

Projet Biosystémique

**Flavie Landomiel, Cathy Guérineau
Anubhav Gupta
Denis Maurel, Anne Poupon**



Parlons Biologie

Le projet Biosystémique côté "biologie"

Projet Biosystémique

□ But ultime

- Construire un réseau de signalisation en biologie systémique
 - reconstruire les voies de signalisation des récepteurs couplés aux protéines G
- À partir des grandes bibliothèques numériques d'articles scientifiques
 - Istex et ses dix-huit millions de documents
 - NCBI resource (PubMed)
- Trouver du "non-dit"... ou plutôt du "non-mis en valeur", deci_delà... à travers un grand nombre d'articles

Projet Biosystémique

□ Hormones et récepteurs

- Les cellules développent des réponses spécifiques aux stimuli envoyés par l'organisme
- Le plus souvent par la mise en circulation d'hormones
 - Ces hormones se lient à des récepteurs spécifiques à la surface des cellules
 - Cette liaison déclenche des cascades de réactions moléculaires
 - Appelé transduction du signal

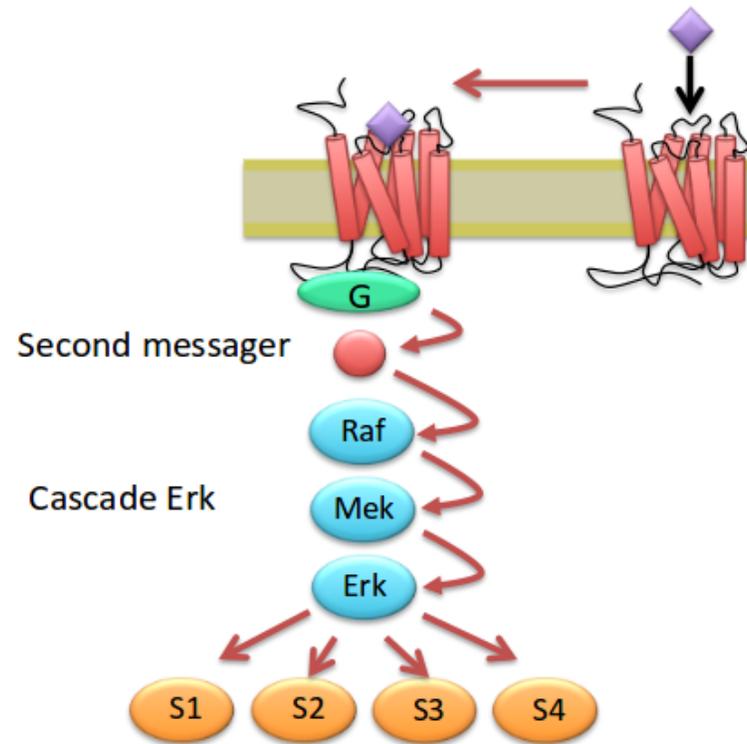
Projet Biosystémique

- Pour construire des réseaux de signalisation, l'équipe BIOS étudie la transduction du signal par les récepteurs couplés aux protéines G (RCPG)
 - plus de 800 récepteurs différents
 - des cibles pharmaceutiques idéales
 - environ 40% des médicaments aujourd'hui sur le marché
 - mais seulement 15% des récepteurs « utilisés » par la pharmacopée
 - quid des effets secondaires ?

Projet Biosystémique

- **Pour construire des réseaux de signalisation, l'équipe BIOS étudie la transduction du signal par les récepteurs couplés aux protéines G (RCPG)**
 - Les voies de signalisation sont encore mal connues
 - Une meilleure connaissance permettra la mise au point de médicaments
 - Plus efficaces
 - Ayant moins d'effets secondaires indésirables

Réseaux de signalisation



Simpl
e

A

1

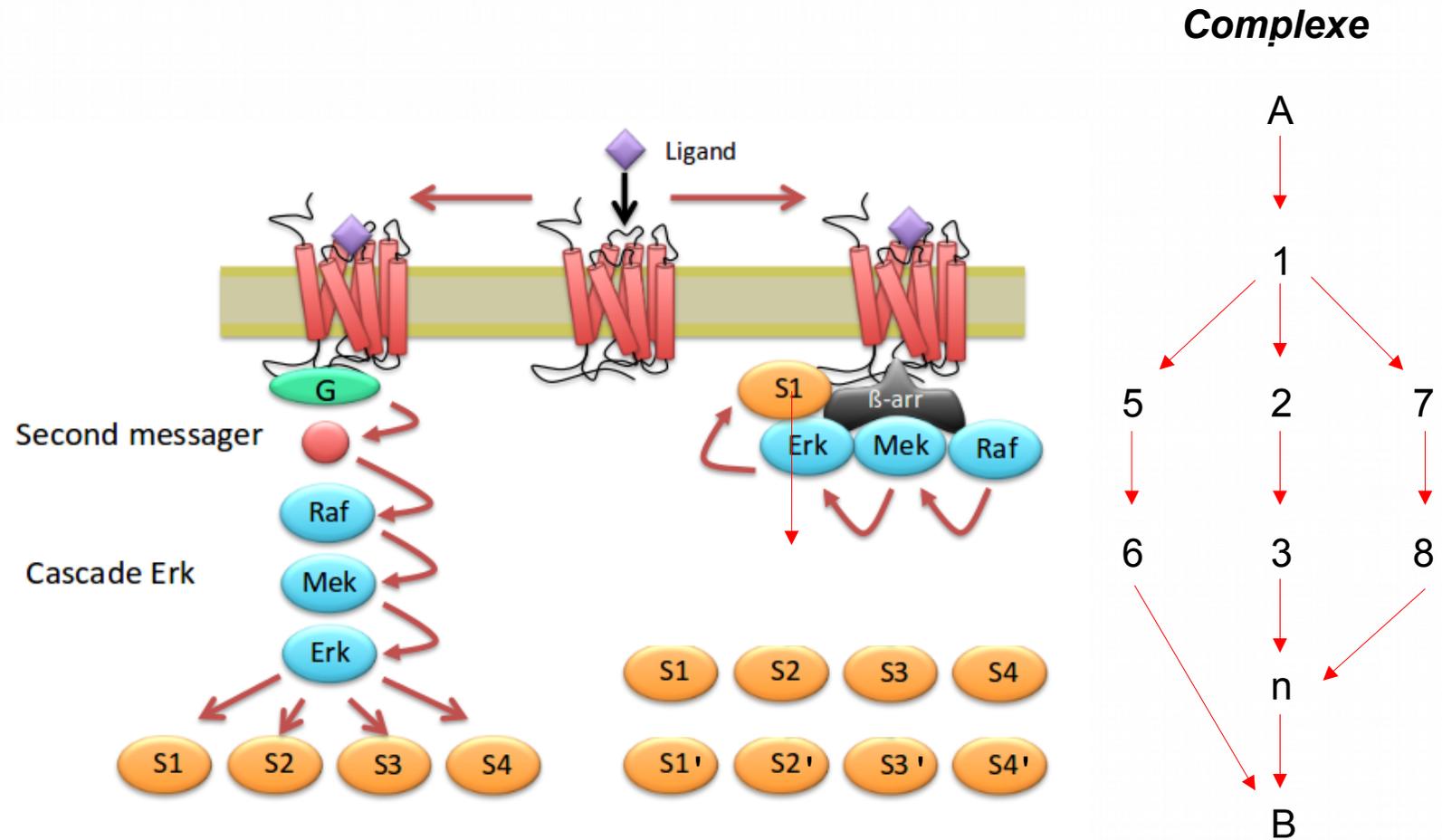
2

3

n

B

Réseaux de signalisation



Nourrir un système expert

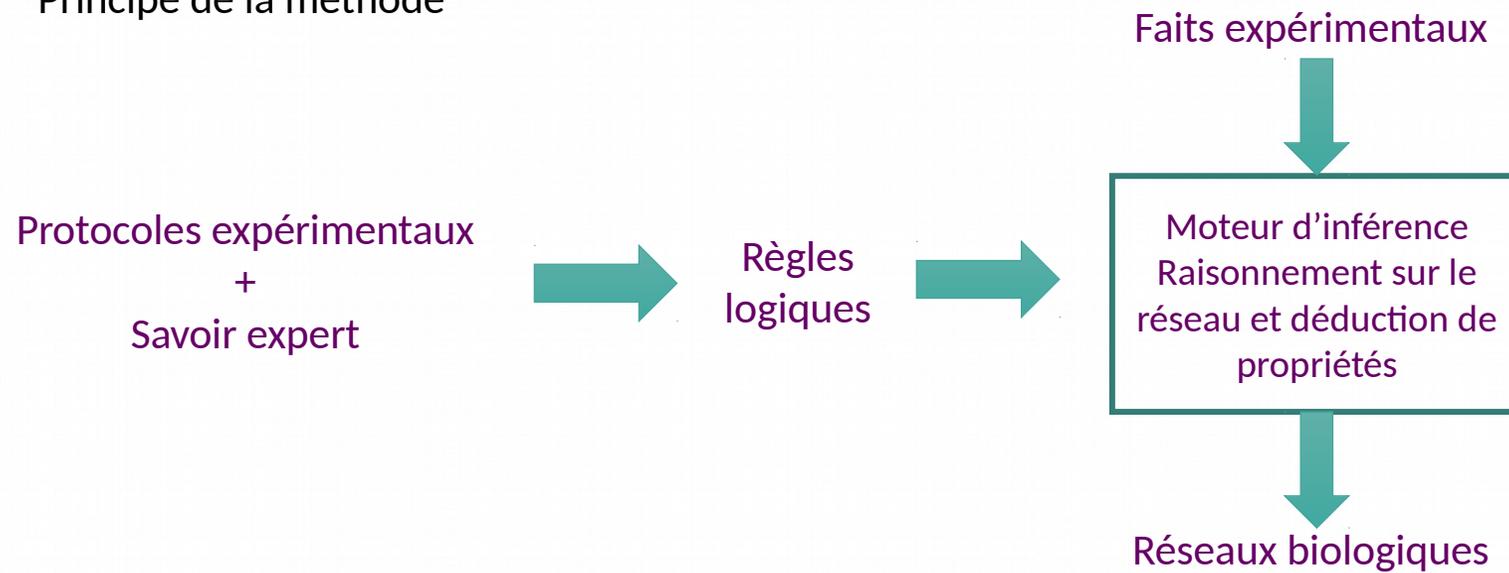
□ Inférence automatique

- L'équipe BIOS a créé un système-expert
 - À base de règles logiques appliquées à des "faits"
 - SI condition(s) ALORS conclusion(s)
 - Simples : conclusion(s) obtenue(s) à partir d'une expérience
 - Complexes : conclusion(s) obtenue(s) à partir de la combinaison de conclusions précédentes

Nourrir un système expert

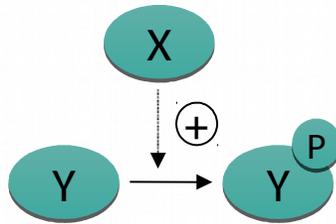
□ Inférence automatique

Principe de la méthode



Nourrir un système expert

Règles primaires



x phosphoryle y

On observe plus de y phosphorylée en présence de x
(phosphorylation assay)

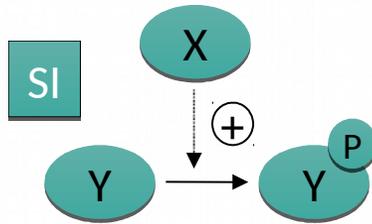
$PA(x;y;py;increase)$

x active la phosphorylation de y
 $PHOSPHORYLATE(x; y; py; increase)$

```
IF PA(x;y;b;increase)
THEN PHOSPHORYLATE(x;y;py;increase)
```

Nourrir un système expert

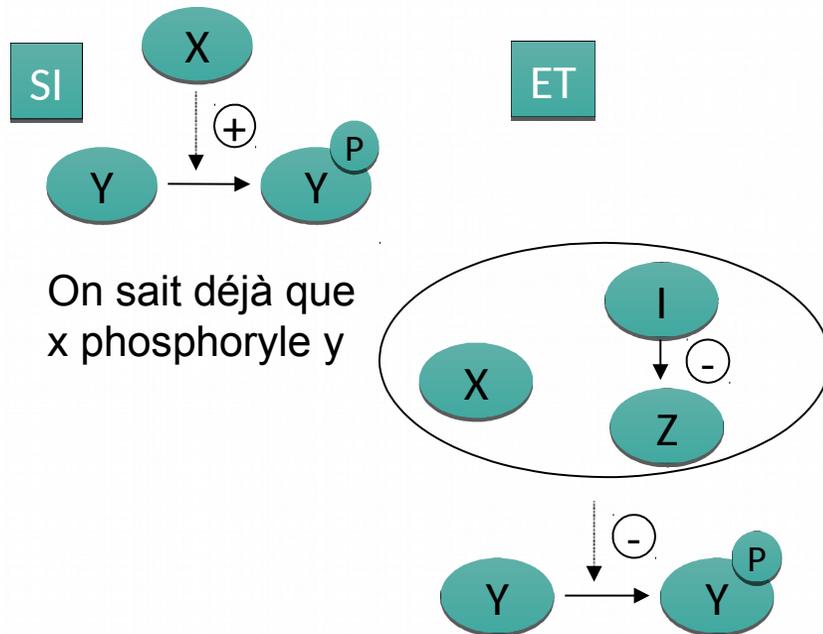
Règles secondaires



On sait déjà que
x phosphoryle y

Nourrir un système expert

Règles secondaires

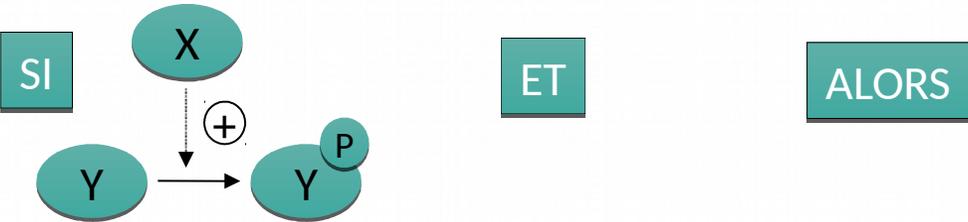


On sait déjà que
x phosphoryle y

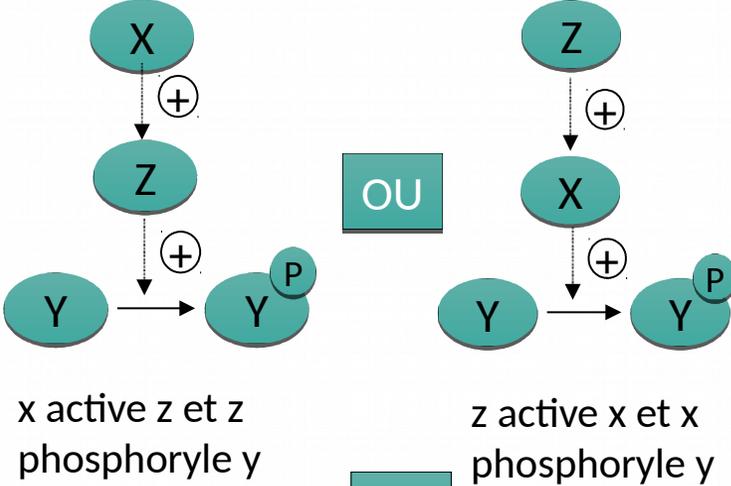
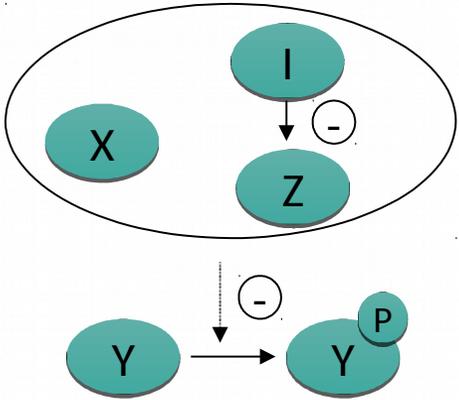
On ajoute un inhibiteur de la
protéine z, on observe que la
quantité de py diminue

Nourrir un système expert

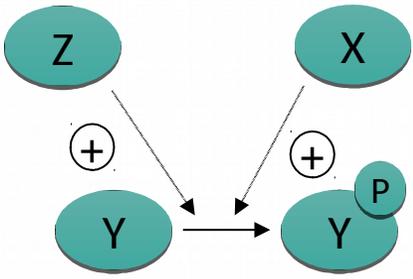
Règles secondaires



On sait déjà que x phosphoryle y



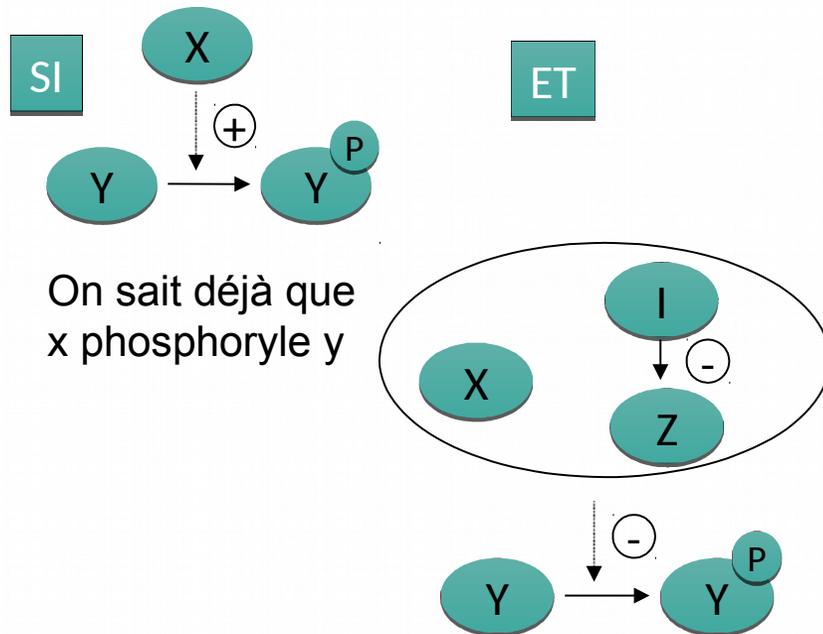
OU
x et z phosphorylent y de manière indépendante



On ajoute un inhibiteur de la protéine z, on observe que la quantité de py diminue

Nourrir un système expert

Règles secondaires



On sait déjà que
x phosphoryle y

On ajoute un inhibiteur de la
protéine z, on observe que la
quantité de py diminue

IF x phosphorylate y
AND inhibitor of z decreases phosphorylation
THEN (x activates z AND z phosphorylates y)
OR (z activates x AND x phosphorylates y)
OR (z phosphorylates y AND x phosphorylates y)

Nourrir un système expert

❑ Test sur le réseau FSH

▪ Faits initiaux extraits de la bibliographie

❑ ~ 150 articles

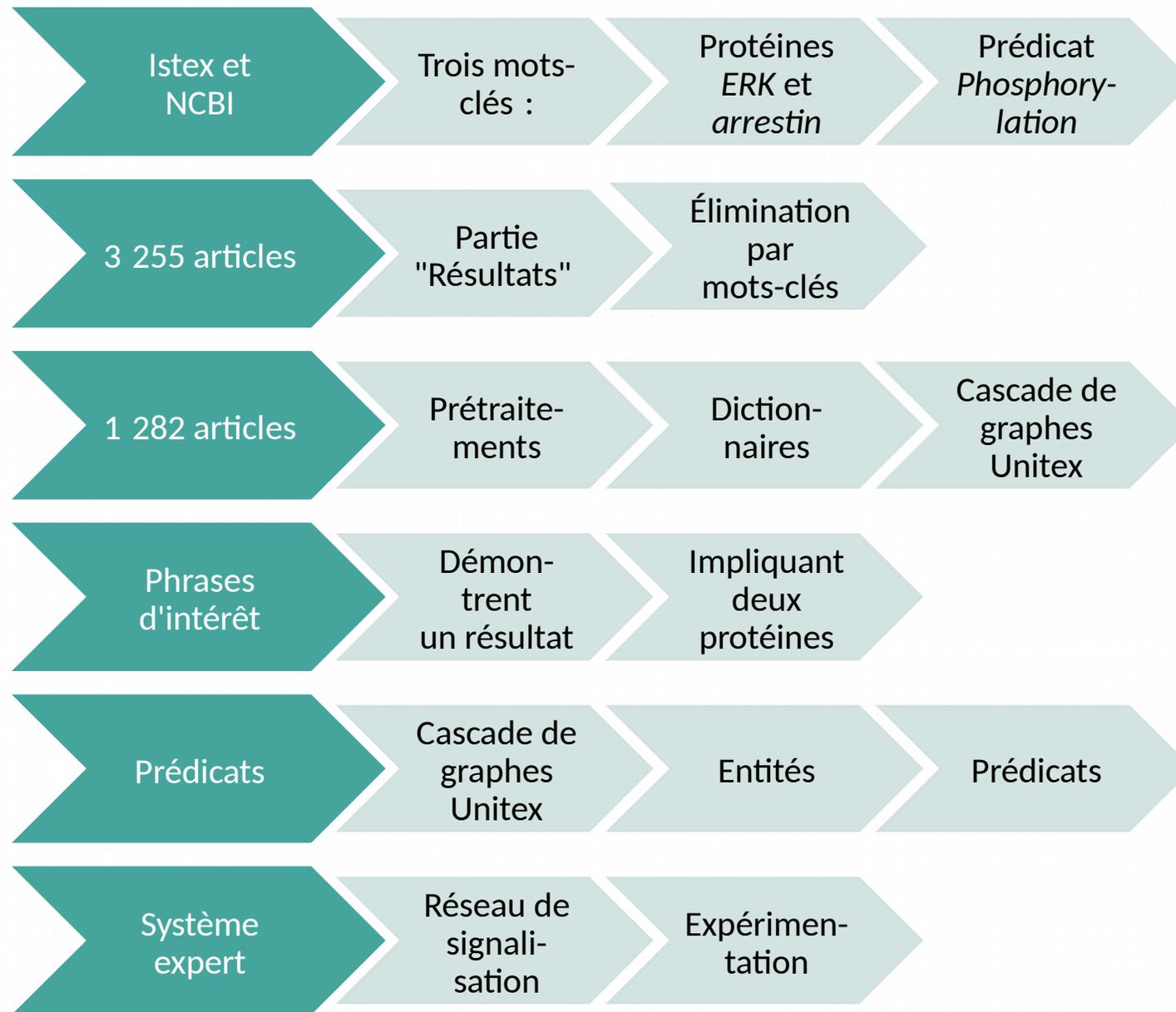
❑ ~ 250 expériences

❑ Actuellement l'extraction des expériences à partir des publications est manuelle !

❑ Peut-on l'automatiser ?

Parlons informatique

Le projet Biosystémique côté "informatique"



Le corpus

□ Un article scientifique en biologie

1. Résumé

- principaux résultats

2. Introduction

- contexte de l'étude et état de l'art

3. Matériel et méthodes

- mise en pratique des expériences

4. Résultats

- liste et détail des expériences réalisées

5. Discussion

- interprétation des expériences et perspectives

Le corpus

□ Un article scientifique en biologie

1. Résumé

- principaux résultats

2. Introduction

- contexte de l'étude et état de l'art

3. Matériel et méthodes

- mise en pratique des expériences

4. Résultats

- liste et détail des expériences réalisées

5. Discussion

- interprétation des expériences et perspectives

Le corpus

❑ Téléchargement ISTEEX

▪ ERK

arrestin

phosphorylation

- ❑ 3 255 documents
(fichier texte)

- ❑ 1 282 documents NCBI
(40%)
(partie *Résultats*)

Bienvenue sur le **démonstrateur ISTEEX**

[En savoir plus](#)

restin AND phosphorylation

Requête `https://api.istex.fr/document/?q=ERK AND arrestin AND phosphorylation &facet=corpusName[*],host.genre[*]>genre[*],pdfVersion[*],refE`

Résultats : **578** (822 ms) 1 / 58

NCBI Resources How To Sign in to NCBI

PMC ERK AND arrestin AND phosphorylation

US National Library of Medicine
National Institutes of Health

[Save search](#) [Journal List](#) [Advanced](#) Help

Search results

Items: 1 to 20 of 2739

<< First < Prev Page 1 of 137 Next > Last >>

Le corpus

- **Téléchargement (fichiers texte)**
 - *ERK, arrestin et phosphorylation*
 - 3 255 documents
 - Sélection de la partie *Résultats*
 - *ERK, arrestin et phosphorylation*
 - 1 282 documents (40%)

Le corpus

□ Identification des phrases d'intérêt qui démontre un résultat

- Identification de **verbes**
- Identification des **molécules**
- Identification des **actions**
- Identifications des **méthodes expérimentales**

C-terminal Ran-binding domain (R)³⁵. ERK phosphorylated Nup50 exclusively in the FG repeat domain *in vitro* (Fig. 4a). LC-MS and tandem MS (MS/MS) analyses of ERK-phosphorylated and unphosphorylated GST-tagged full-length Nup50 demonstrated that ERK phosphorylates Nup50 stoichiometrically at Ser221 and Ser315 *in vitro* (Supplementary Fig. 3). *In vitro* kinase assays with wild-type and Ser-to-Ala mutants of Nup50 confirmed this finding (Fig. 4b).

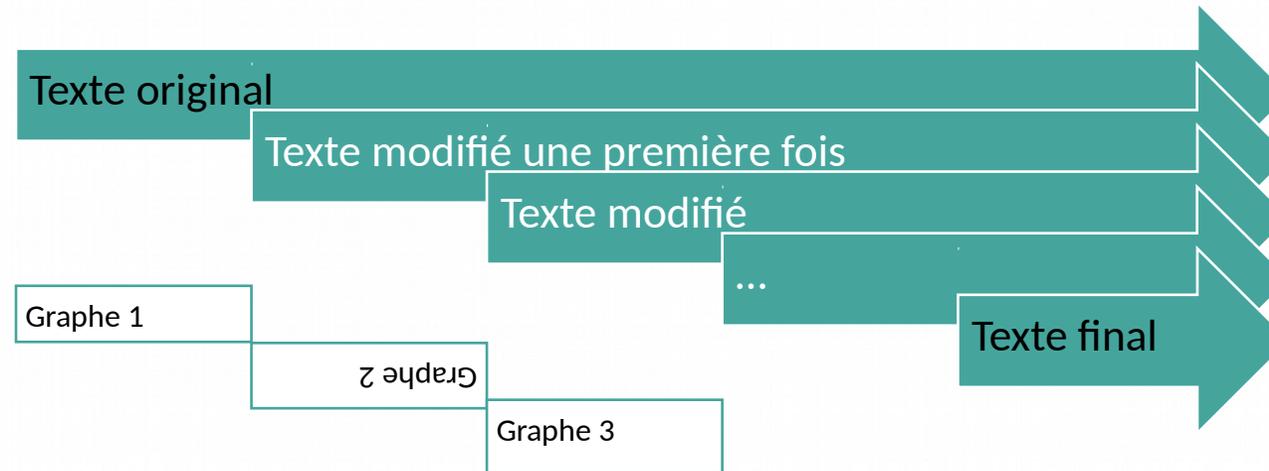
□ Il y a en moyenne 45 phrases d'intérêt par articles

Unitex

- **Unitex est un logiciel libre d'analyse lexicale automatique auquel contribue le Lifat**
- **Unitex allie un système informatique performant**
 - *des réseaux de transitions "augmentées"*
 - *opérations sur le texte*
 - *utilisation de variables*
 - *compilation*
- **et une interface conviviale**
 - *des graphes*

CasSys

- Un module pour la constitution et l'utilisation de cascades de graphes intégré à Unitex (réalisé par le Lifat)



Organisation

□ Prétraitement

- Sélection, normalisation et nettoyage
- Application de quatre dictionnaires (verbes d'action, cellules, composés et protéines)

□ Phrases d'intérêt

- Lancement d'une cascade de graphes sur les 1 282 parties *Résultats*
- 7 graphes dont 3 spécifiques

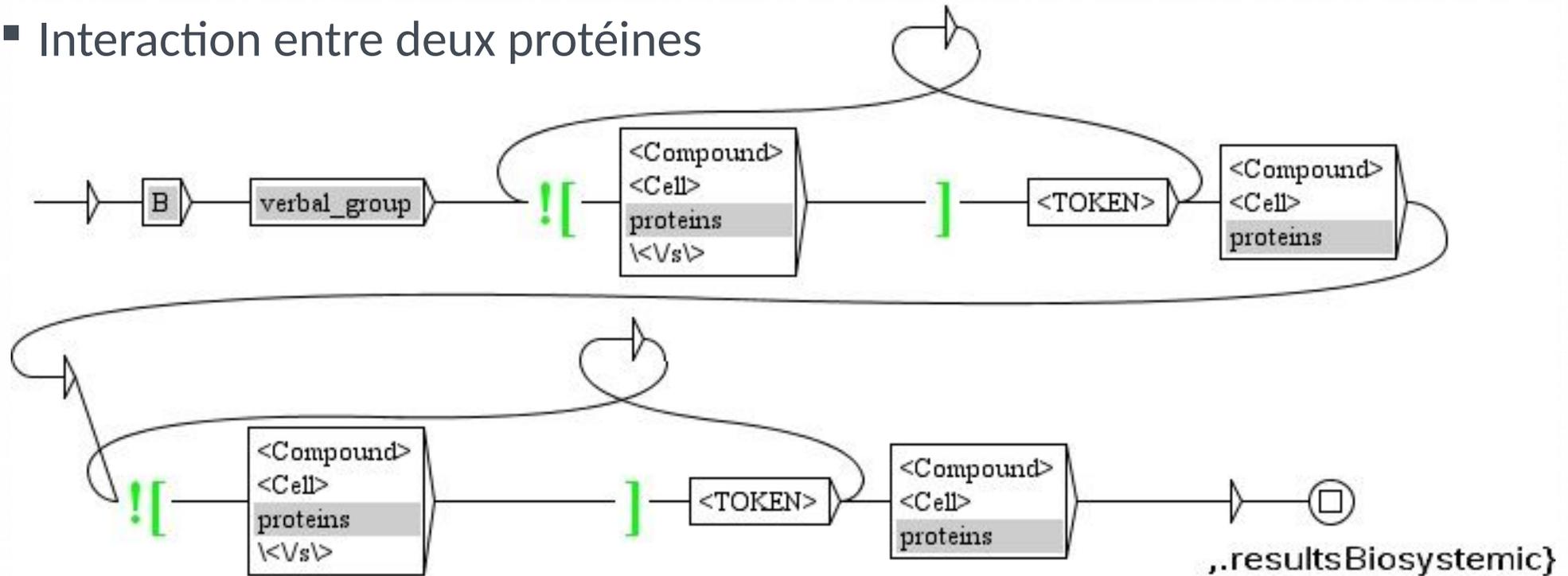
□ Création des faits

- Lancement d'une cascade de graphes sur les 62 655 *Phrases d'intérêt*
- Balisage des entités et des prédicats

Graphes

□ Exemple de graphe pour la recherche des phrases d'intérêt

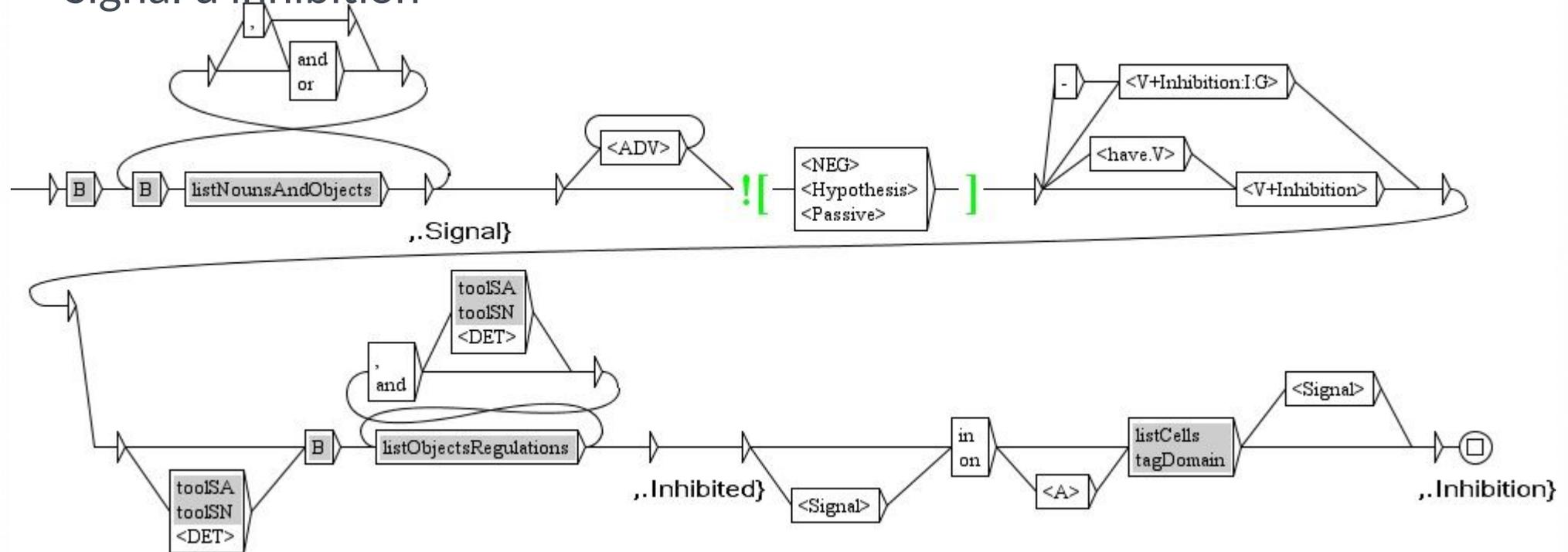
- Interaction entre deux protéines



Graphes

□ Exemple de graphe pour la construction des faits

▪ Signal d'inhibition



Évaluations

□ La recherche des phrases d'intérêt

- 27 textes choisis au hasard (année de parution et journal de publication)

	Phrases d'intérêt
Rappel	90%
Précision	81%

Évaluations

□ La recherche des entités et prédicats

- 5 textes choisis au hasard, soit un total de 226 phrases d'intérêt

	Entités et prédicats	Prédicats
Rappel	88%	65%
Précision	76%	79%

Un exemple réel

Similar to DSLET, EGF (5 ng/ml) produces a robust activation of ERKs in δ -OR-expressing HEK-293 cells after a 5 minute treatment (Figure 2a).

a robust activation
of ERKs
in δ -OR-expressing
HEK-293 cells

molecule(ERKs)	particle(ERKs)
protein(ERKs)	cell(HEK-293)
molecule(δ -OR)	particle(δ -OR)
protein(δ -OR)	

expressed(δ -OR, HEK-293, positive)
reactionModulation(unknown signal, ERKs_inactive,
ERKs_active, increase, δ OR-expressing HEK-293)

Merci !

Ce projet a été financé par le programme "Chantiers d'avenir"
instauré dans le cadre du projet Istex (ANR-10-IDEX-0004-02)