

Transition conformationnelle des protéines : Topologie “Dynamique”, TOPDYN

FARMAN 2013 (BiMoDyM, LBPA – CMLA)

- *Luba Tchertanov*
- *Isaure Chauvot de Beauchêne, étudiante en thèse*
- *Florent Langenfeld, étudiant en thèse*
- *Ariane Allain, Stage Master*

- *Alain Trouvé*
- *Frédéric Pascal*
- *Benjamin Charlier, PostDoc*
- *Jérôme Jacquenod, élève*
- *Mathieu Kohli, élève*
- *Edwin Hamel-de-le-Court, élève*

Topdyn: Problématiques LPBA/CMLA et Faux-Amis

Mise en place d'un **vocabulaire commun** et éliminations des faux amis (**topology changes** versus **shape changes**, **transitions conformationnelles** versus **simulations de dynamique moléculaires**, **communication paths** versus **IDS**, etc...)

Gros travail de compréhension de la problématique biologique de la communication allostérique dans les protéines kinases et **d'extraction des problématiques mathématiques pertinentes** pour l'étude des dynamiques par **élimination des idées trop naïves ou des outils encore inadapté à ce stade** (difféomorphométrie 3D des surfaces).

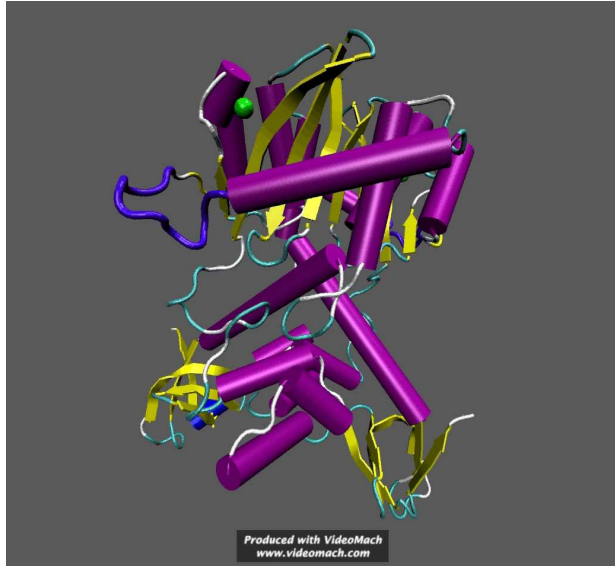
Création de communications entre les outils **de calcul et de visualisation** du LPBA et du CMLA (*VMD to Matlab* et *Matlab to VMD*) et mise en place d'une **plateforme commune de calcul** et de partage (Machine **Topdyn**, 32 cœurs + carte GPU) hébergé au CMLA Cournot.

Protein motion and Associated Allostery

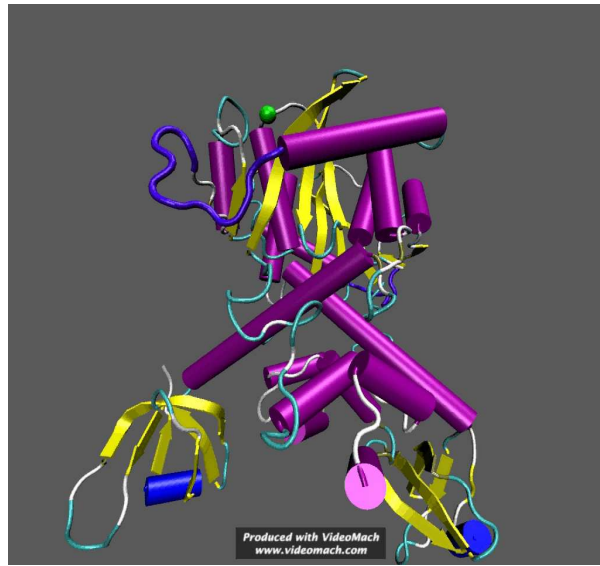
- Protein structure is a snapshot of one equilibrium state
 - Average of conformational microstates of native proteins
- Proteins are allosteric and achieve a wide range of motion to perform a biological function

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 Å
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 Å
Large-scale Motions • Domain Motion • Subunit Motion	Hinge-bending motion Allosteric transition	μ s to ms $10^{-6} - 10^{-1}$ s	5-10 Å
Global Motions • Helix-coil transition • Folding/unfolding • Subunits association	Hormone activation Protein functionality	ms to hours $10^{-1} - 10^4$ s	> 10 Å

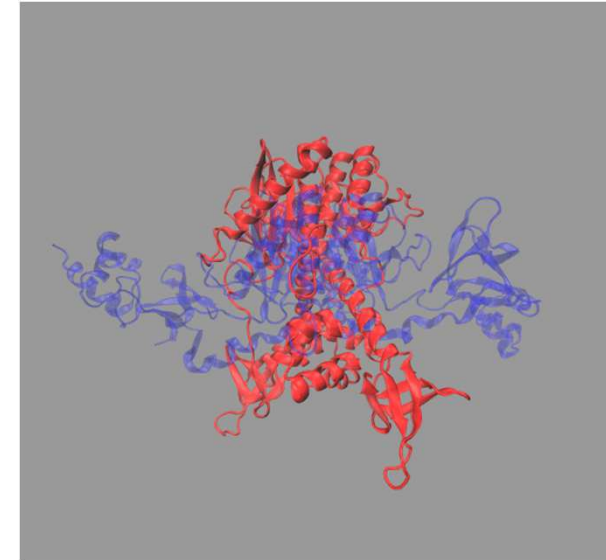
Simulations de Dynamique Moléculaire



Local Motions



Medium-scale Motions



Large-scale Motions

fs to ps

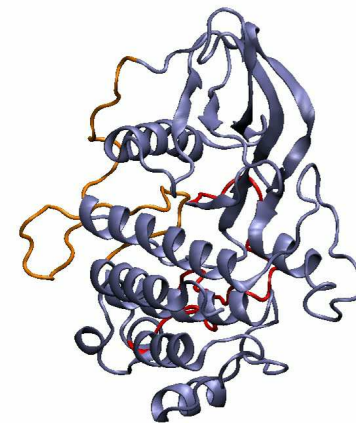
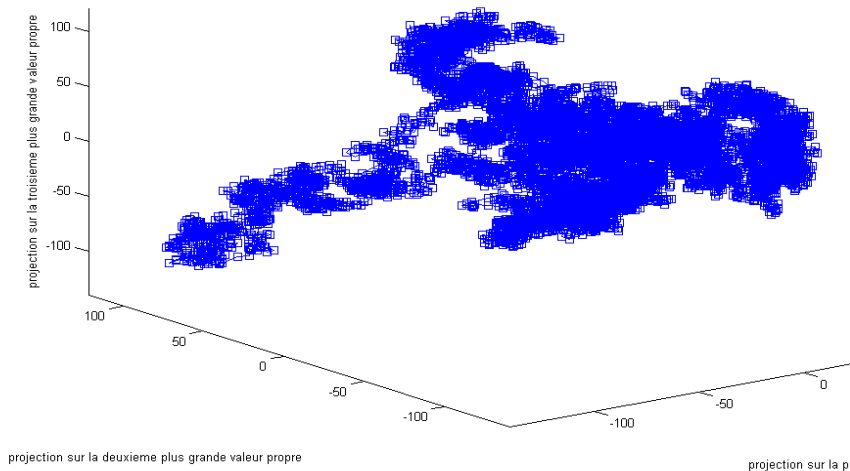
ns to μ s

μ s to ms

Transition conformationnelles versus DM: De la μ s à la ps

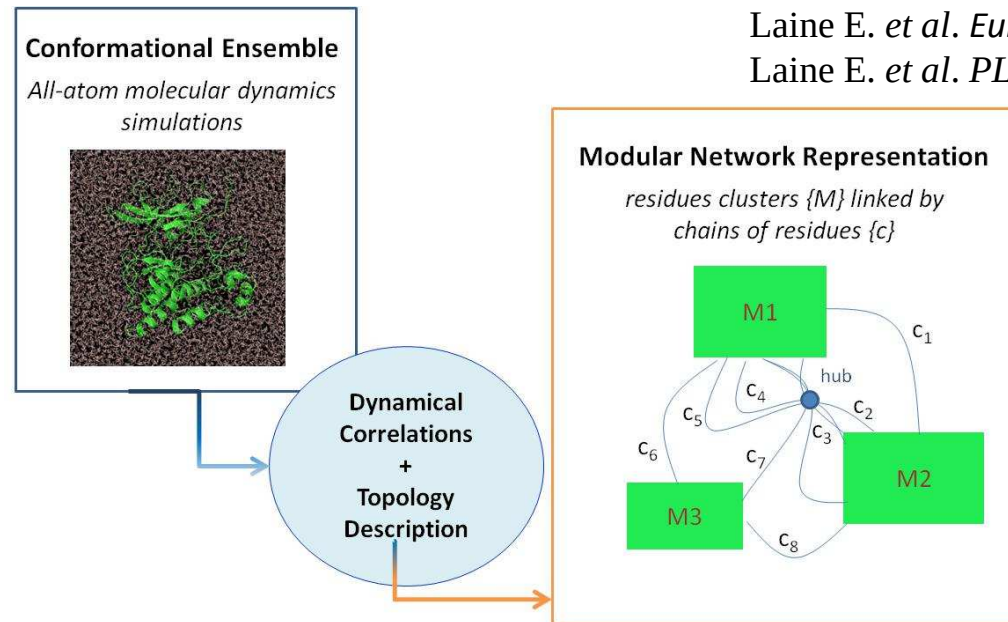
- Décrire la structure pour un grand nombre de conformations simulées (~1 500- 15 000 par objet)
 - Surface variable
 - Evolution des poches (espace vide).

Projection de la dynamique moléculaire (3 premiers axes)



- Décrire les voies de propagation d'une perturbation induite par effecteur (modulateur) à travers un système anisotrope et flexible a partir de simulations de Dynamique Moléculaire

MOdular Network Analysis (MONETA)

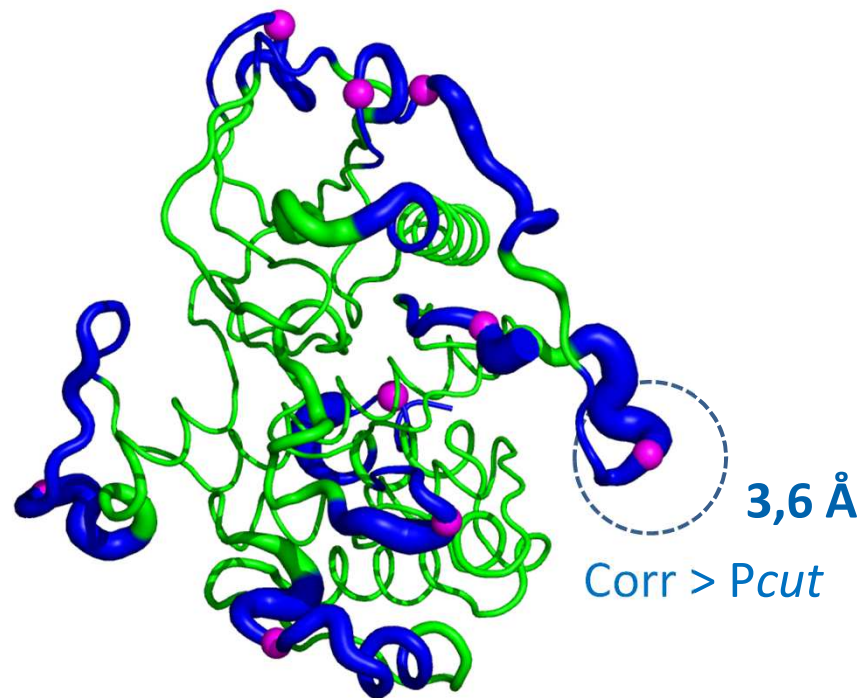


Laine E. *et al.* *Eur. Biophys. J. Biophys. Lett.* **2011**

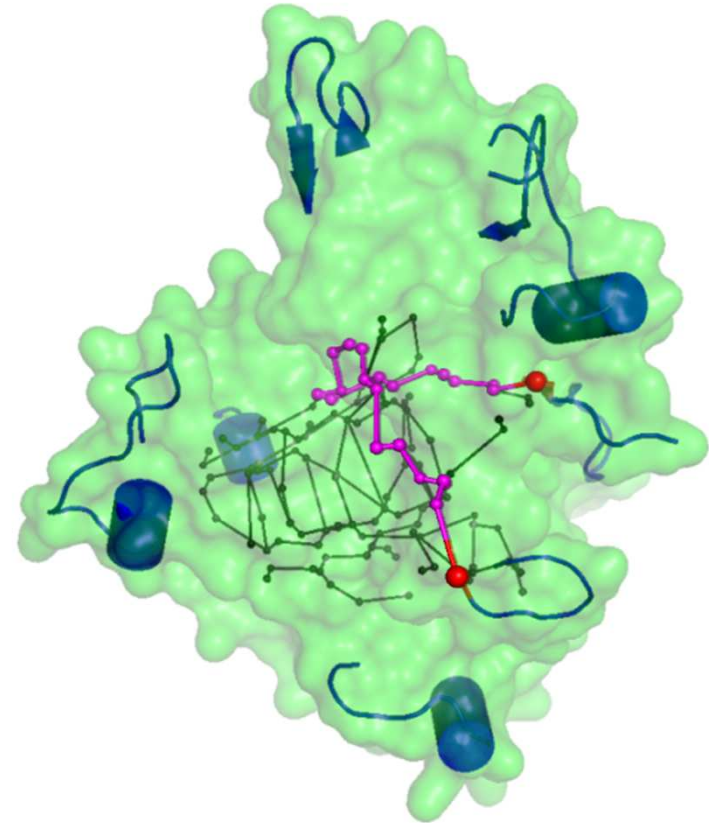
Laine E. *et al.* *PLOS Comp. Biol.* **2012**

- **Independent Dynamic Segments (IDSs)**, presenting the most striking local features of the protein internal dynamics are identified by a statistical technique *Local Feature Analysis (LFA)*.
- **Communication pathways (CPs)** are determined using the concept of communication propensity. The CPs are grown ensures that any two adjacent residues are connected by non-covalent interactions and that every residue in the CP is connected to any other point by a short **commute time (CT)**.

Independent Dynamic Segments (IDSs)



Communication pathways (CPs)



Chauvot de Beauchêne *et al.* **PLoS Comp. Biol.** (2013). submitted at 18 Nov2013

Recherche des IDS : PCA versus LFA

Independent Dynamical Segments:

Sources *localisées* et *indépendantes* de variabilités de la dynamique
Manque d'une conceptualisation vraiment opérationnelle.

Composantes indépendantes mais non localisées par l'ACP

L'ACP fait émerger des facteurs de variabilités non locaux (projection sur directions principales). *Inadaptée* pour la recherche des IDS.

Local Feature Analysis (LFA) (Zhang et Wriggers, 06)

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 IDS: Une nouvelle approche par PFA

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

Principal Feature Analysis comme
recherche de sous-espaces
optimaux de prédicteurs *constraints*

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) :**

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

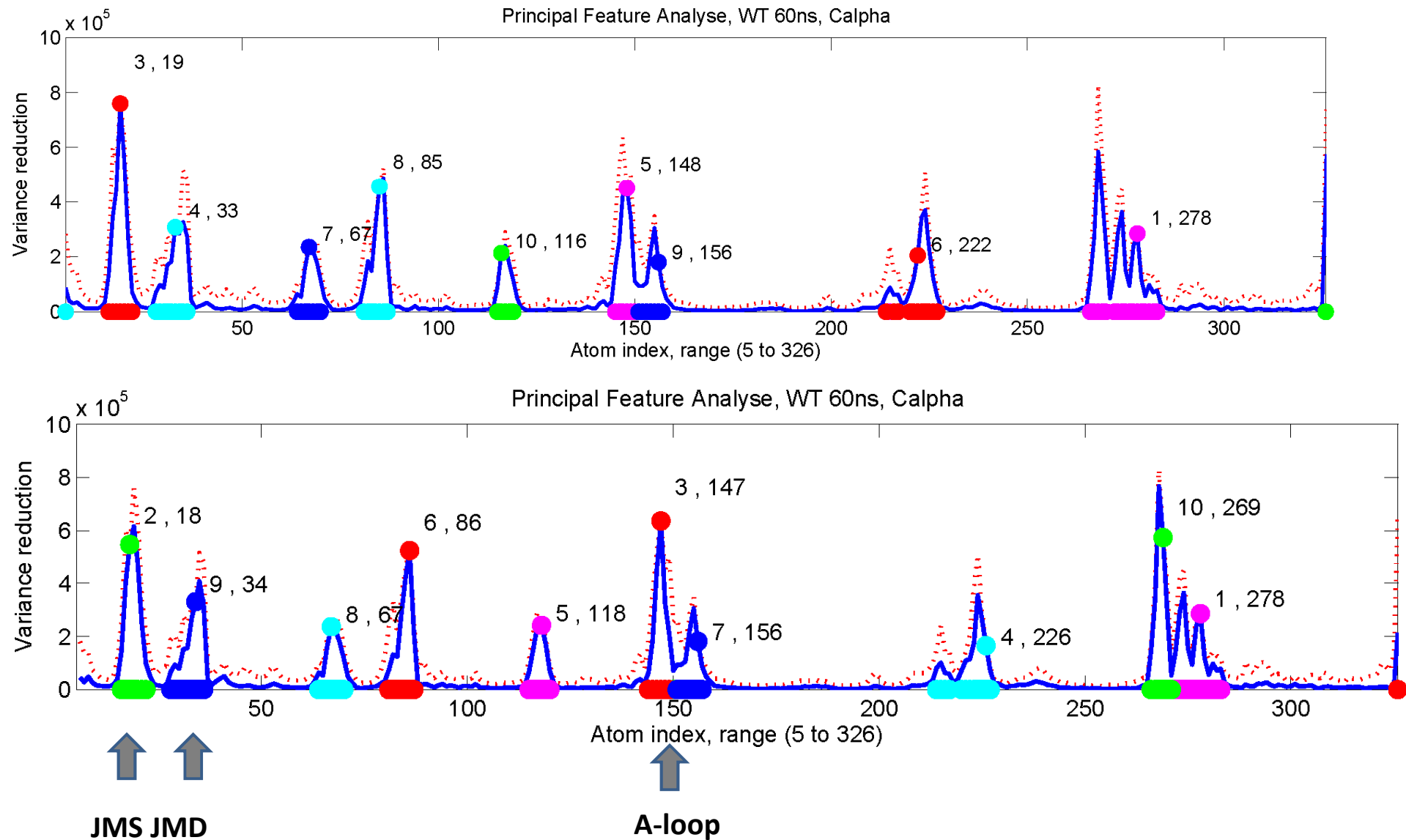
E_j : coordonnées du $j^{\text{ième}}$ atome C_α

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

Algorithme glouton :

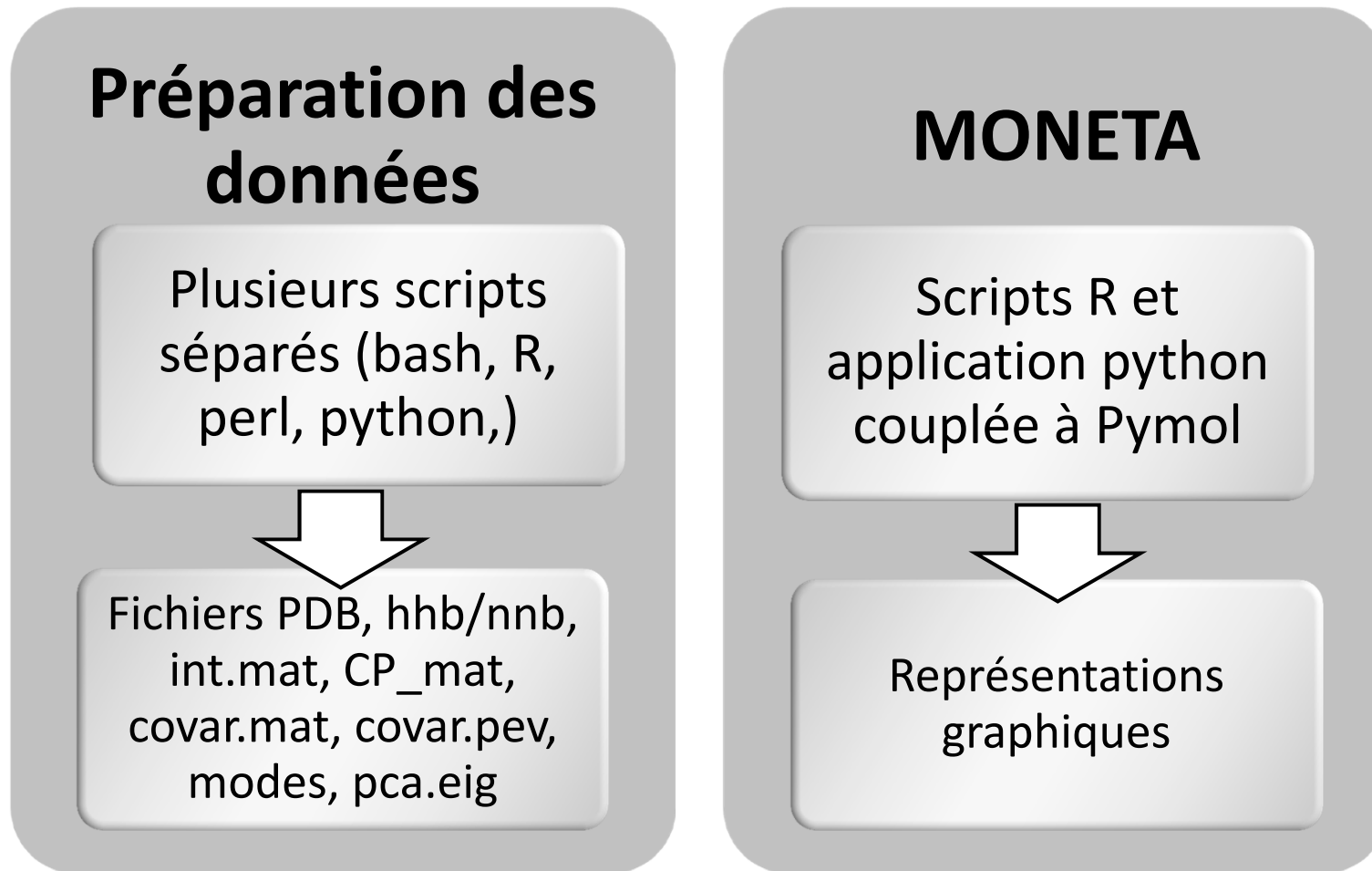
$$\hat{j}_1, \hat{j}_2, \cdots, \hat{j}_k$$

Principal Feature Analysis sur la DM du KIT WT



Optimisation de l'application MONETA

Structure



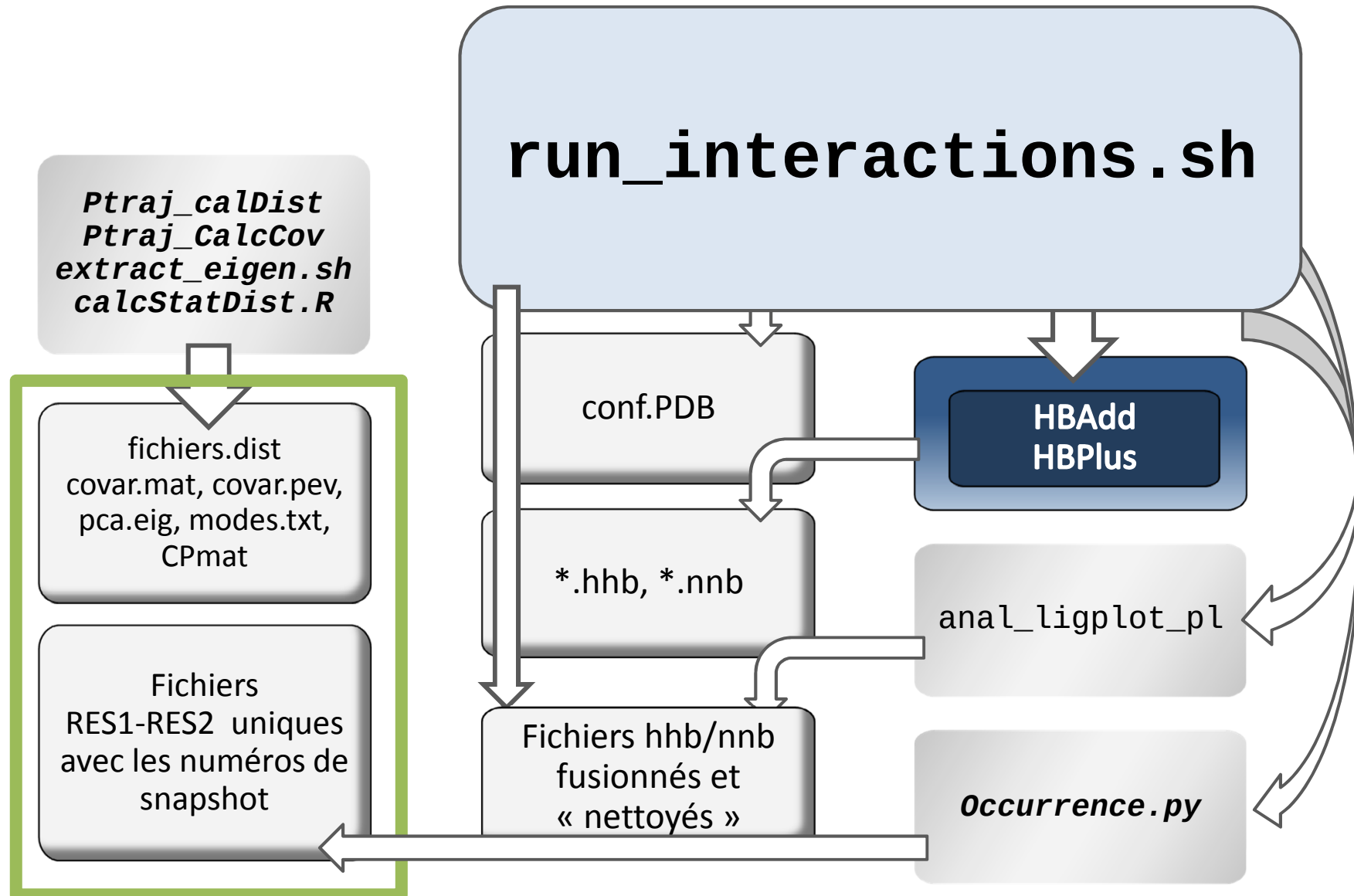
Objectives et Résultats obtenu en 2013

- Optimization of the MONETA application

« A la recherche du temps perdu »

- Parallel computing
- Graphical representation

SOLUTIONS APPORTEES:



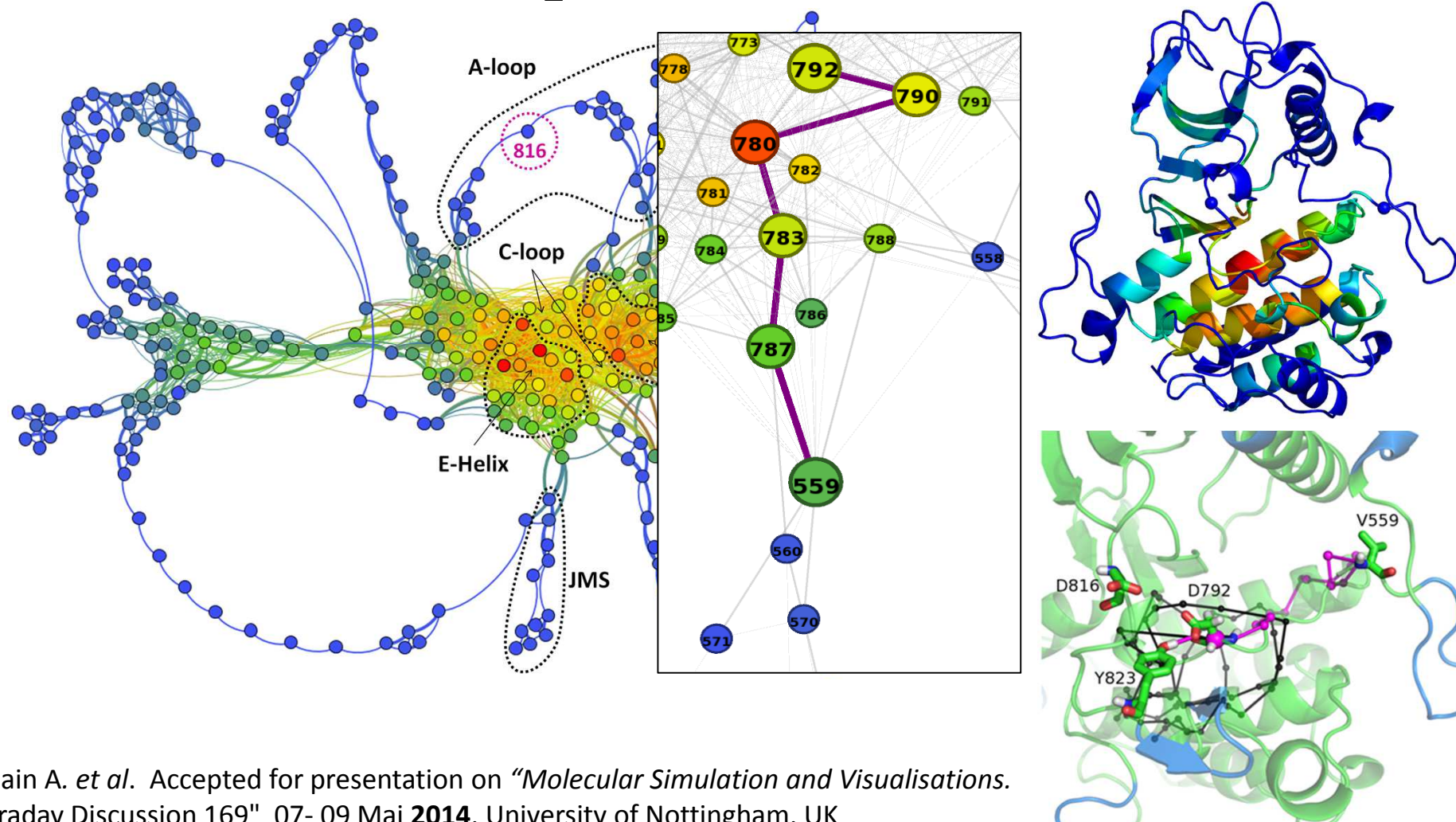
SOLUTIONS APPORTEES: Temps Estimes

	AVANT	WT dt10 6000 frames	WT complet 60 000 frames
Extraction des PDB	Plusieurs jours à plusieurs semaines	≈ 2 minutes	≈ 12 minutes
Création des fichiers avec les liaisons pour chaque snapshot		≈ 75 minutes	≈ 505 minutes
Création des fichiers interactions résidus- résidus		≈ 18 secondes	≈ 10 minutes
TEMPS TOTAL		≈ 80 minutes = 1h20	≈ 524 minutes = 8h45

- computation : performance is improved by factor 30;
- performance/history data (disc space consumption) is improved by factor 10.

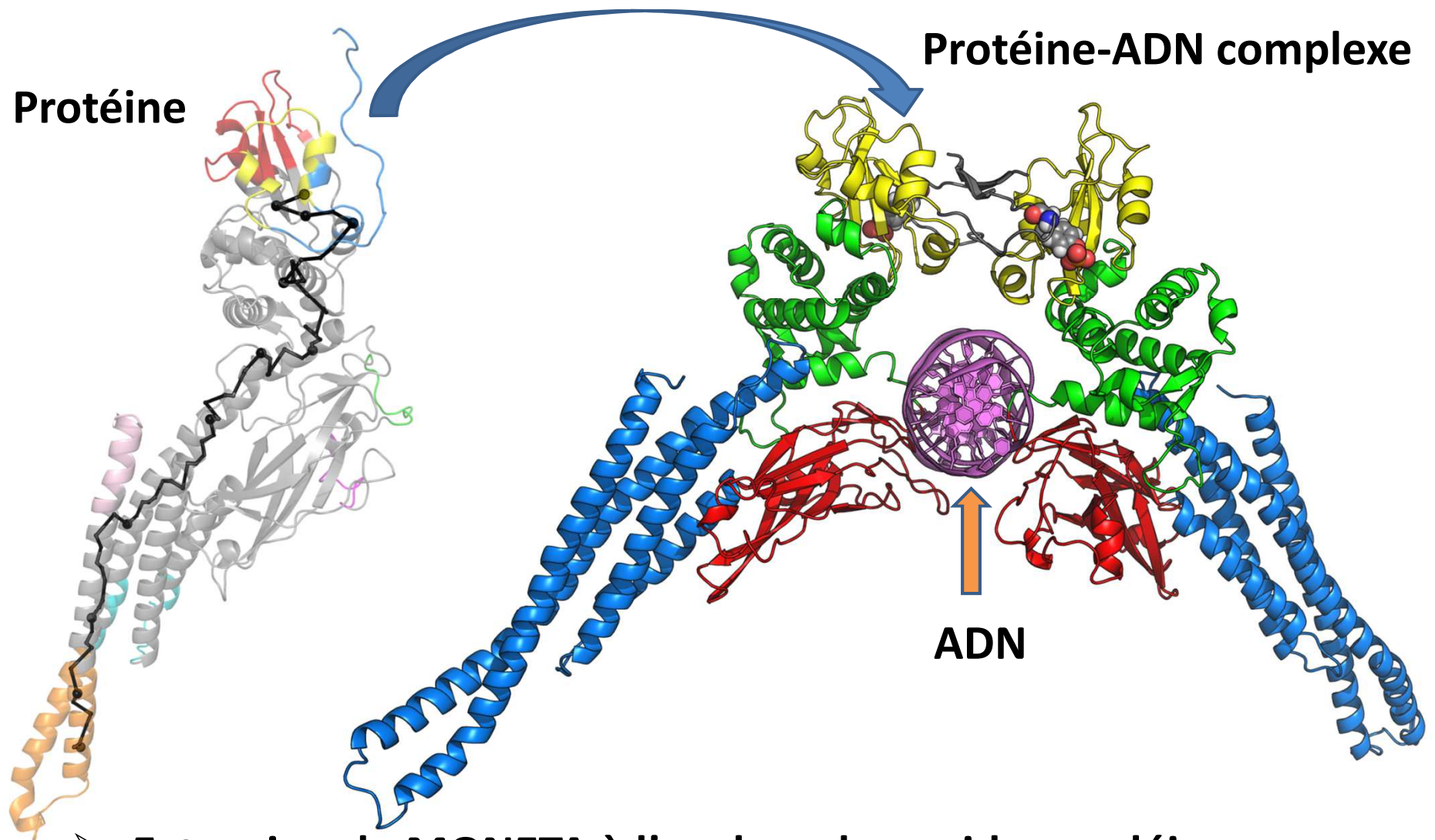
AUTRES APPORTS: 2D- and 3D- graphical representations

➤ « *Dessine-moi une protéine* »



Allain A. *et al.* Accepted for presentation on "Molecular Simulation and Visualisations. Faraday Discussion 169" 07- 09 Mai **2014**, University of Nottingham, UK

En cours de développement:



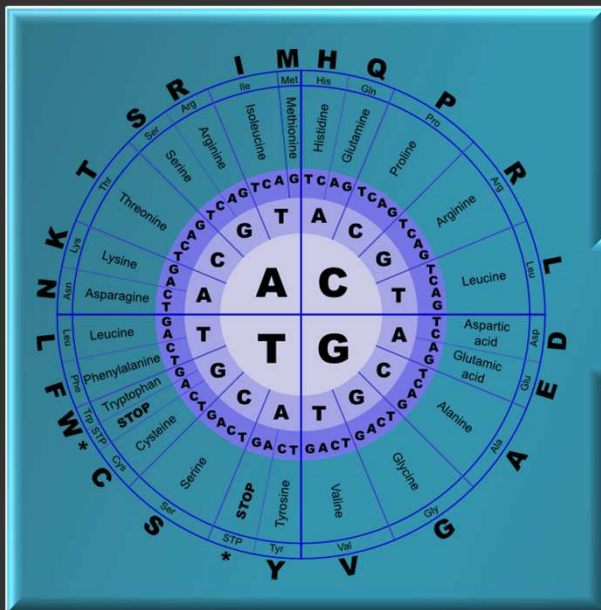
➤ **Extension de MONETA à l'analyse des acides nucléiques**

Short-time project

- The Web-Based interactive platform tools, the first utility to investigate allosteric effects/mechanisms in biological macromolecules

Long-time projects

MONETA + Genetic code



- Forecasting of mutations able to prompt protein dysfunction

- “Interactome” of the RTKs → Cell Signaling Pathways