

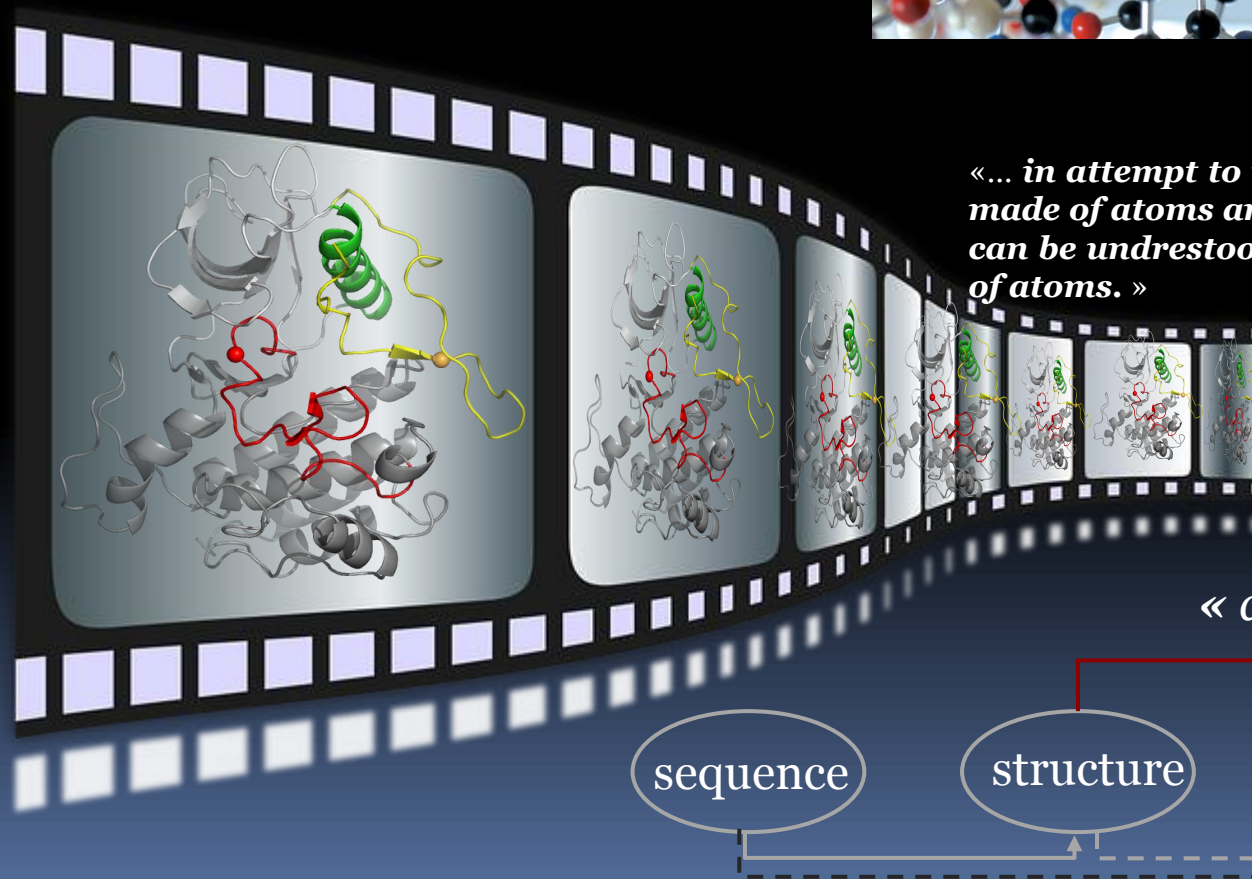
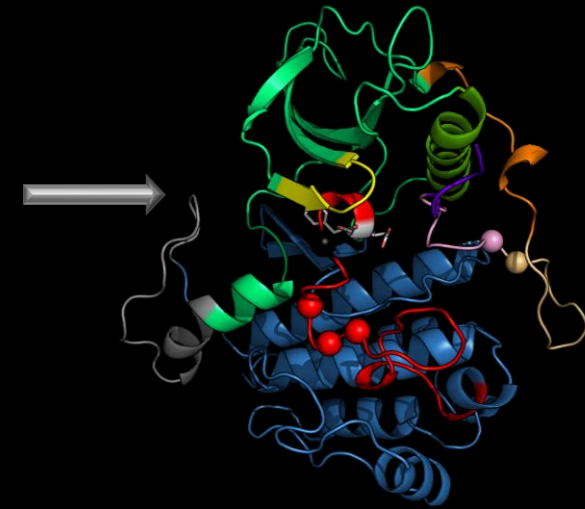
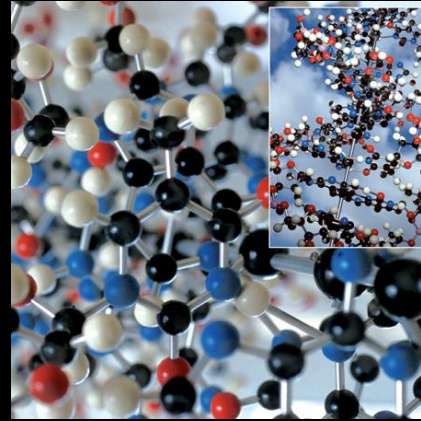
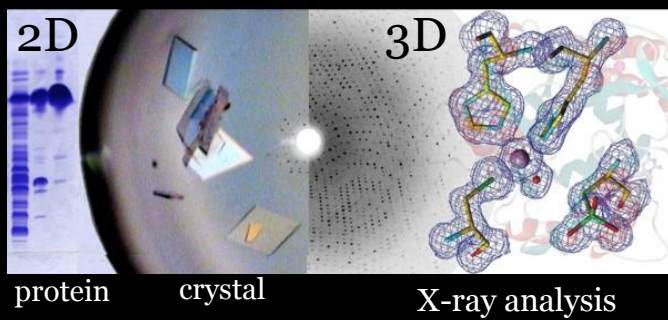
Transition conformationnelle des protéines : Topologie “Dynamique”, TOPDYN-2

FARMAN 2014 (LBPA – CMLA)

- *Luba Tchertanov*
- *Isaure Chauvot de Beauchêne, postdoc à l'Université
Technique de Munich*
- *Florent Langenfeld, étudiant en thèse*
- *Yann Guarracino, étudiant en stage Master*
- *Alain Trouvé*
- *Ouriel Bloede (Elève ENSC, License)*
- *Andrei Bengus-Lasnier (Eleve ENSC, License)*
- *Benjamin Charlier*
- *Nicolas Pajor*

Serveur TopDyn
Serveur de sauvegarde

Protein structure and dynamics



«... in attempt to understand life, it is that all things are made of atoms and that everything that living things do can be understood in terms of the jiggling and wiggling of atoms. »

Feynman Lectures on Physics (1964)

« dynasome » descriptors

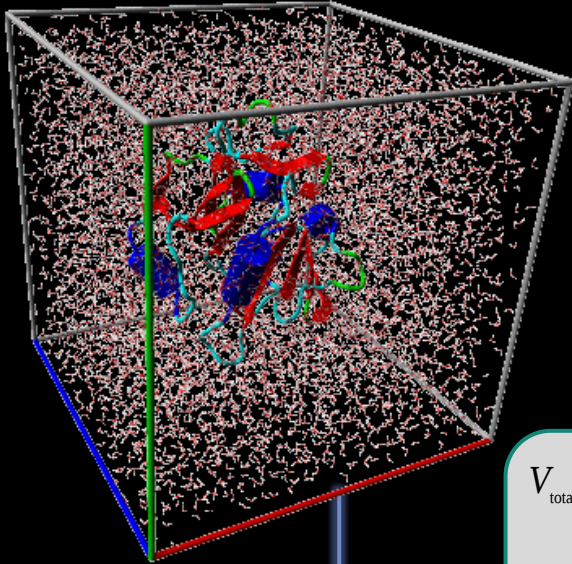
sequence

structure

dynamics

function

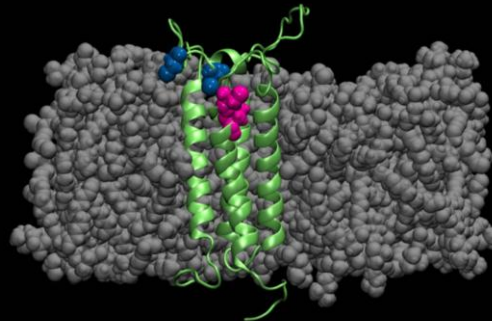
Molecular dynamics simulations experiment (MM, tout-atomes)



AMBER
GROMACS
Desmond
CHARMM
ACEMD
NAMD

$$m \frac{d^2 \vec{r}}{dt^2} = -\nabla V(\vec{r})$$

$$V_{\text{total}} = \sum_{\text{bonds}} K_b (r - r_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \gamma)] \\ + \sum_{\text{Hbonds}} \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \sum_{\text{van der Waals } i, j \text{ pairs}} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \sum_{\text{electrostatic } i, j \text{ pairs}} \frac{q_i q_j}{\epsilon r_{ij}}$$



Concept ?

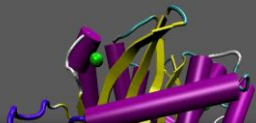
Results representation

Trajectory analysis

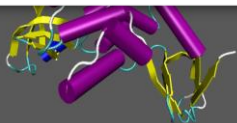


Protein motions

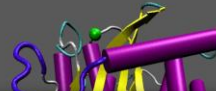
Type of motion	Functional Examples	Time	Amplitude Scale
Local Motions • Atomic fluctuations • Side Chain motion	Ligand docking flexibility Temporal diffusion pathways	fs to ps $10^{-15} - 10^{-12}$ s	$< 1 \text{ \AA}$
Medium-scale Motions • Loop motion • Terminal-atom motion • Rigid-body motion	Active site conformation adaptation Binding specificity	ns to μ s $10^{-9} - 10^{-6}$ s	$1-5 \text{ \AA}$
Large-scale Motions • Domain Motion • Subunit Motion	Hinge-bending motion Allosteric transition	μ s to ms $10^{-6} - 10^{-1}$ s	$5-10 \text{ \AA}$
Global Motions • Helix-coil transition • Foldin/unfolding • Subunits association	Protein activation Protein functionallity	ms to hours $10^{-1} - 10^4$ s	$> 10 \text{ \AA}$



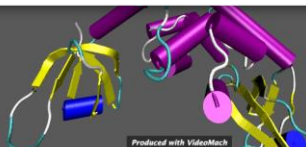
fs to ps



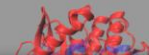
Produced with VideoMach
www.videomach.com



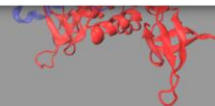
ns to μ s



Produced with VideoMach
www.videomach.com

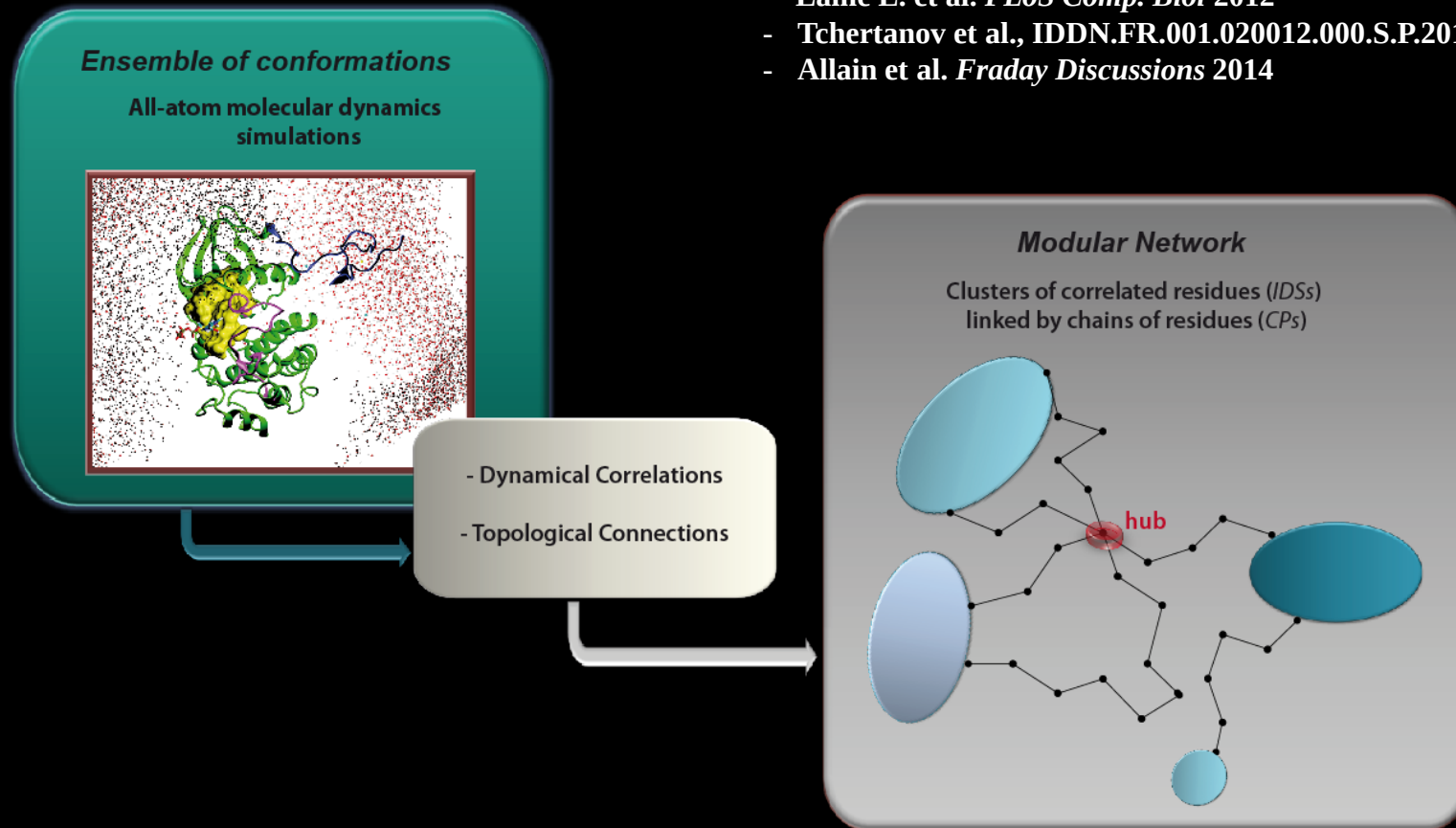


μ s to ms



MOdular NEtwork Analysis (MONETA)

- Laine E. et al. *PLoS Comp. Biol* 2012
- Tchertanov et al., *IDDN.FR.001.020012.000.S.P.2014.000.31235*
- Allain et al. *Fraday Discussions* 2014

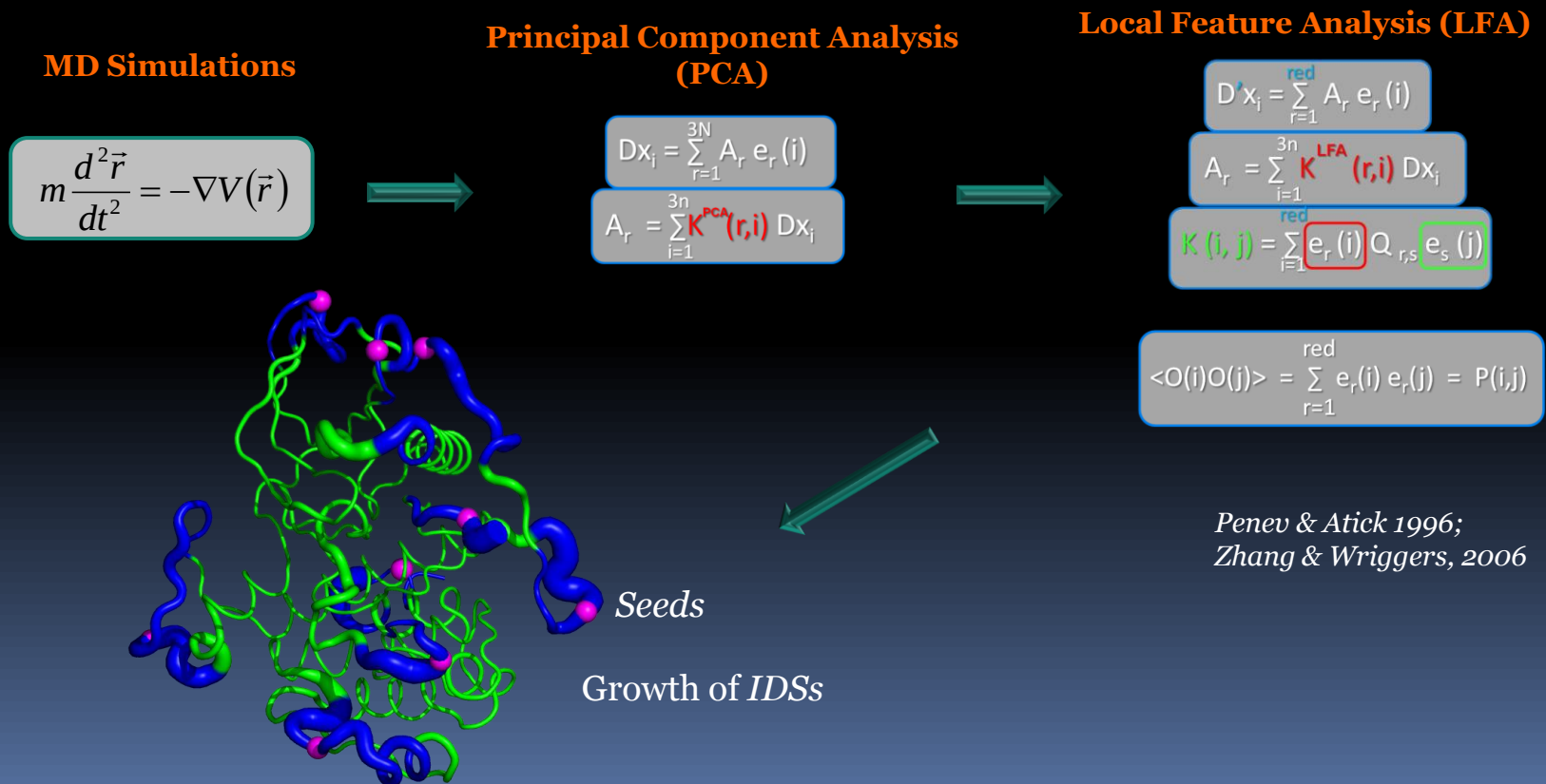


- **Independent Dynamic Segments (IDSs)** – locally coupled residues
- **Communication pathways (CPs)** – chains of connected residues

1. Independent Dynamic Segments

IDSs – de groupes de résidus présentant des fluctuations atomiques concertées et localement corrélées (sous ensembles indépendants).

- Approche initialement utilisée:



Problèmes:

- Sources *localisées* et *indépendantes* de variabilités de la dynamique
Manque d'une conceptualisation vraiment opérationnelle.
- L'ACP fait émerger des facteurs de variabilités non locaux (projection sur directions principales). *Inadaptée* pour la recherche des IDSs

Blanchiment *partiel* des données $X = (X(i)) \rightarrow O = (O(i))$ suivi d'une sélection des atomes sources par un algorithme de *poursuite de projection*

$$O^{rec}(i) = \sum_{m=1}^n \hat{a}_m(i) O(i_m)$$

$$i_{\{n+1\}} = \operatorname{argmax}_i \langle |O(i) - O^{rec}(i)|^2 \rangle$$

PB: Paramètres cachés, perte de localité dans le processus de blanchiment et difficulté d'analyse et d'interprétation statistique.

Recherche des *IDSs*: Une nouvelle approche par PFD

Principal Component Analysis
comme recherche de sous-espaces
optimaux de prédicteurs *arbitraires*

Principal Features Decomposition
comme recherche de sous-espaces
optimaux de prédicteurs *contraints*

Espace des configurations:

$$X \in E = \mathbb{R}^{3N}$$

Predicteurs :

$$p_V(X) \text{ avec } V \text{ sev de } E$$

Erreur de Prediction Résiduelle :

$$\text{EPR}(V) \stackrel{\text{def}}{=} \min_{A \in L(V, E)} \langle \|X - Ap_V(X)\|^2 \rangle$$

Prédicteurs optimaux :

$$\begin{aligned} \hat{V} &= \underset{\dim(V) \leq k}{\operatorname{argmin}} \langle \|X - p_V(X)\|^2 \rangle \\ &= \underset{\dim(V) \leq k}{\operatorname{argmin}} \text{EPR}(V) \end{aligned}$$

**Decomposition en espaces de
prédicteurs caractéristiques
(features) :**

$$E = E_1 \oplus E_2 \cdots \oplus E_N$$

E_j : coordonnées du $j^{\text{ième}}$ atome C_α

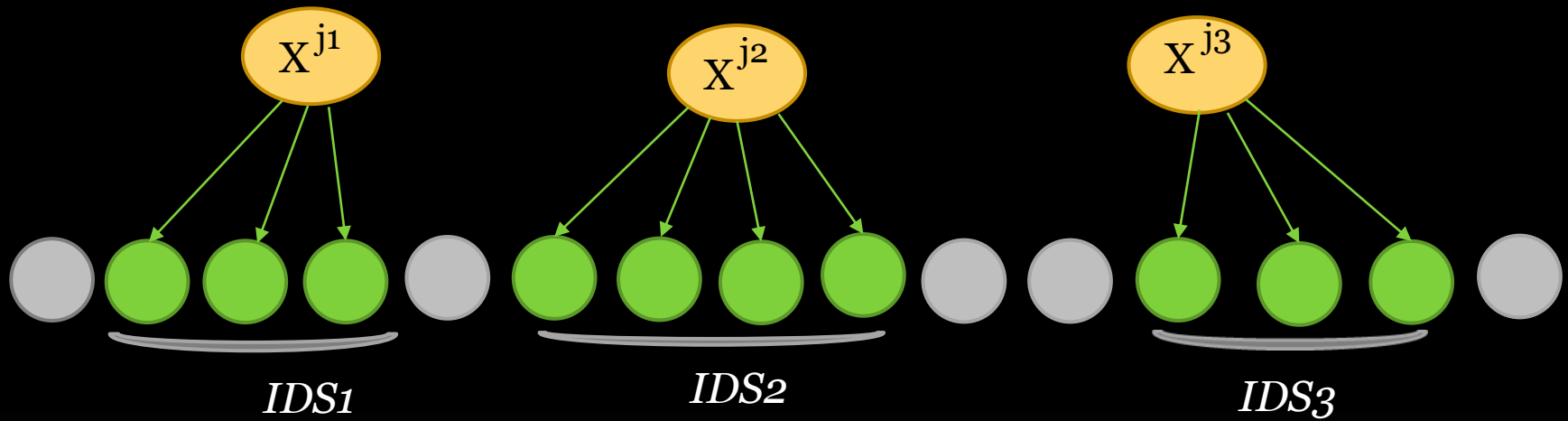
$$\hat{V} = \underset{V = E_{j_1} \oplus \cdots \oplus E_{j_k}}{\operatorname{argmin}} \text{EPR}(V)$$

Algorithme glouton :

$$\hat{j}_1, \hat{j}_2, \cdots, \hat{j}_k$$

Modèle probabiliste sous-jacent des *IDSs* :

La métaphore de l'orchestre

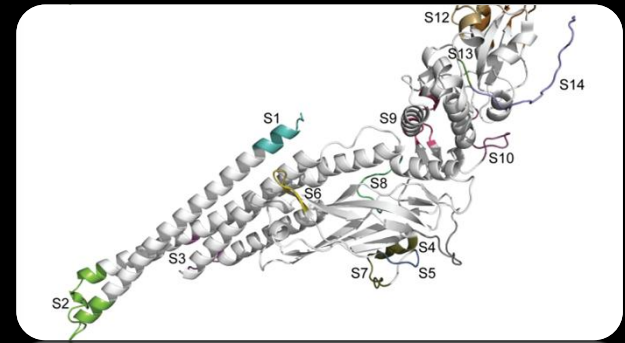
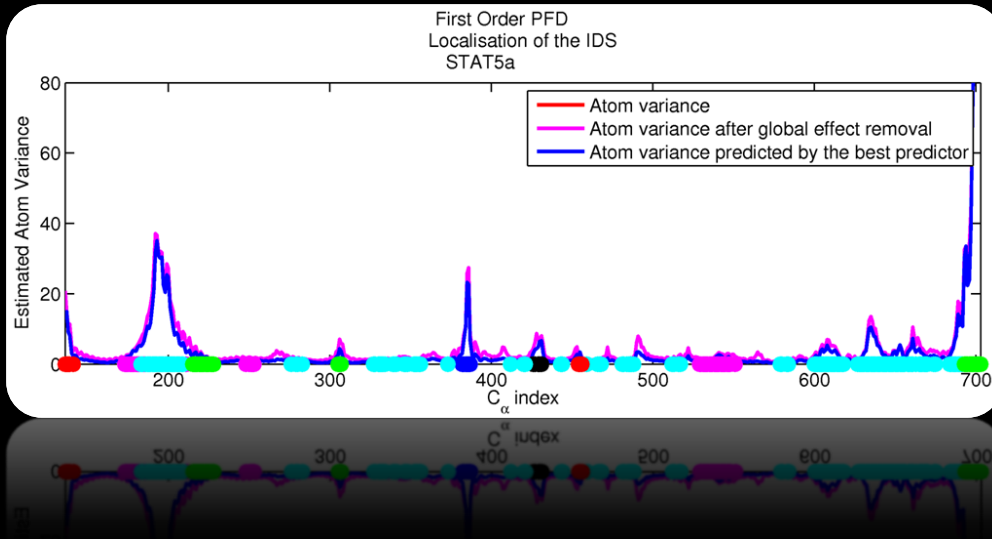


 chef de chœur

 choriste

 auditeur

Très bons résultats sur KIT mais résultats déroutants sur STAT5



Récupération de très grands segments couvrant de très larges portions de la molécule !

La propriétés de *localisation* et d'*indépendance* des sources mise en défaut sur STAT5

Présence de déformations globales de la protéine rendant dépendant tous les atomes entre eux même en présence d'un recalage rigide préalable sur la configuration initiale

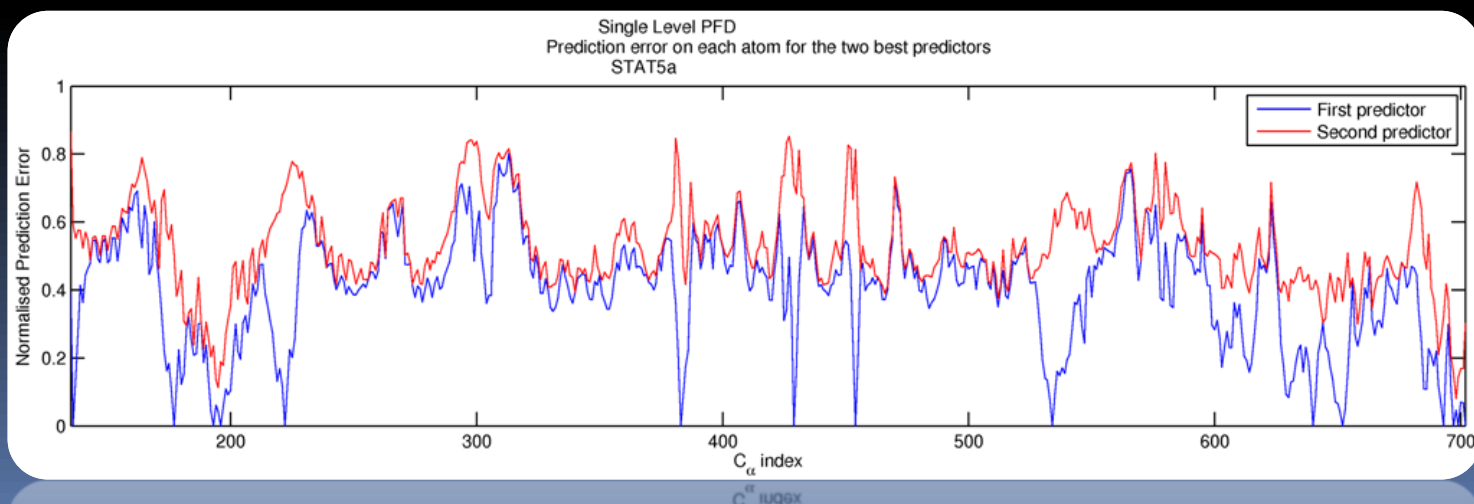
Quantitativement sur STAT5:

Chaque atome prédicteur est capable d'expliquer environ 30% à 60% de la variance de n'importe quel autre atome, même très éloigné dans la molécule

Conséquence:

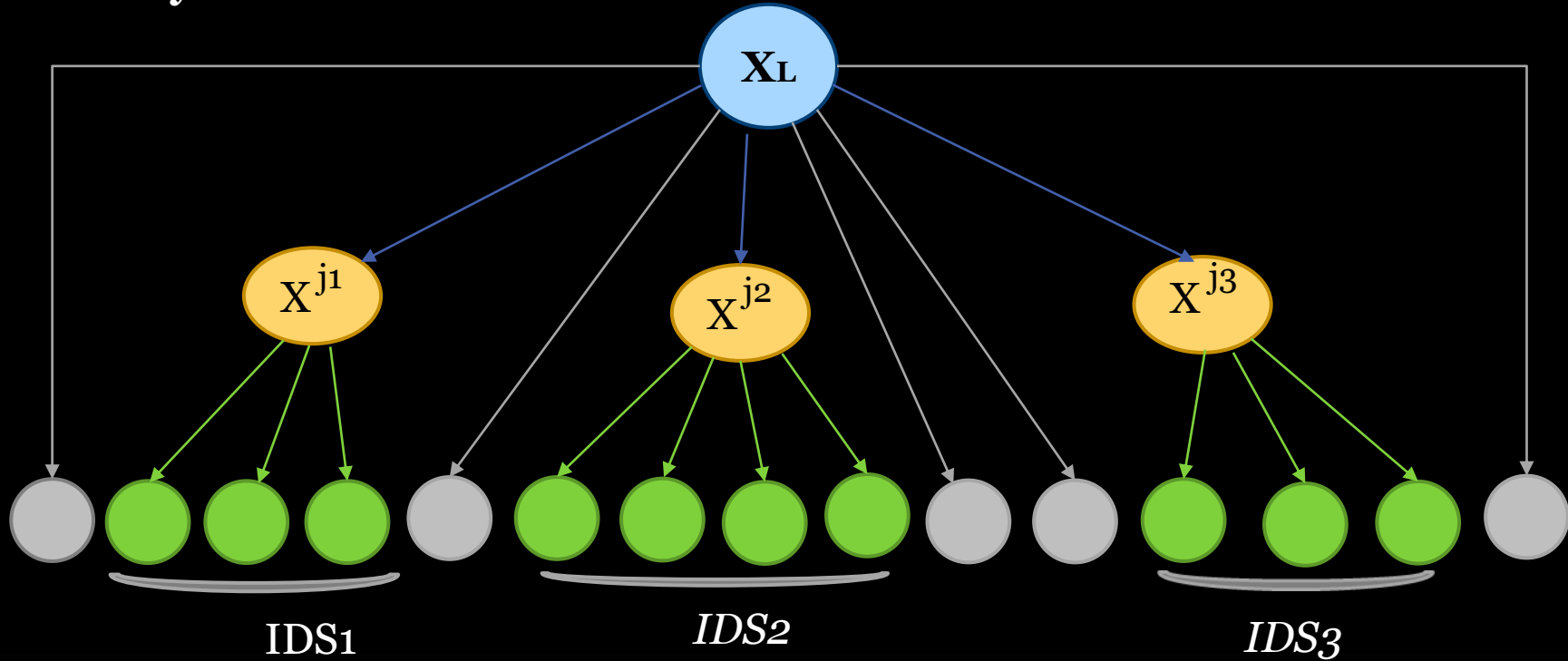
1. Il n'existe pas deux atomes prédicteurs indépendants.
2. Pas de localisation sur des segments de taille raisonnable

Contradiction avec la notion habituelle de segment dynamique indépendant



Solution proposée: un nouveau niveau

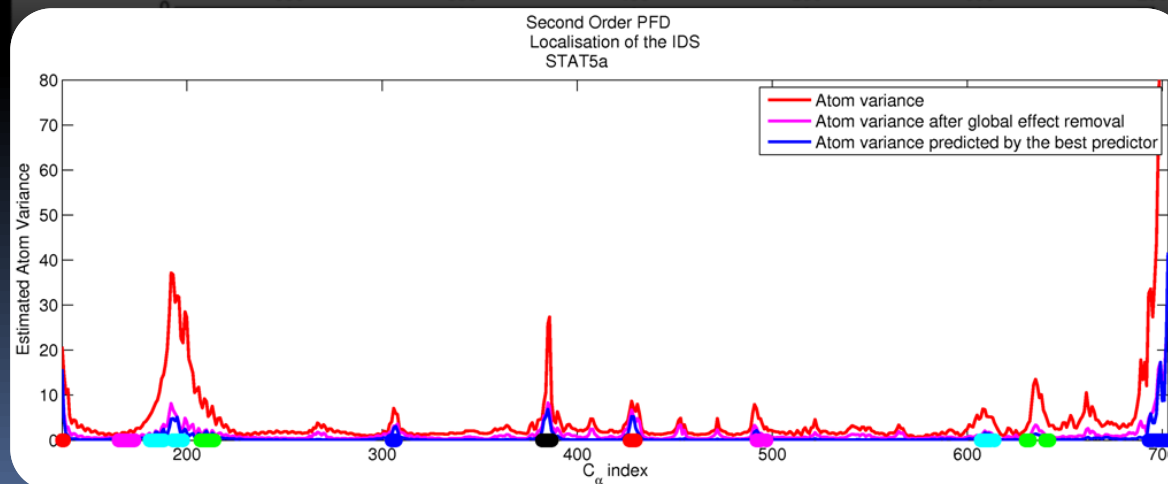
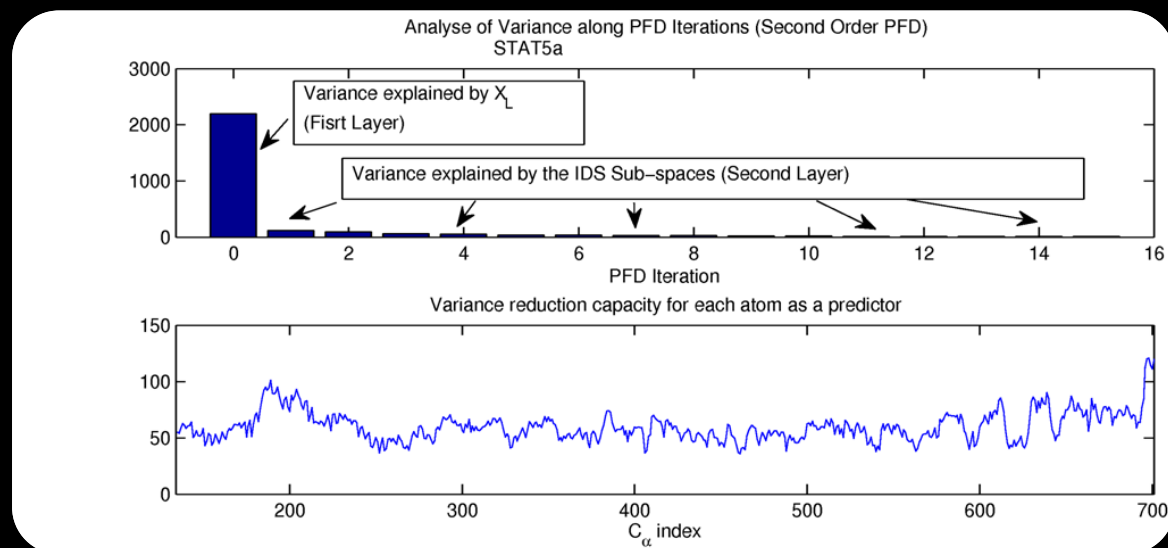
Extension de la PFD dans un formalisme de **Réseau Gaussien Bayésien** sur un arbre à deux niveaux :



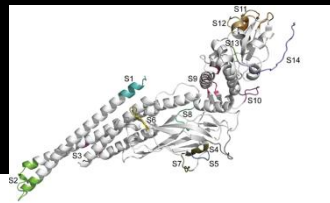
- X_L est un vecteur aléatoire codant les transformations conformationnelles dynamiques globales
- Le niveau suivant correspond à une analyse des effets locaux **conditionnellement** à des transformations globales

Nouvelle approche sur STAT5:

- Des effets globaux massifs (80% de la variance globale)
- Existence d'*IDS*s localisés conditionnels pour les 20% restants.

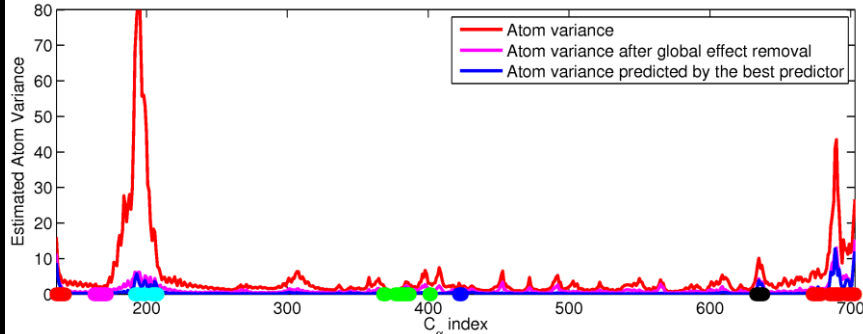


STAT5

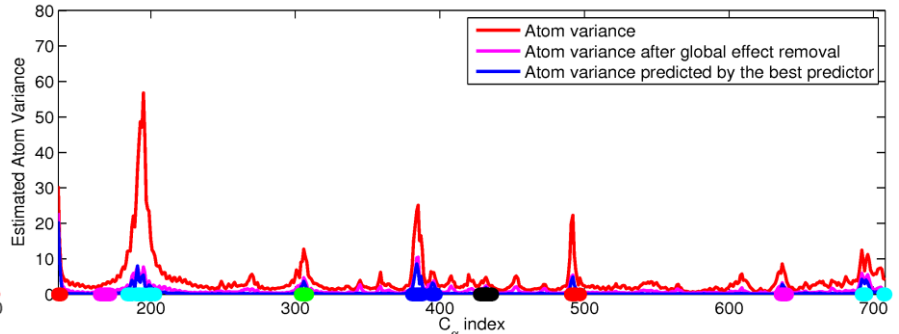


PFD de second ordre

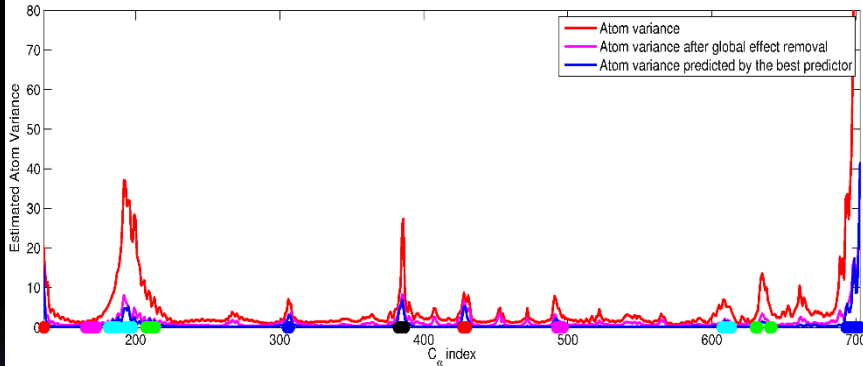
Second Order PFD
Localisation of the IDS
STAT5aP



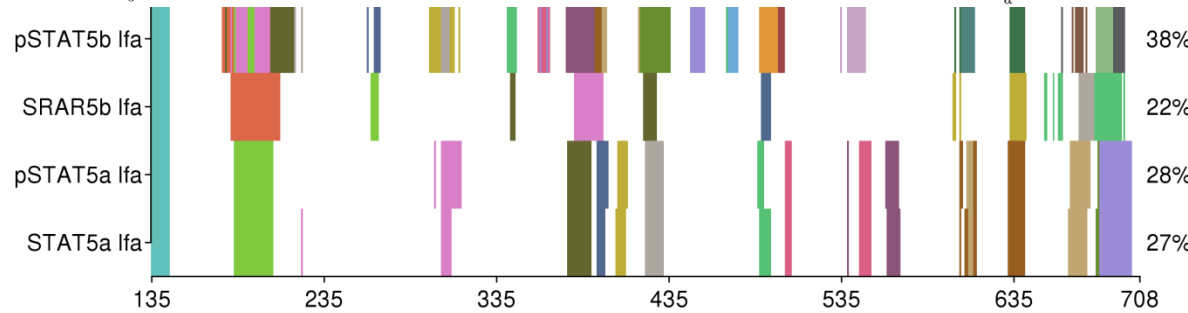
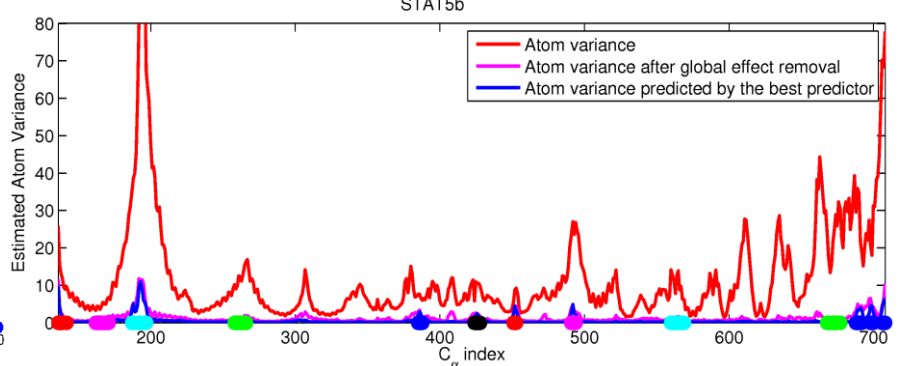
Second Order PFD
Localisation of the IDS
STAT5bP



Second Order PFD
Localisation of the IDS
STAT5a

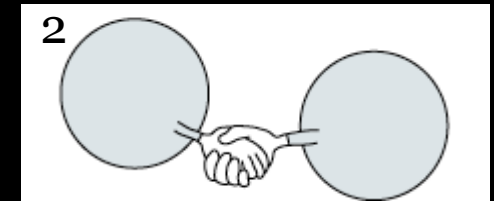
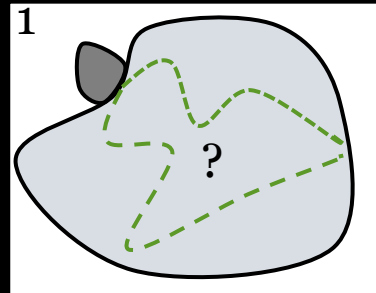
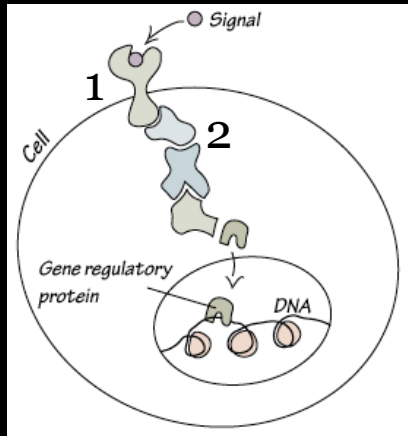


Second Order PFD
Localisation of the IDS
STAT5b

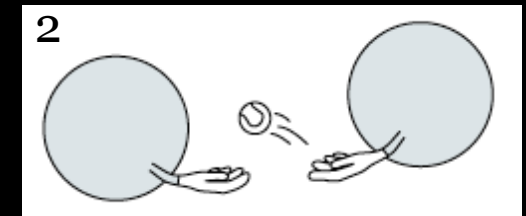


LFA

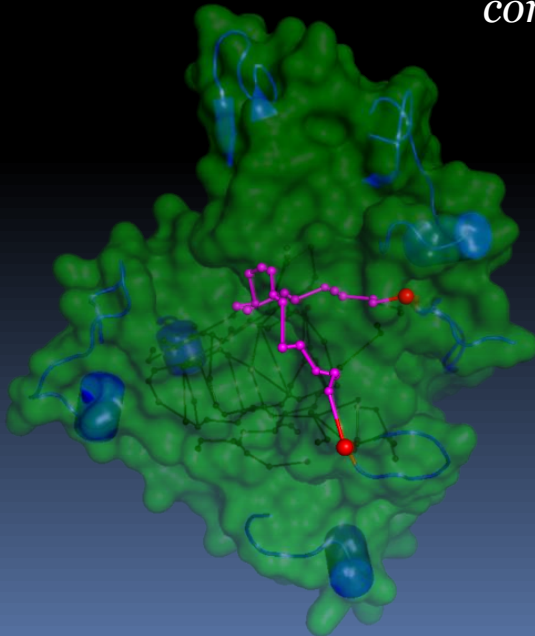
2. Signal transmission in protein



or



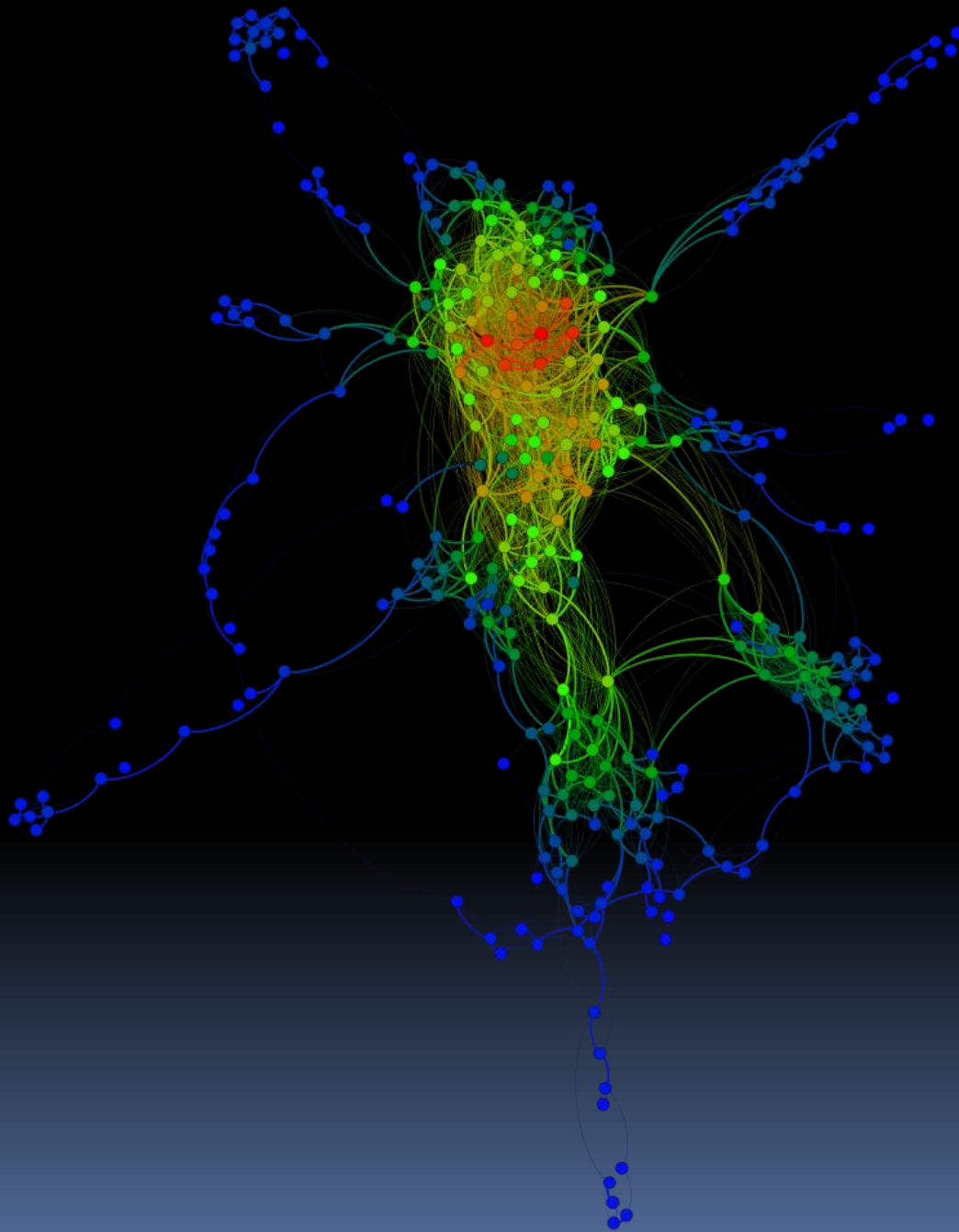
Communication pathways (CPs) are chains of neighboring residues whose communication propensities between each other are high.



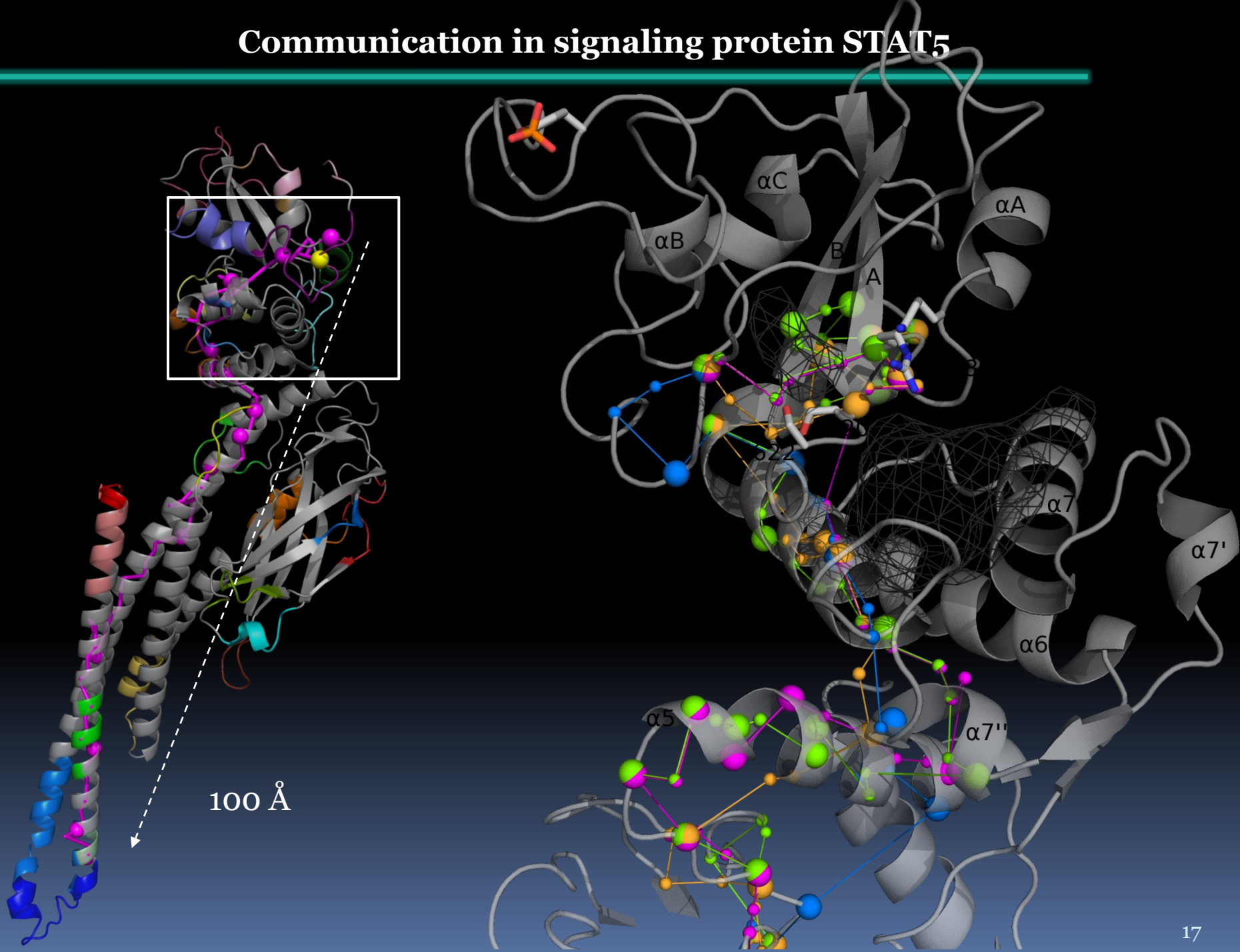
Commute time (CT)

$$CT(i,j) = \langle (d_{ij} - d_{ij}) \rangle$$

Chennubhotla & Bahar, 2007



Communication in signaling protein STAT5



Publications communes LBPA-CMLA

1. Chauvot de Beauchêne I., Alain A., Panel N., Laine E., Trouvé A., Dubreuil P. and Tchertanov L. (2014). Oncogenic mutations of KIT receptor differentially modulate tyrosine kinase activity and drug susceptibility. *PLOS Comput. Biol.*10(7):e1003749. doi: 10.1371/journal.pcbi.1003749.
2. Langenfeld F., Guarracino Y., Arock M., Trouvé and Tchertanov L. (2015). Structure, dynamics and communication features of STAT5, a protein involved in cytosolic signaling. *Science Signaling*. Submission Dec. 2014.

Communication

1. Langenfeld F., Guarracino, Y., Allain A., Laine E., Bengus-Lasnier A., Bloede O., Trouvé A. and Tchertanov L. MODular NETwork Analysis (MONETA) of Allosteric Regulation in Proteins and Protein-DNA complexes. *RICT 2014 – Interfacing Chemical Biology and Drug Discovery*. Rouen, France. 2-4 Juillet 2014.

Autres publications (MONETA)

1. Allain A., Chauvot de Beauchêne I., Langenfeld F., Guarracino Y., Laine E., and Tchertanov L. (**2014**). Allosteric Pathway Identification through Network Analysis from Molecular Dynamics Simulations to Interactive 2D and 3D Graphs. *Faraday Disc.*, DOI: 10.1039/C4FD00024B. 169, 303-321
2. Da Silva Figueiredo Celestino Gomes P., Panel N., Laine E., Pascutti P. G., Solary E. and Tchertanov L. (**2014**). Differential effects of CSF-1R D802V and KIT D816V homologous mutations on receptor tertiary structure and allosteric communication. *PLOS ONE*. May 14; 9(5):e97519. doi: 10.1371/journal.pone. 0097519.

INTERDEPOSIT CERTIFICATION **2014**. Certificat délivré par Agence pour la Protection des Programmes. Inter Deposit Digital Number IDDN.FR.001.020012.000.S.P.2014.000.31235 pour l'œuvre: MOdular NETwork Analysis (MONETA) version 2.0 en date du 31 juillet 2013. Les auteurs: Tchertanov L, Laine E., Allain A., Chauvot de Beauchêne I.

Guarracino, Y., Langenfeld F., Panel N. and Tchertanov L. Extension of MONETA to analysis of Allosteric communication in Protein-Protein and Protein-DNA complexes. Faraday Discussions 169. 7-9 May **2014**. Nottingham, UK. **Best Poster Award, The Royal Chemical Society**.

<http://www.rsc.org/conferencesandevents/rscconferences/fd/fd169/index.asp>

- **MONETA est utilisé dans 15-20 laboratoires publics à travers du monde.**

- **Proposition d'achat par les entreprises privés: 2**

Merci FARMAN !