : la cytokine IL2

Où elle intervient

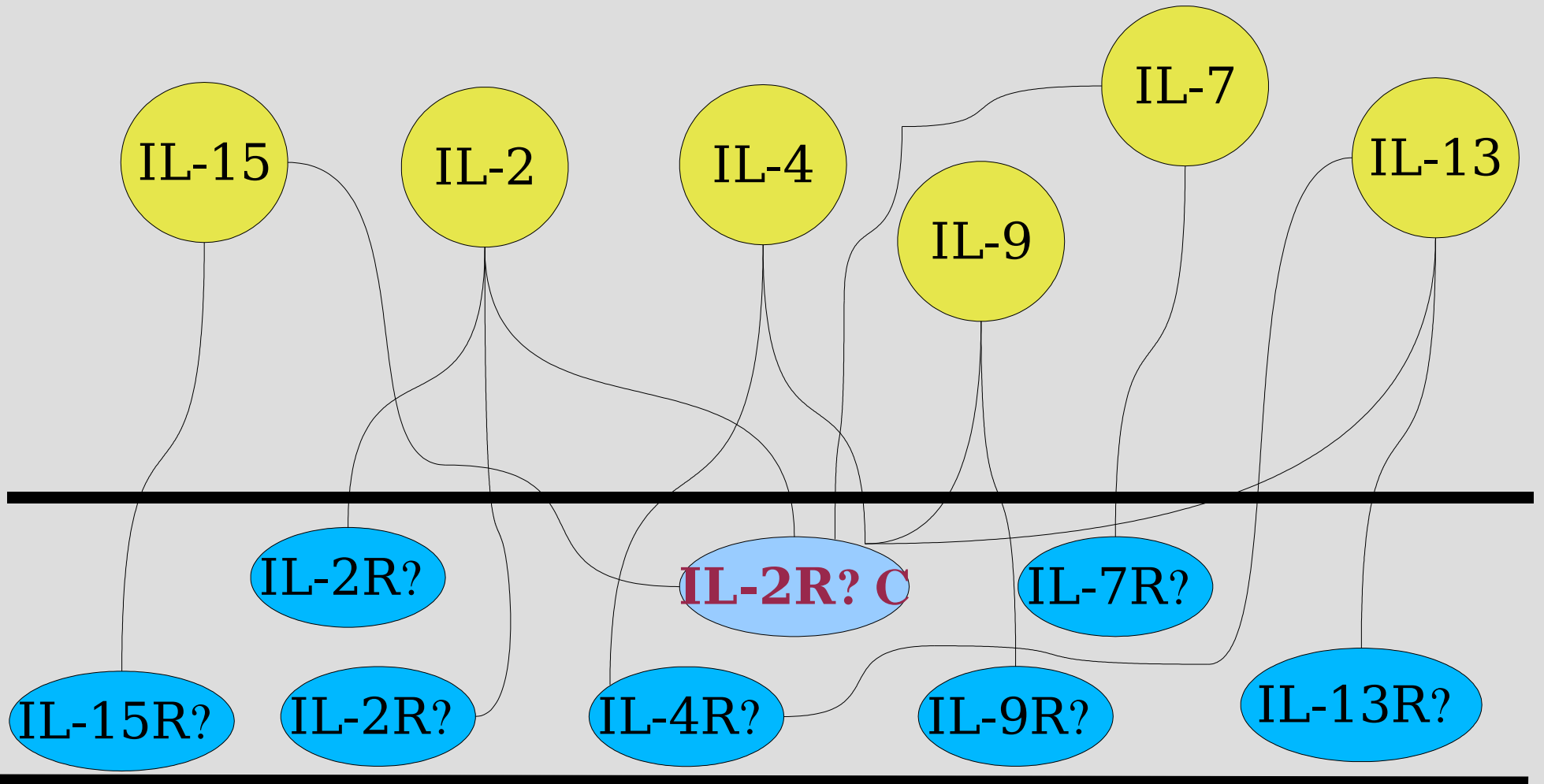
Les recherches réalisées dessus

# Application : Interaction

## IL2 - IL2 R



# Application : Interaction IL2 - IL2 R

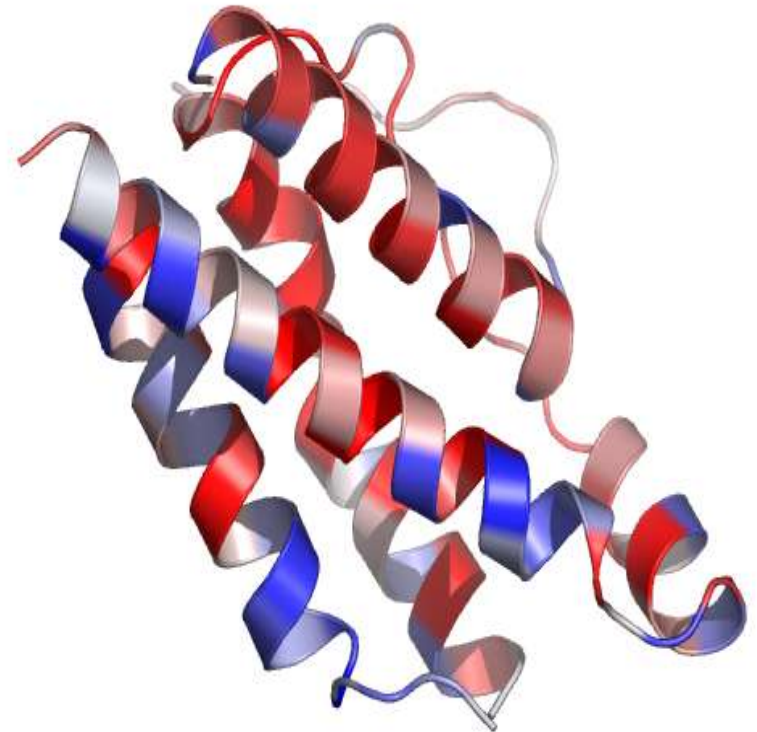
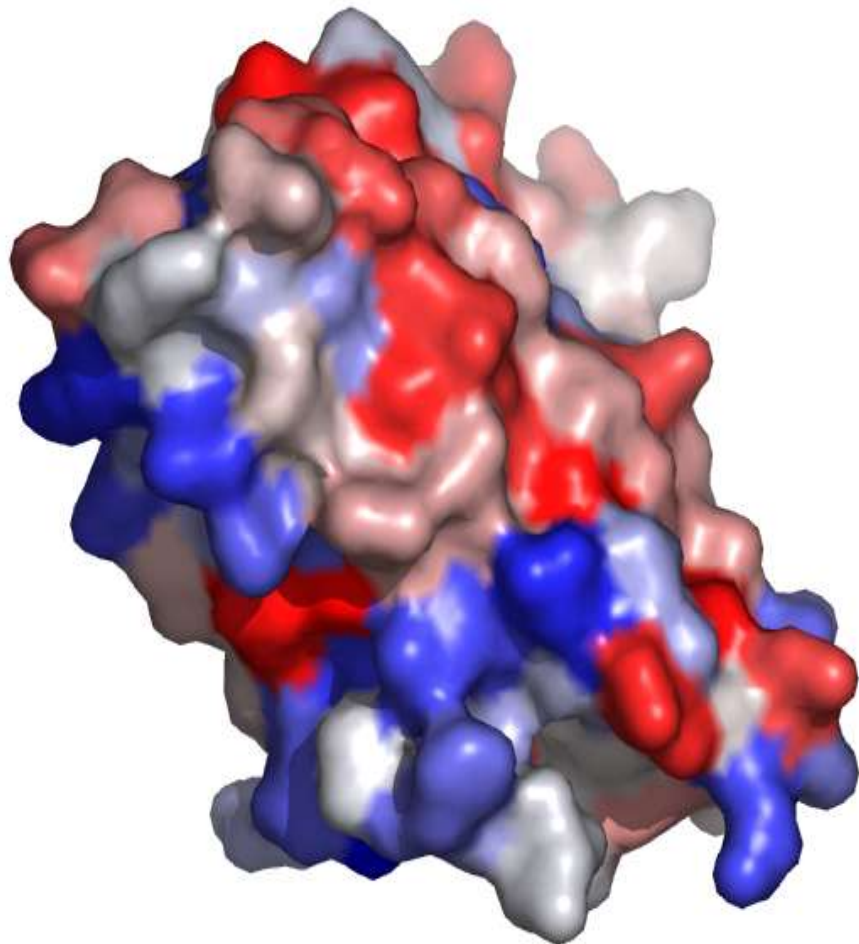


# Application : reconnaissance de l'IL2 par son récepteur

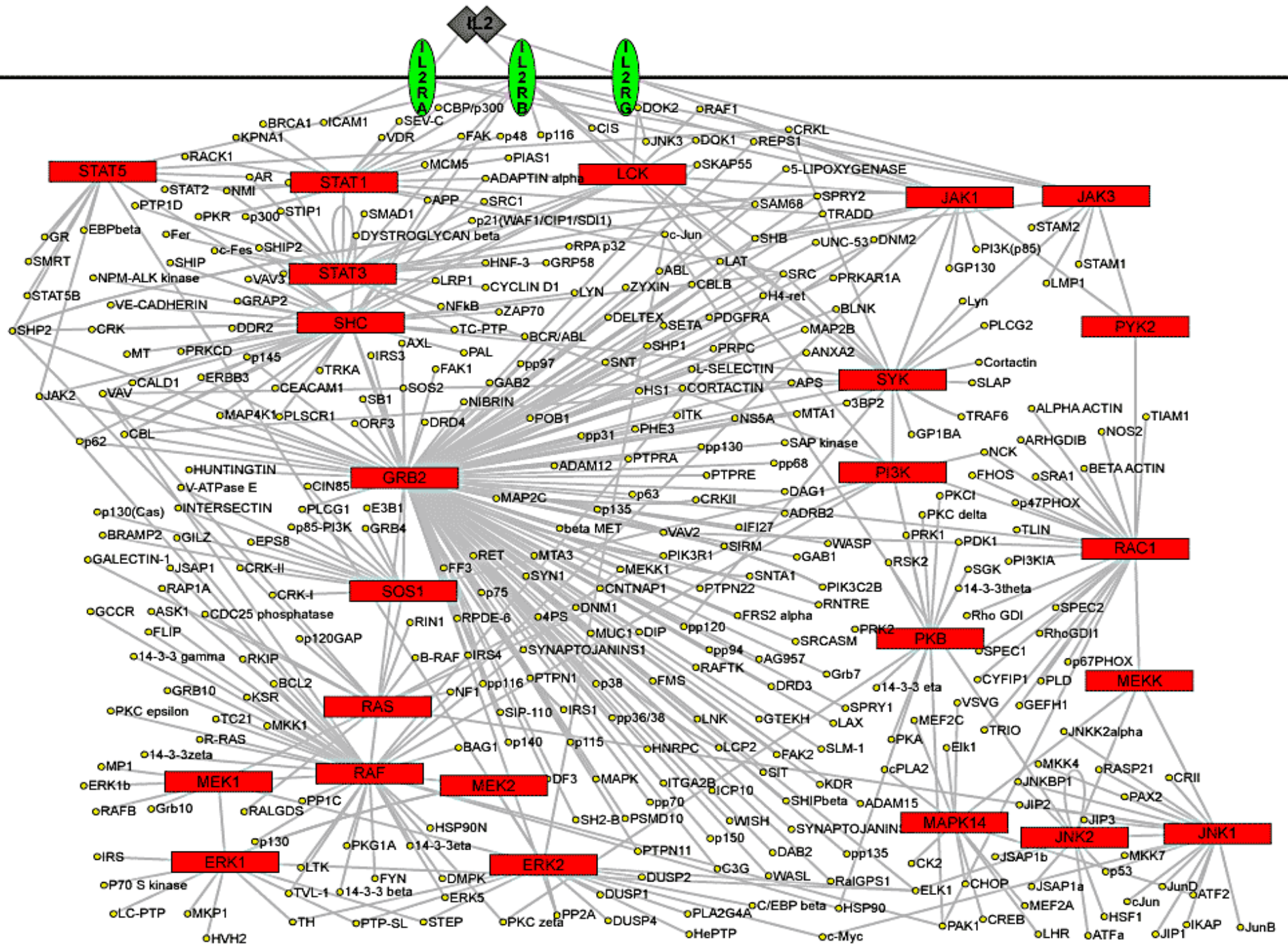
- Recherche des séquences orthologues
- Alignements multiples
- Réseaux d'interactions
- Arbres phylogénétiques
- Recherche des co-évolutions de positions
- Figurer les réseaux

## 2. Alignement Multiple

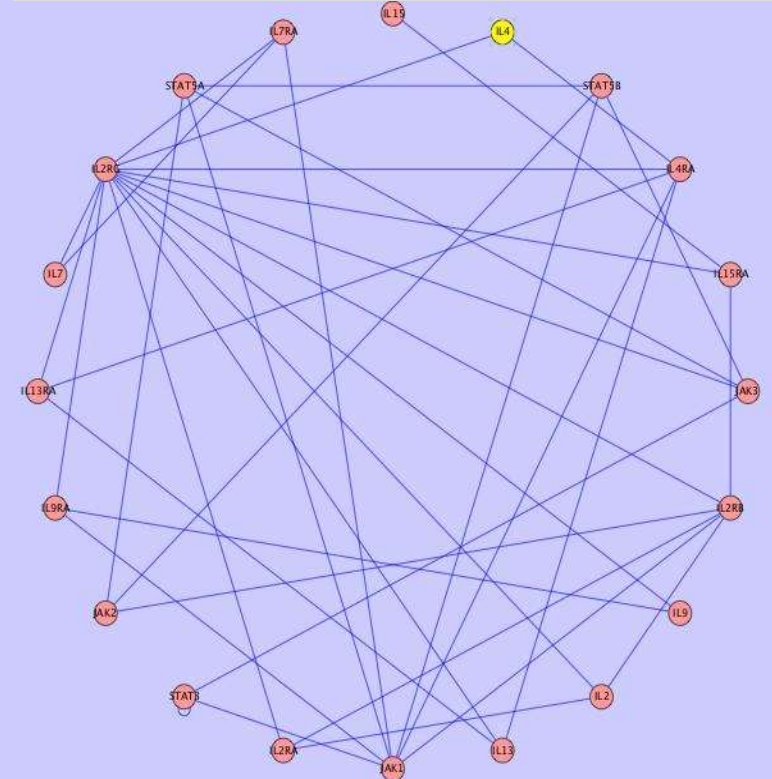
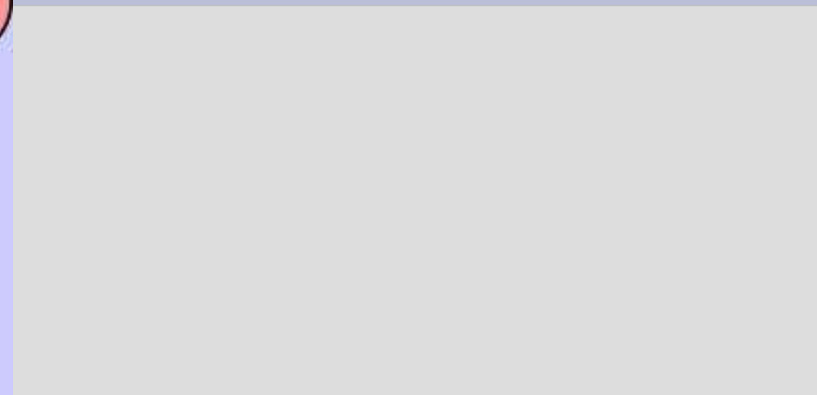
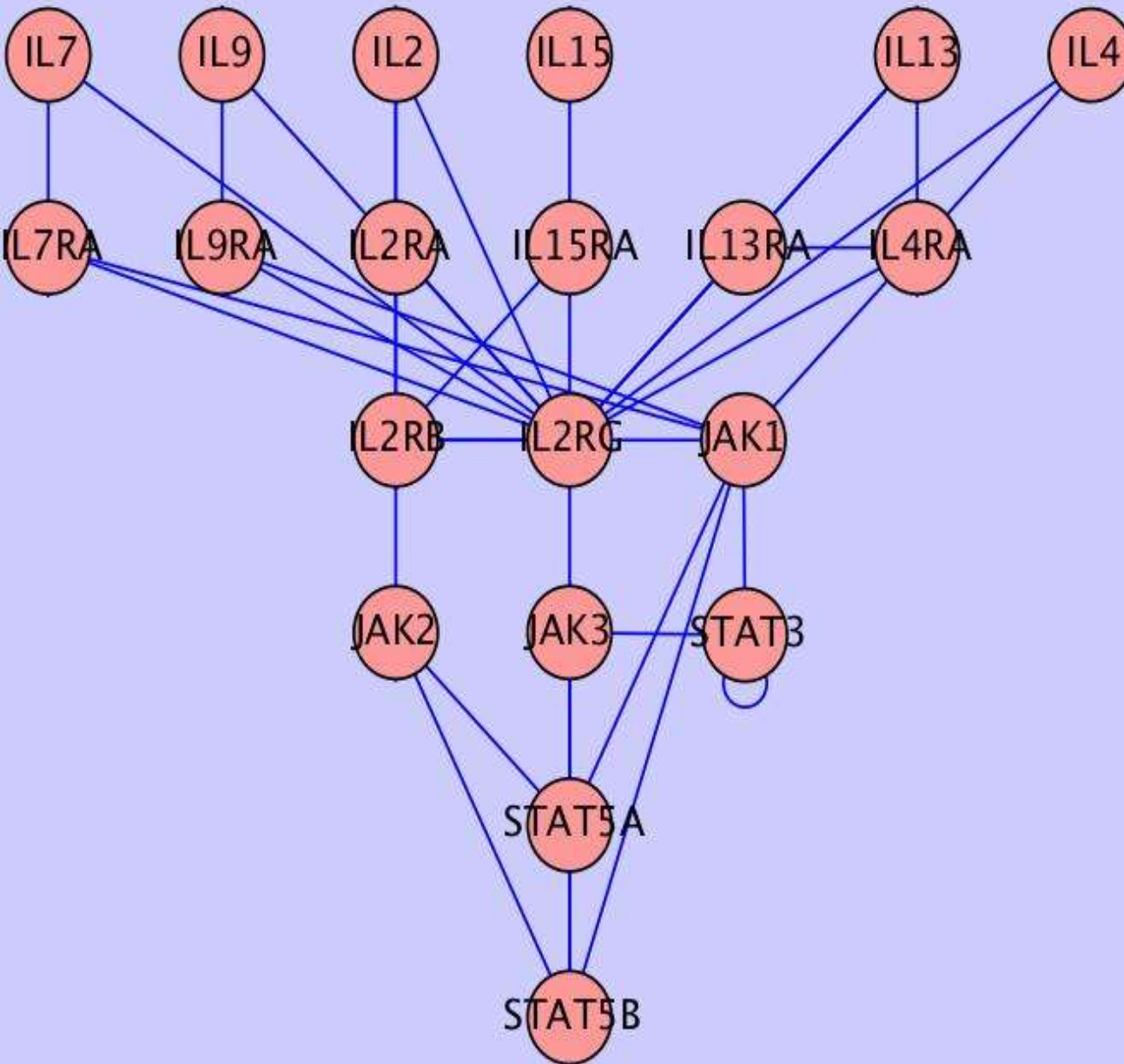
**Conservation projetée sur la structure cristalline de l'IL2**



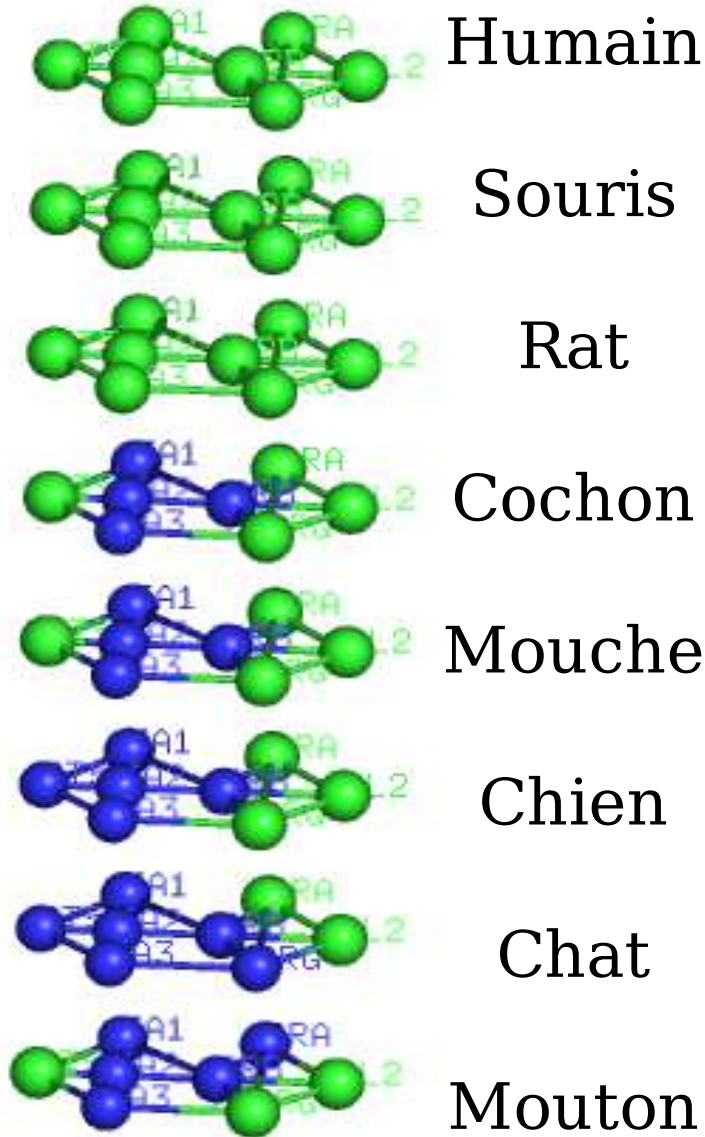
# 3. Réseaux d'interactions



# Sous Cytoscape



# Réseau d'interactions multicouche



# APPLICATION À LA RECONNAISSANCE DE L'IL2 PAR SON RÉCEPTEUR:

Élargissement du séquençage aux autres cytokines de la famille des  
hématopoïétines

IL2, IL4, IL6, IL7, IL9, IL13, IL15, EPO, GH, IL3, IL5, prolactine

SCHEMA DES DIFFÉRENTS COMPLEXES CYTOKINE-RÉCEPTEUR

REPRÉSENTATION DES COMPLEXES

REPRÉSENTATION DES RÉSEAUX

POSITIONS CONSERVÉES PAR GROUPES ORTHOLOGUES

ENTRE IL2, IL4, IL6, IL7, IL9, IL13, IL15

VARIATION DE RÉSIDUS ENTRE LES GROUPES

ORTHOLOGUES IL2, IL4, IL6, IL7, IL9, IL13, IL15

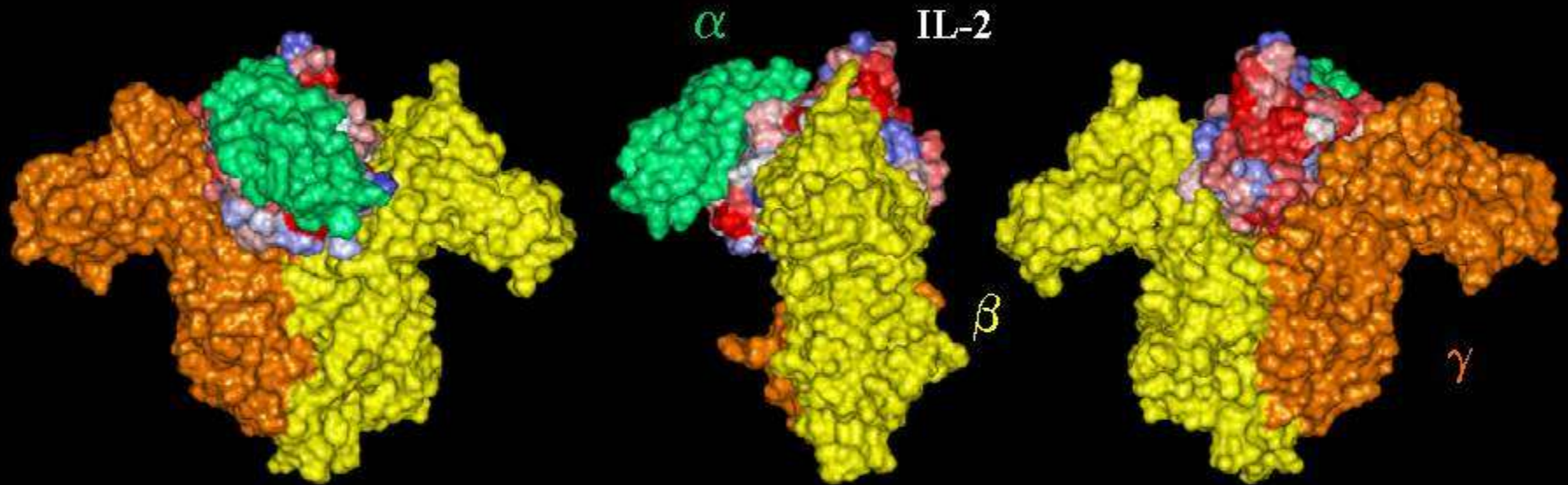
AUX POSITIONS CONSERVÉES

ANALYSE DES ÉCHANGES ACIDES BASES ENTRE CHAÎNES

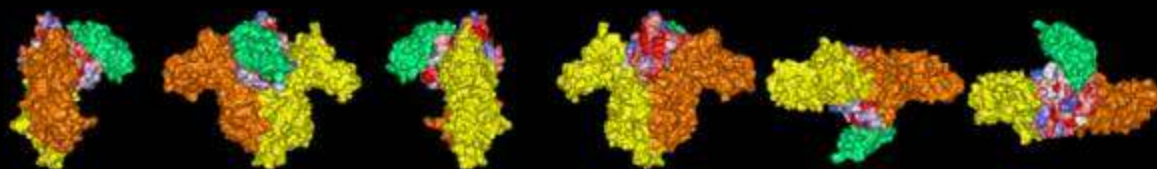
AUX INTERFACES D'UNE ESPÈCE À L'AUTRE :

COMPENSATION DE MUTATION DURANT L'ÉVOLUTION

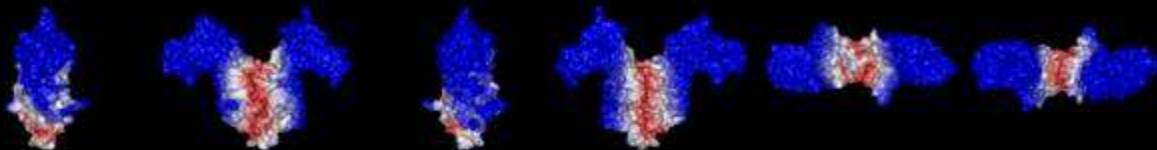
# Interface IL-2/IL2R $\alpha\beta\gamma$



IL-2R $\alpha\beta\gamma$



IL-2R $\beta\gamma$

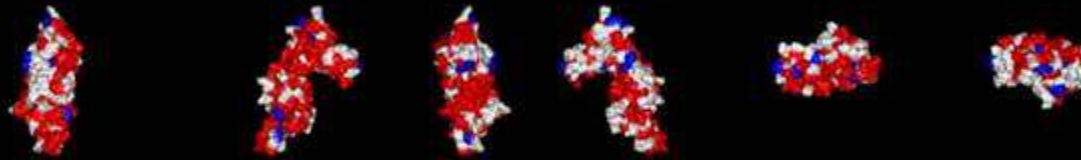


IL-2R $\beta$

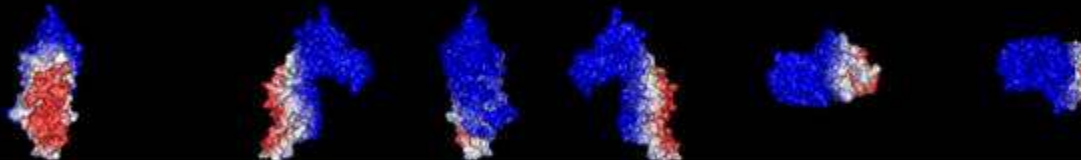
proprietes



conservation



contact



IL-2R $\gamma$

proprietes



conservation



contact

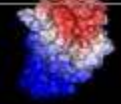
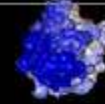
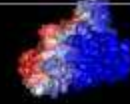
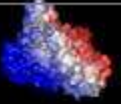


# Interface IL-2/IL2R $\alpha\beta\gamma$

propriétés



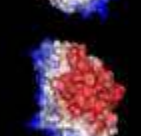
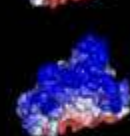
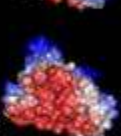
$\alpha$



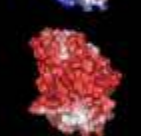
$\beta$



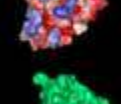
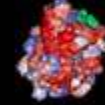
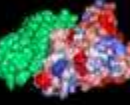
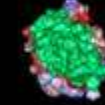
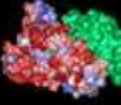
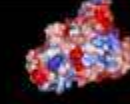
$\gamma$



$\alpha\beta\gamma$



conservation

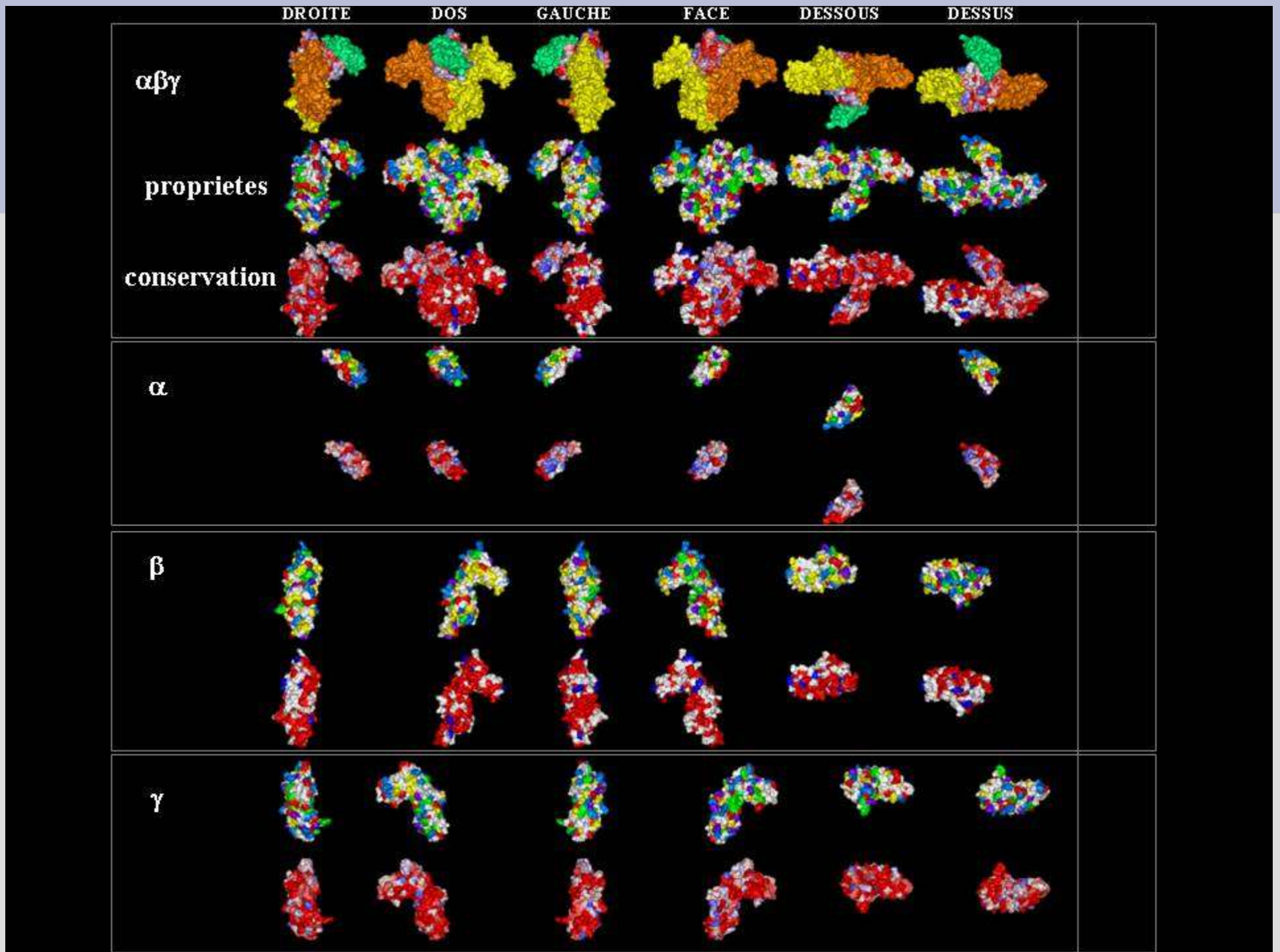


IL-2

$\beta$

$\gamma$

$\alpha$



## Contacts électrostatiques dans IL2-IL2Rabg:

### Interface IL2 et IL2Ra

2:K35 (1ex)

2:R38

2:K43 (2ex)

2:K64

A: E9

A: D6

A: D4 (1/2ex)

A: D56

human K-D

bovine D-H

### Interface IL2 et IL2Rb

2:E95 (15ex)

2:D220

B: K71, R91 (1/2 ex) human E-K

B: R15

mouse K-S

### Interface IL2 et IL2Rb

2:K110 (2ex)

2:K116

G: K36

G: K91

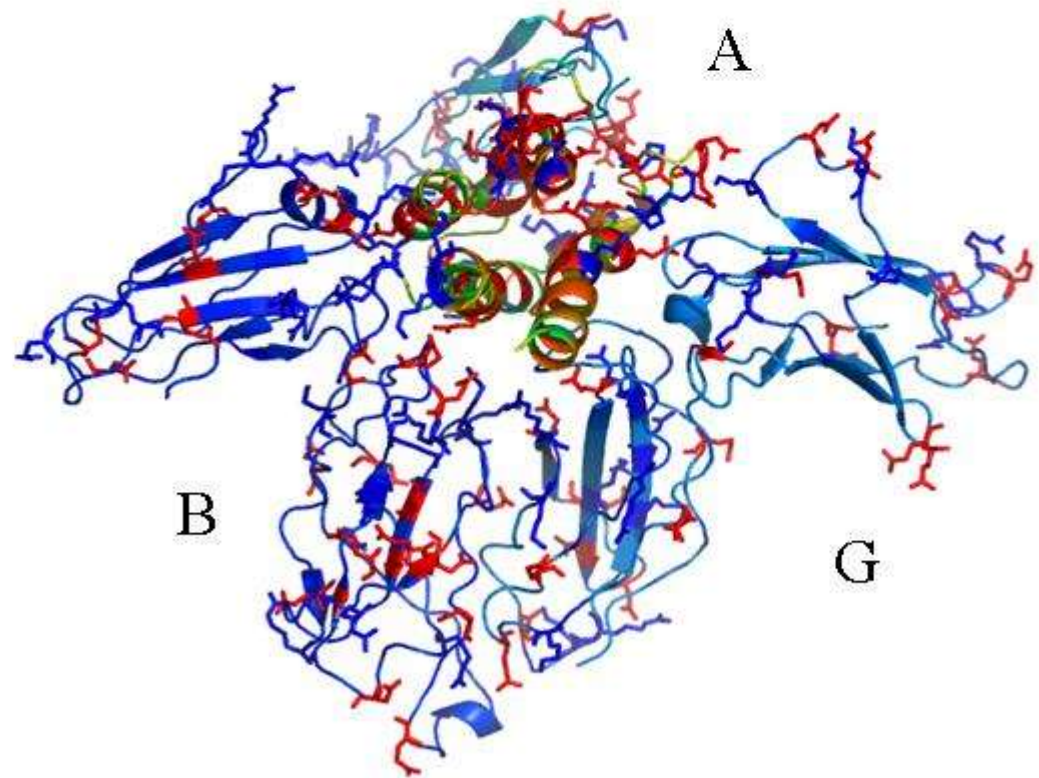
### Interface IL2b et IL2Rg

B:K163

B:E136

G: D147

G: R149



## Contacts électrostatiques dans IL2-IL2Rabg:

### Interface IL2 et IL2Ra

2:K35 (1ex)

2:R38

2:K43 (2ex)

2:K64

A: E9

A: D6

A: D4 (1/2ex)

A: D56

human K-D

bovine D-H

### Interface IL2 et IL2Rb

2:E95 (15ex)

2:D220

B: K71, R91 (1/2 ex) human E-K

B: R15

mouse K-S

### Interface IL2 et IL2Rb

2:K110 (2ex)

2:K116

G: K36

G: K91

### Interface IL2b et IL2Rg

B:K163

B:E136

G: D147

G: R149

