

**UNIVERSITÉ
PARIS-SUD 11**

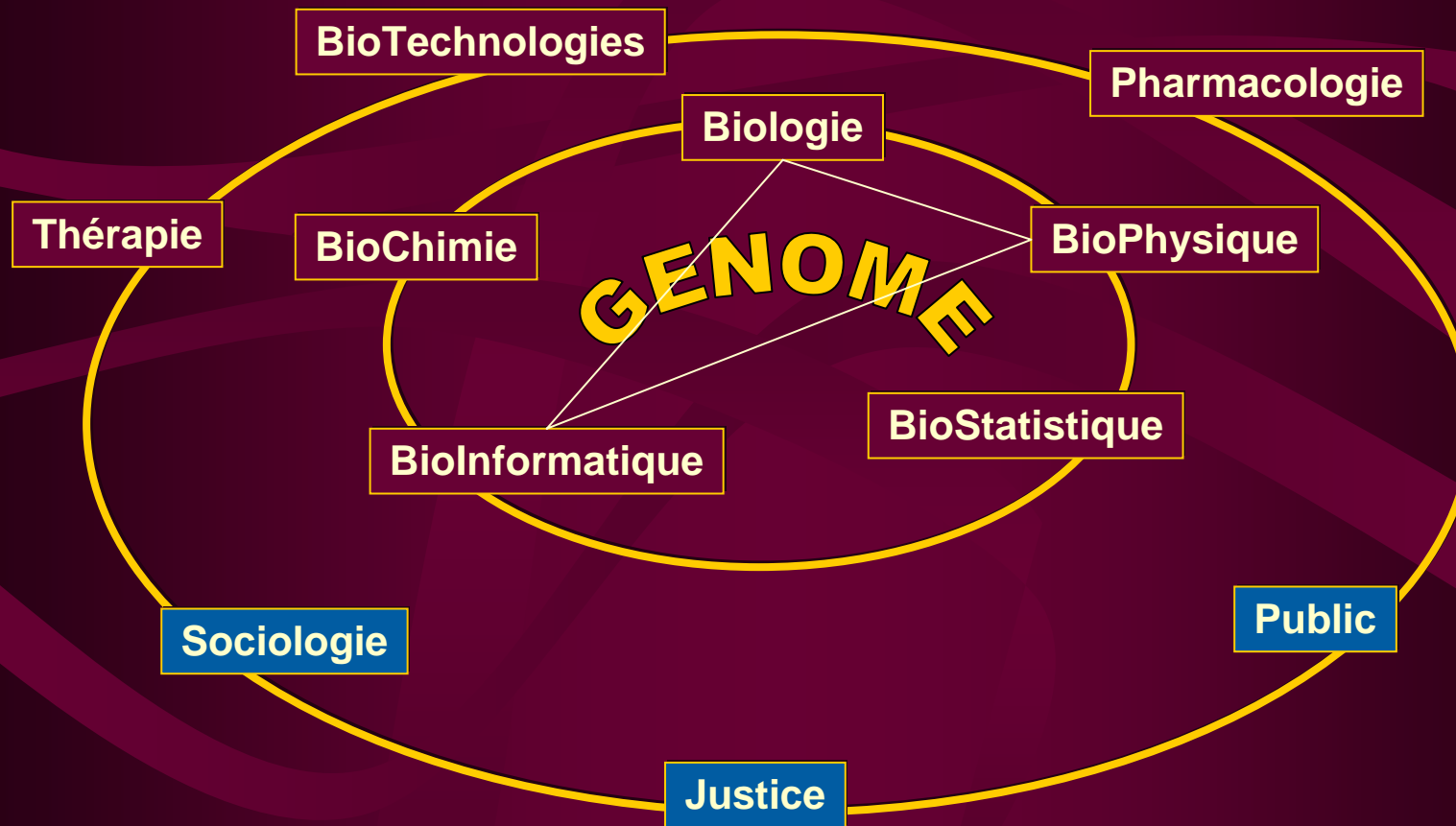


**Représentation spatiale des génomes/protéomes
et
Exploration virtuelle de données
factuelles et textuelles en biologie moléculaire**



Génome

objet scientifique et sociétal



Génome : Objet Scientifique & Societal

- Objet de discussion collective
 - Entre spécialistes
 - Entre multi-spécialistes
 - Entre spécialistes et non spécialistes
- Objet en construction (révision) permanente
 - Faits
 - Points de vue
 - Publications
 - Communiqués
- Objet médiatique
- Informations factuelles
 - Séquences annotées (acides ADN, ARN, Protéines...)
 - Structures Spatiales (double hélice ADN, Protéines...)
 - Données d'expression (Images Transcriptome et Protéome, Puces à ADN,...)
- Informations textuelles
 - Publications scientifiques
 - Documents spécialisés
 - Documents généralistes

Objets Complexes Médiatiques et Médiatisés

de Taille immense, Formes variées, Etablies/Personnelles, Réparties...

Difficilement accessibles au(x) utilisateurs(s)

Exploration Immersive : Pistes...

- Immersion
 - Représentations multi-niveaux, communes et homogènes
 - Documents, Contenus, Objets biologiques...
 - Exploration cohérente et continue
 - Interaction, Navigation...
 - Liens entre factuel et textuel
- Problématiques
 - Clusterisation/Traitement textuel (datamining)
 - Gestion de scène (visualisation)
 - Paradigmes et patterns d'interaction H-M
 - ...
- Evaluations?

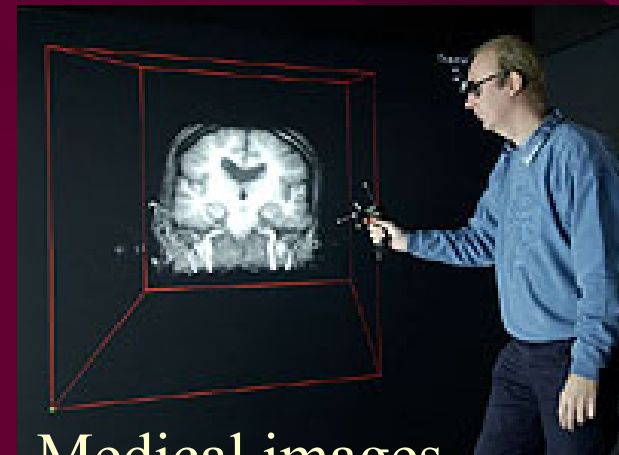
Exemples



Multi disciplinary Discussion of complex datasets

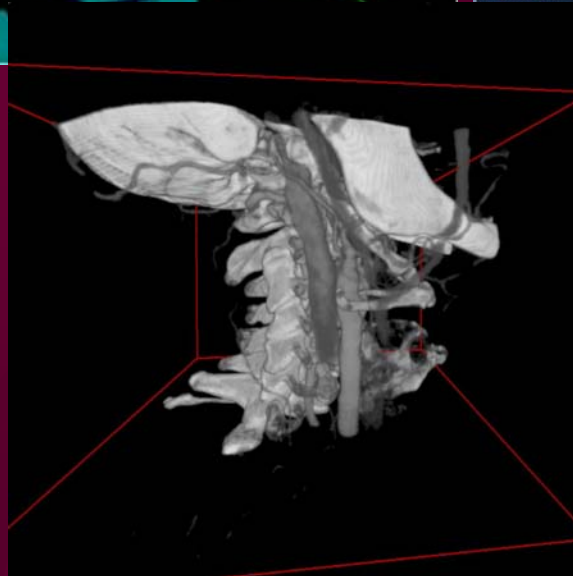
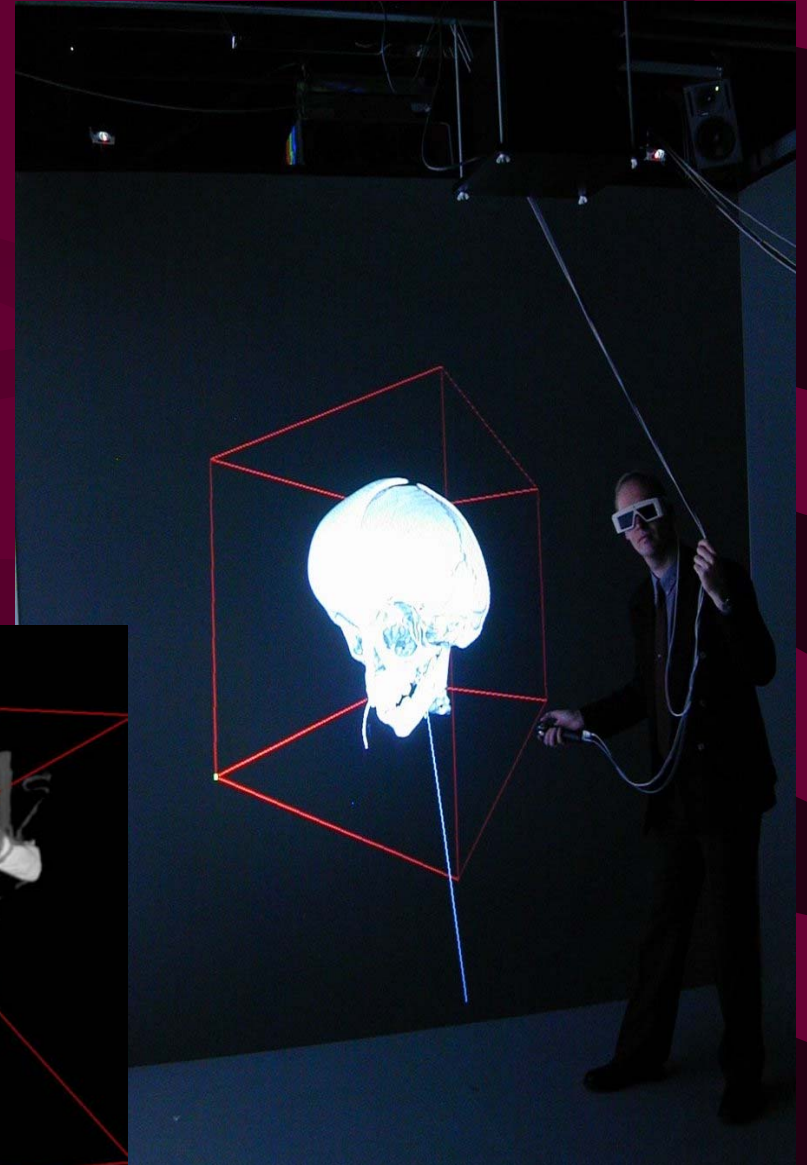
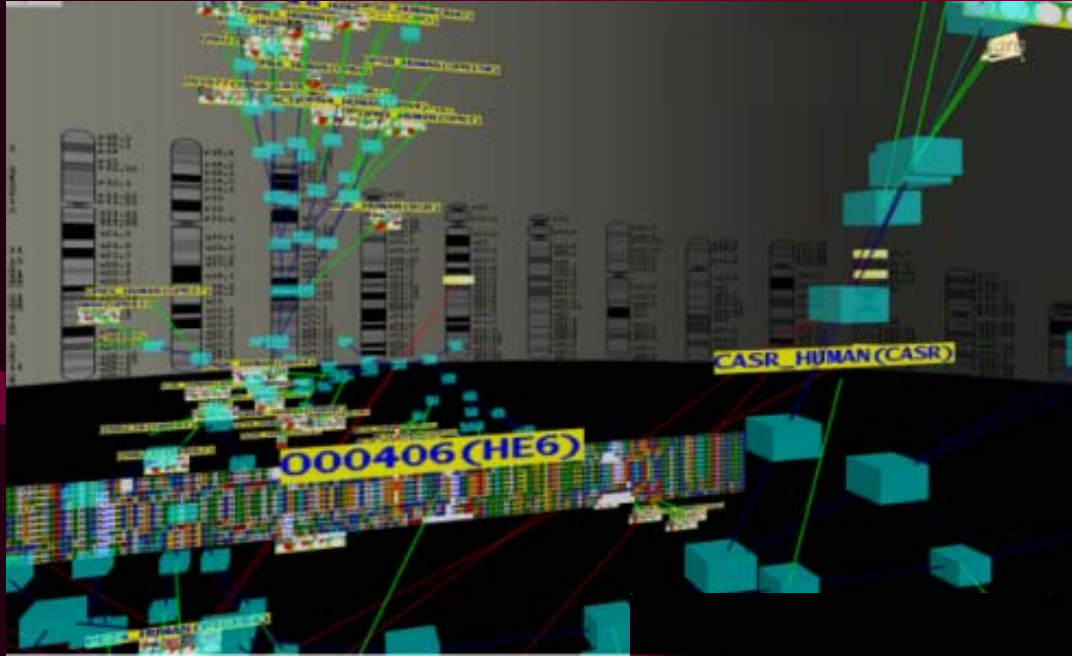


Barco Stereoscopic visualization techniques
and immersive data analysis
to support cancer research
at Erasmus MC, the Netherlands



Medical images

Exemples



Représentation spatiale des génomes

Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique

Outils d'analyse

Applications biologiques

Démonstration *ADN-Viewer*

Problématiques

- **Informatique**
 - Gestion de grosses masses de données
 - Visualisation
 - Interface (bureau, immersif)
 - Performances
 - Outils
- **Génomique**
 - Représentation pertinente des données
 - Outils complémentaires
 - Applications à des problèmes biologiques
- **Bioinformatique**
 - Nouveau point de vue
 - Approche nouvelle et alternative

Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique

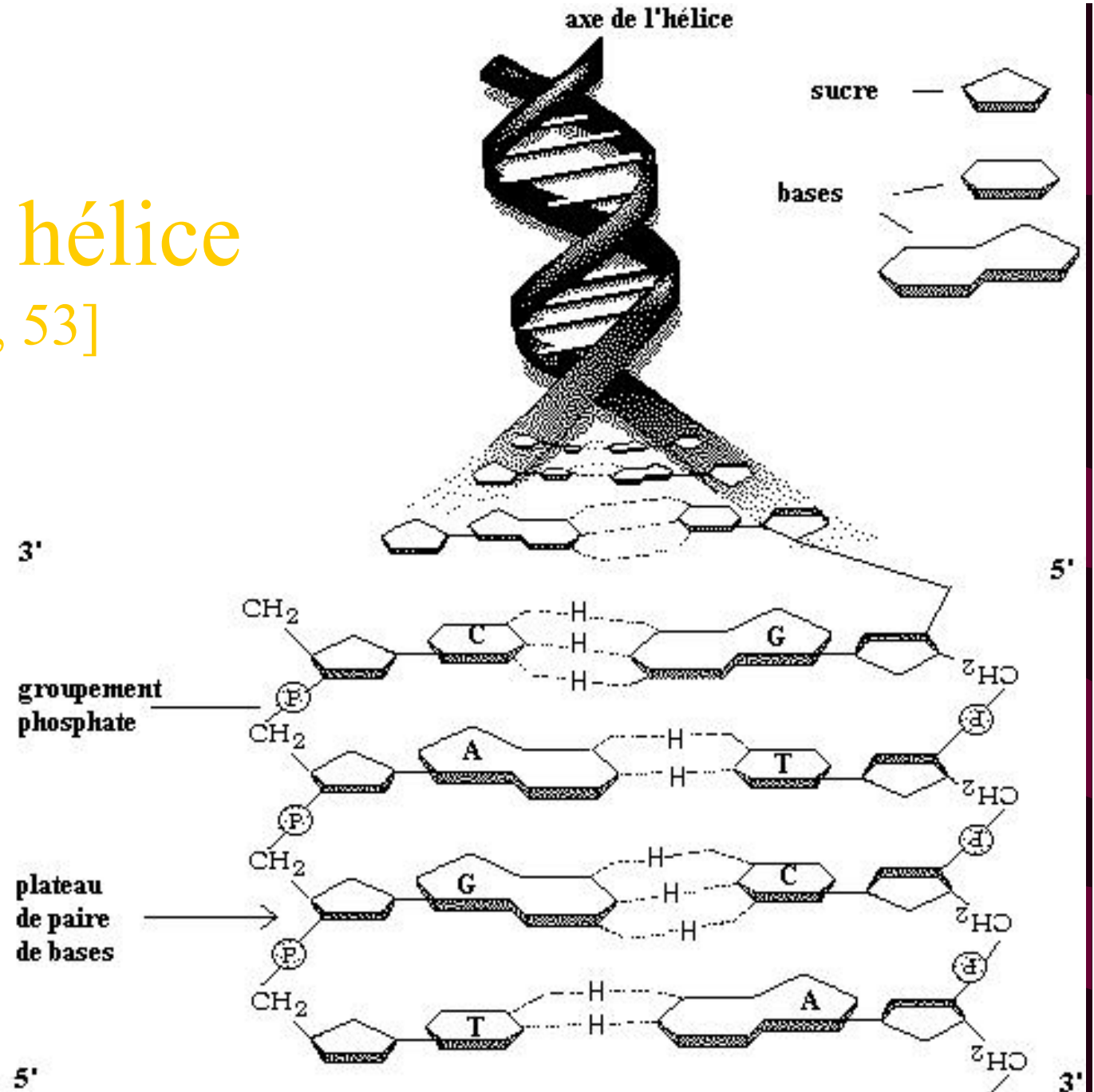
Outils d'analyse

Applications biologiques

Démonstration *ADN-Viewer*

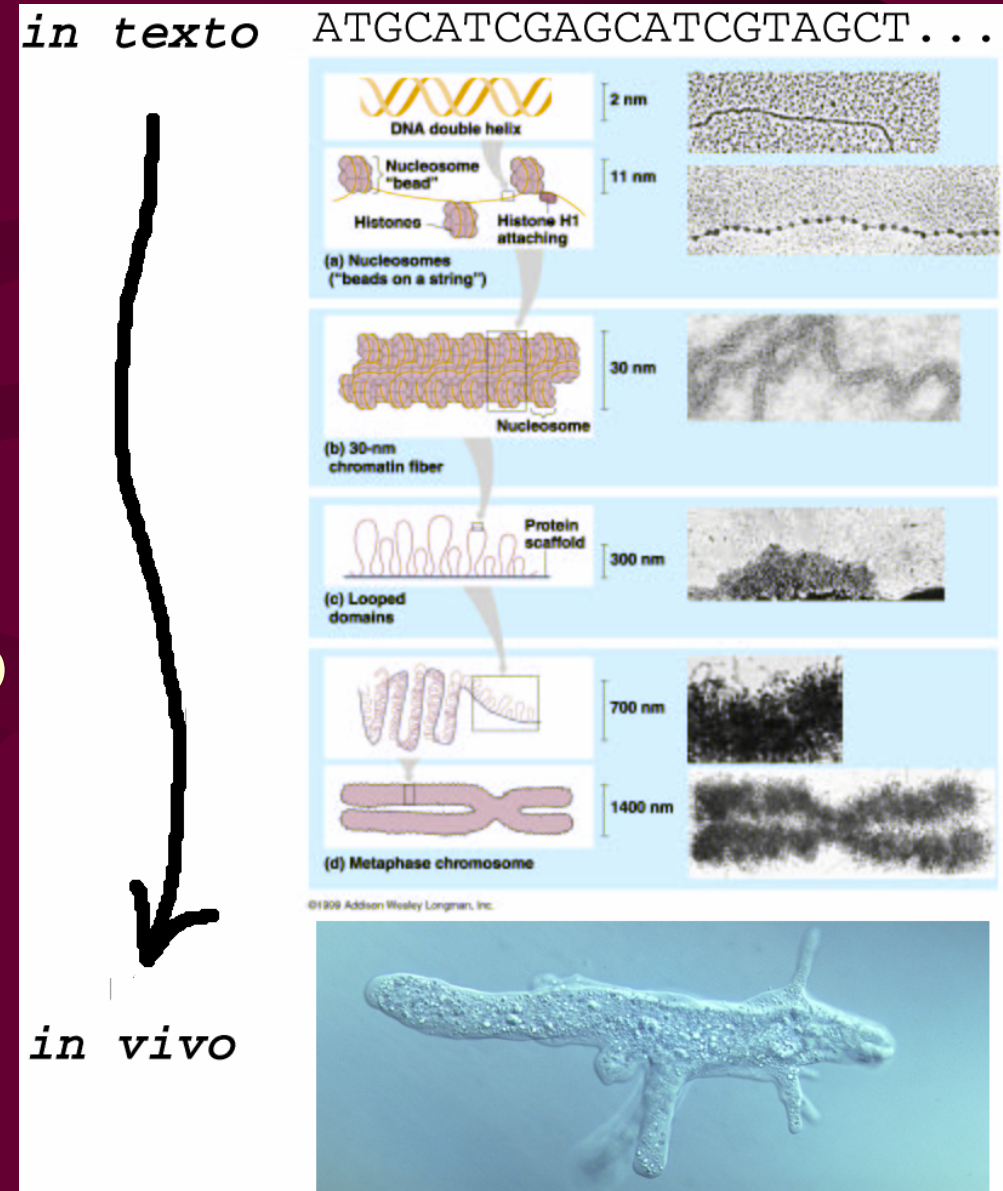
L'ADN : une double hélice

[Watson & Crick, 53]



L'ADN dans tous ses états

- **ADN *in vivo*** (4 dimensions)
 - Chromatine
 - Observation difficile
- **ADN *in texto*** (1 dimension)
 - Succession de lettres (A, C, G, T)
 - Représentation éprouvée avec succès
 - Réalité chimique
- **ADN *in visio*** (3 dimensions)
 - Visualisation 3d



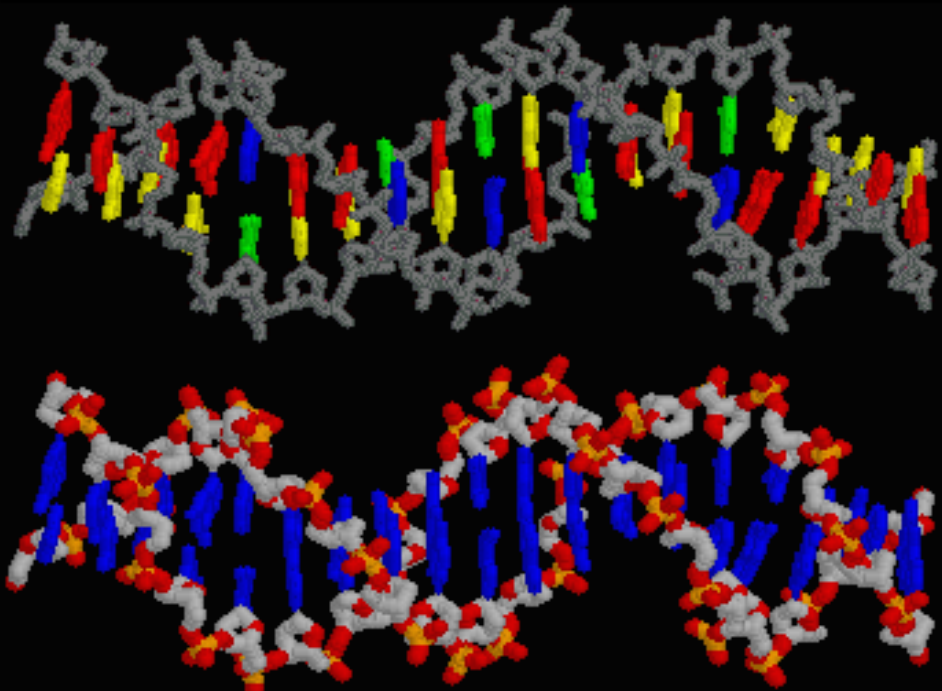
Visualisation en bioinformatique

RasMol [Sayle, 93]

Observation

DNA

A T C G

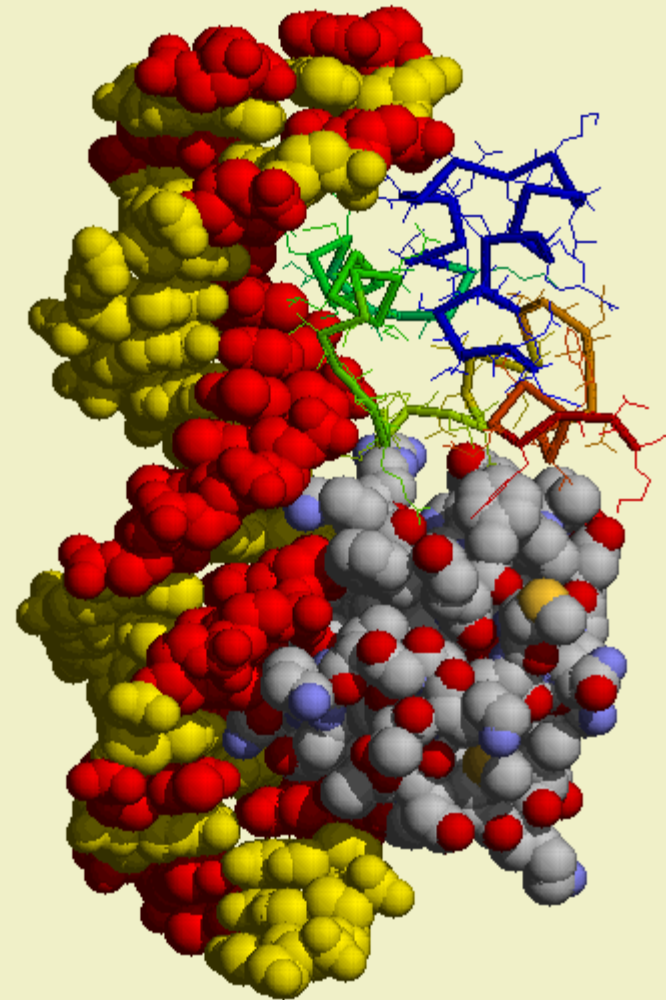


Helix Backbone

Carbon Oxygen Phosphorus

RasMol

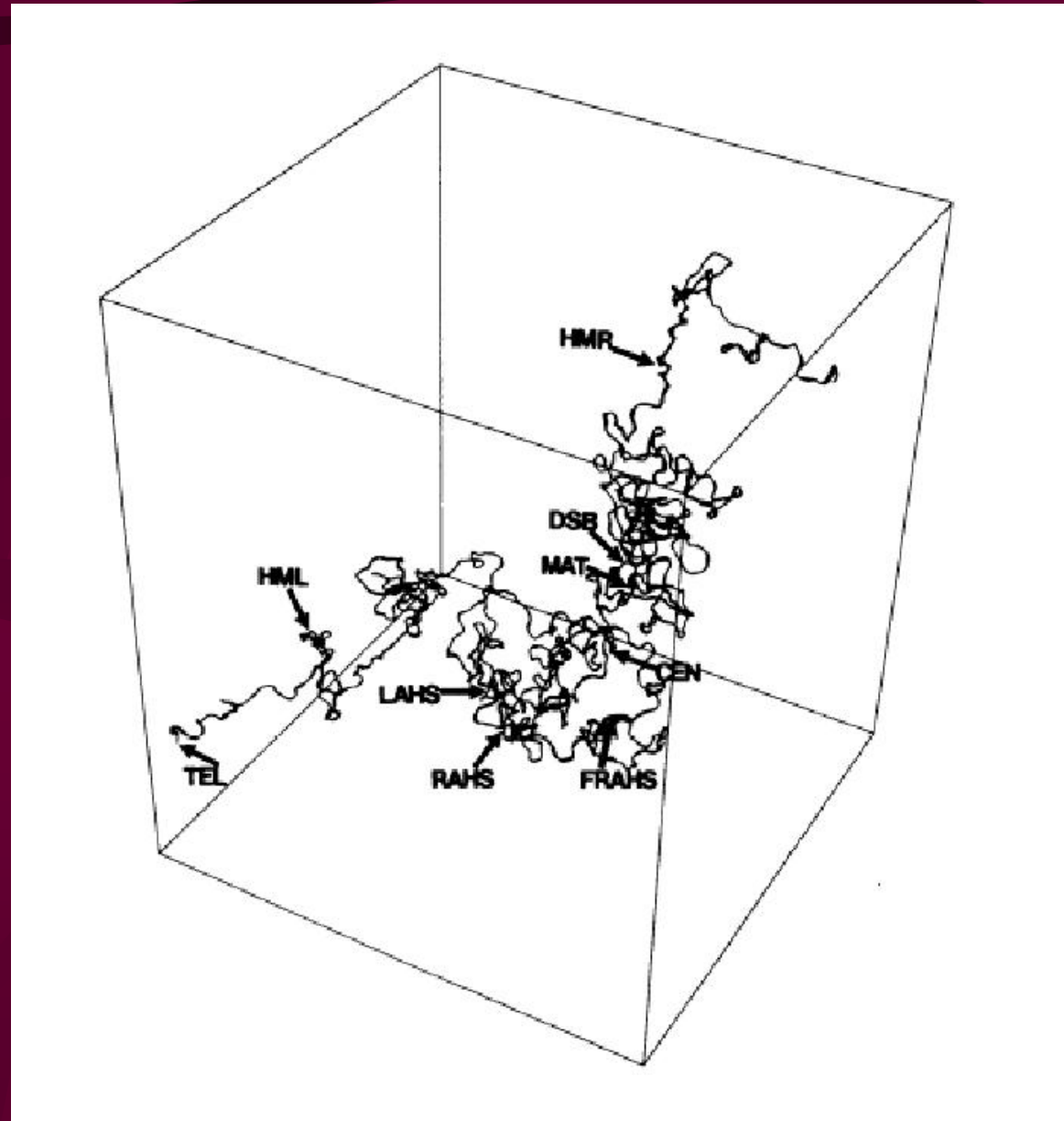
Phage CRO Repressor on DNA. Andrew Coulson & Roger
Sayle with RasMol, University of Edinburgh, 1993



Visualisation en bioinformatique

Prédiction

[King, 93]



Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique

Outils d'analyse

Applications biologiques

Démonstration *ADN-Viewer*

Le moteur géométrique

DONNEES EN ENTREE

Fichier séquence

Règles de construction 3d [Shpigelman et al., 93]

```
ACCATGGCT
GCGCTATGC
ATTTCGCGT
AATT.....
```

```
AA =  $\theta_x, \theta_y, \theta_z$ 
AC =  $\beta_x, \beta_y, \beta_z$ 
AG =  $\lambda_x, \lambda_y, \lambda_z$ 
:
Trans = 3.39 Å
```

Représentation interne

Chaîne
de
caractères

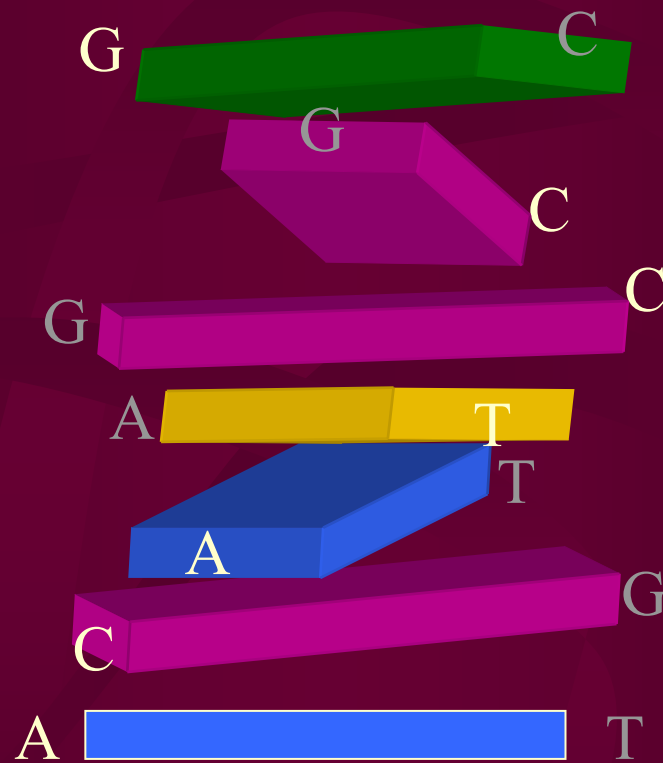
...CCAT

Matrices
de
transformations

RESULTAT
EN
SORTIE

Coordonnées
3d

Le moteur géométrique

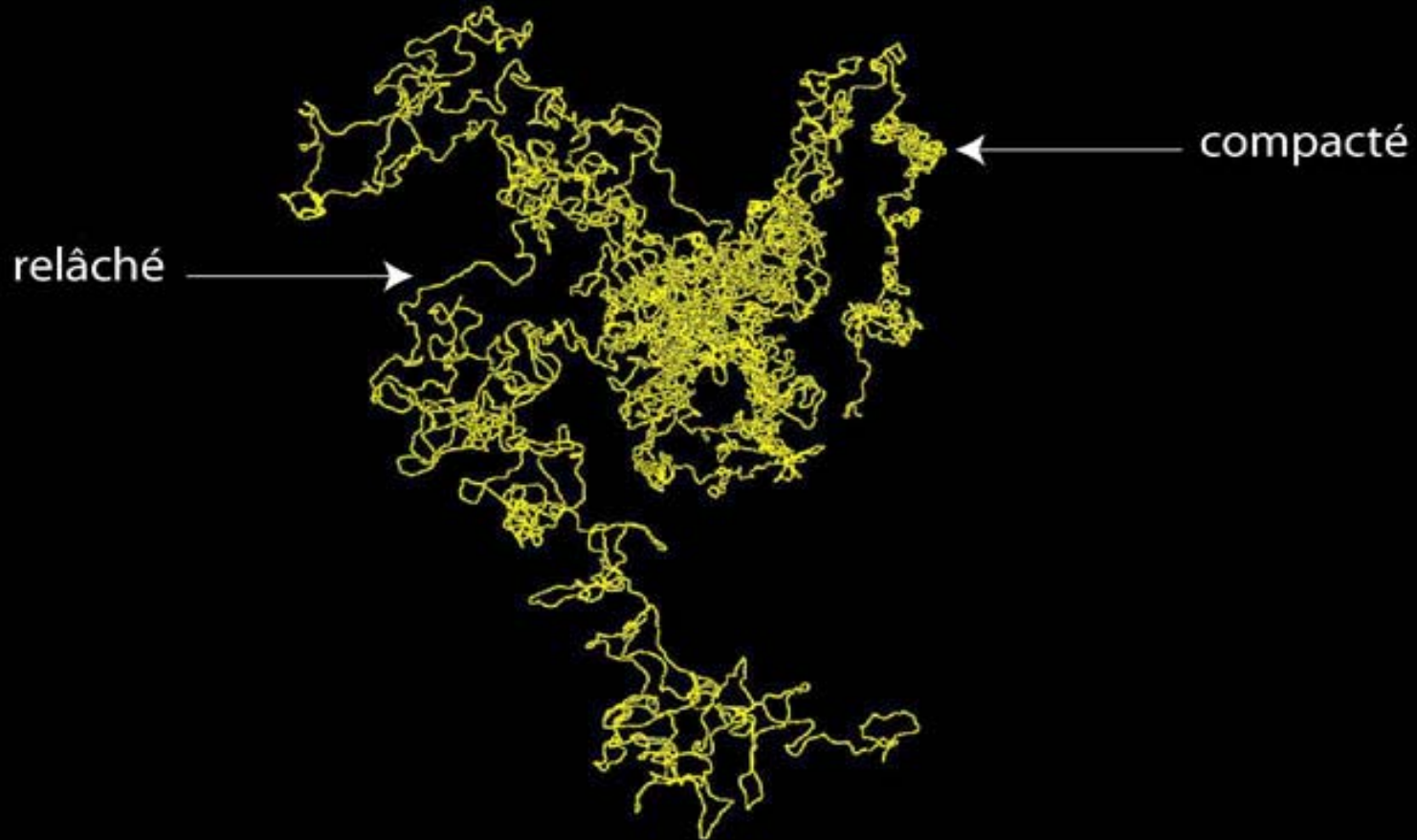


Visualisation de l'ADN

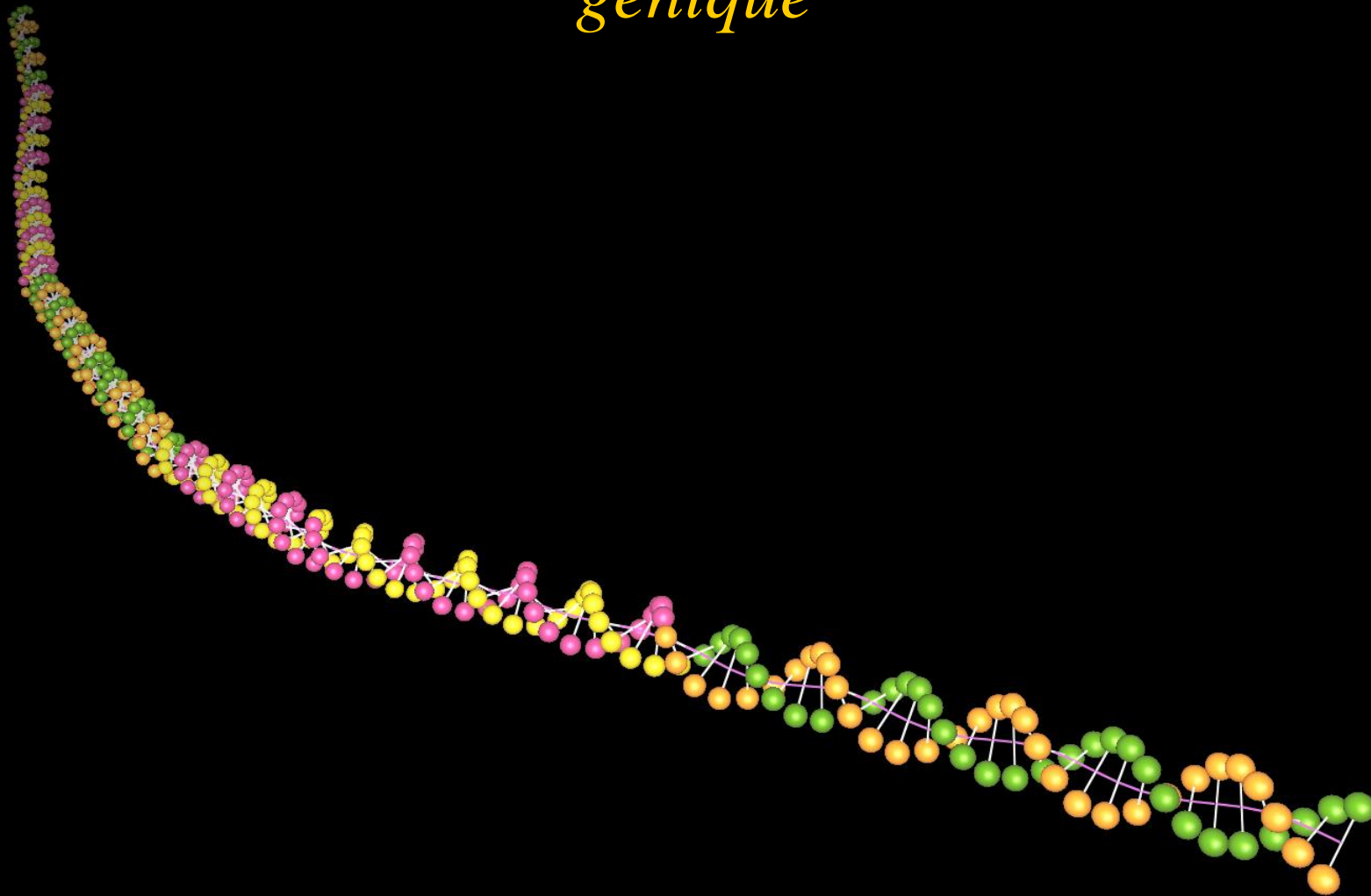
ADN-Viewer

- **Chromosome**
- **Gène**
- **Nucléotide**
 - Structure PDB
- **Génome**

Visualisation de l'ADN *génomique*

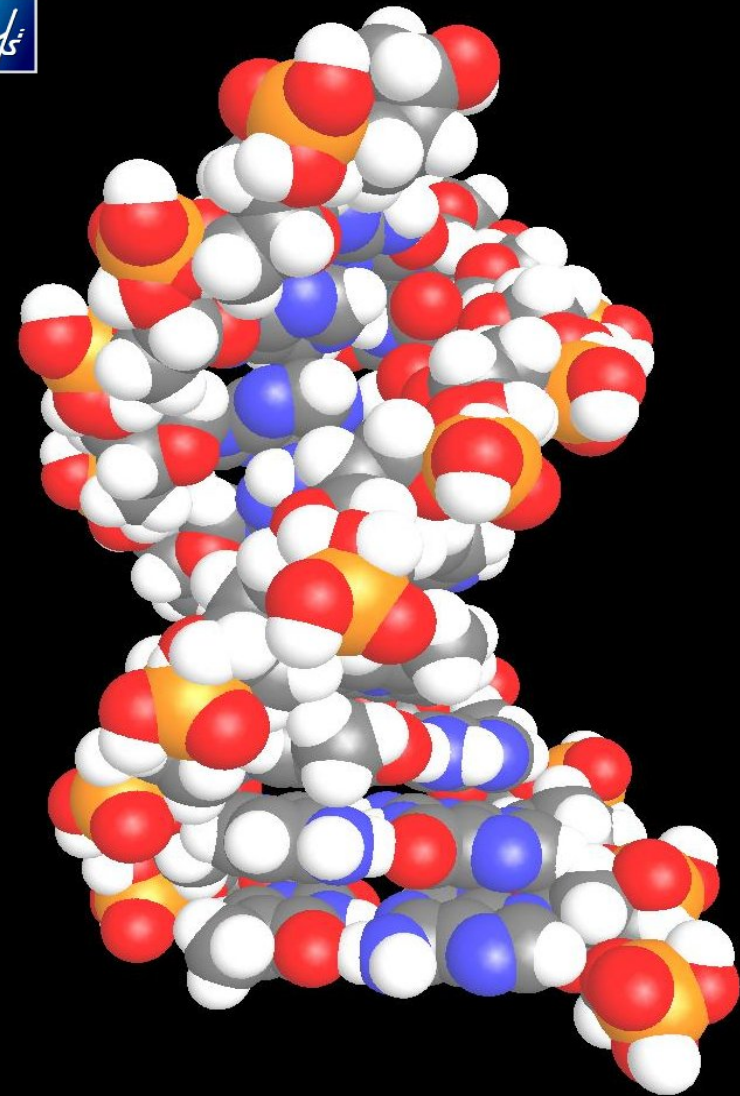
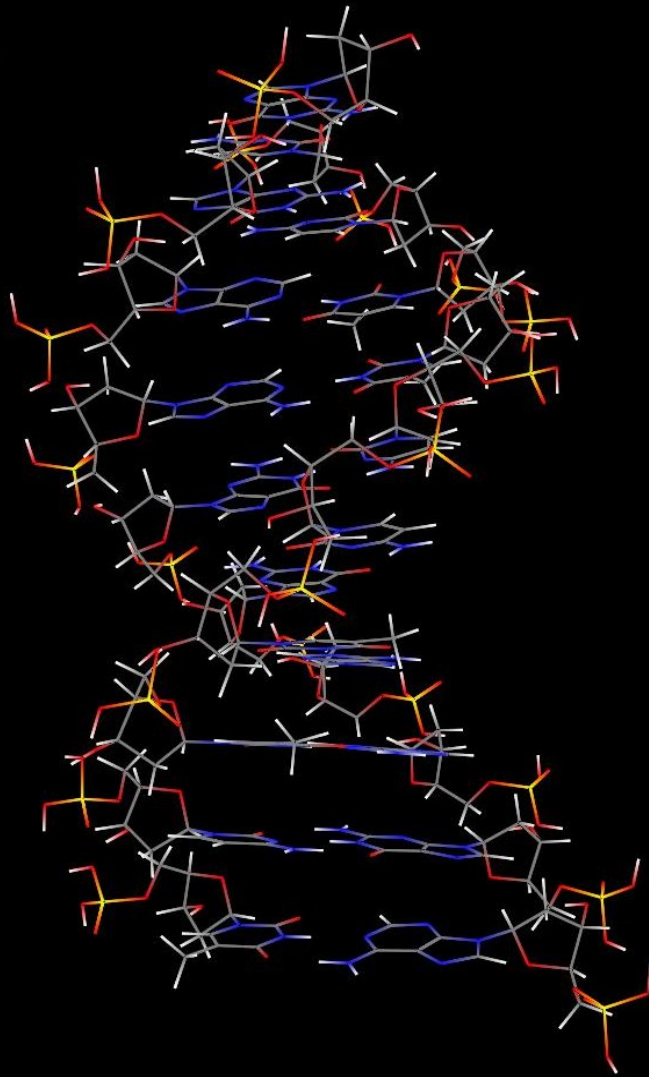


Visualisation de l'ADN *génique*



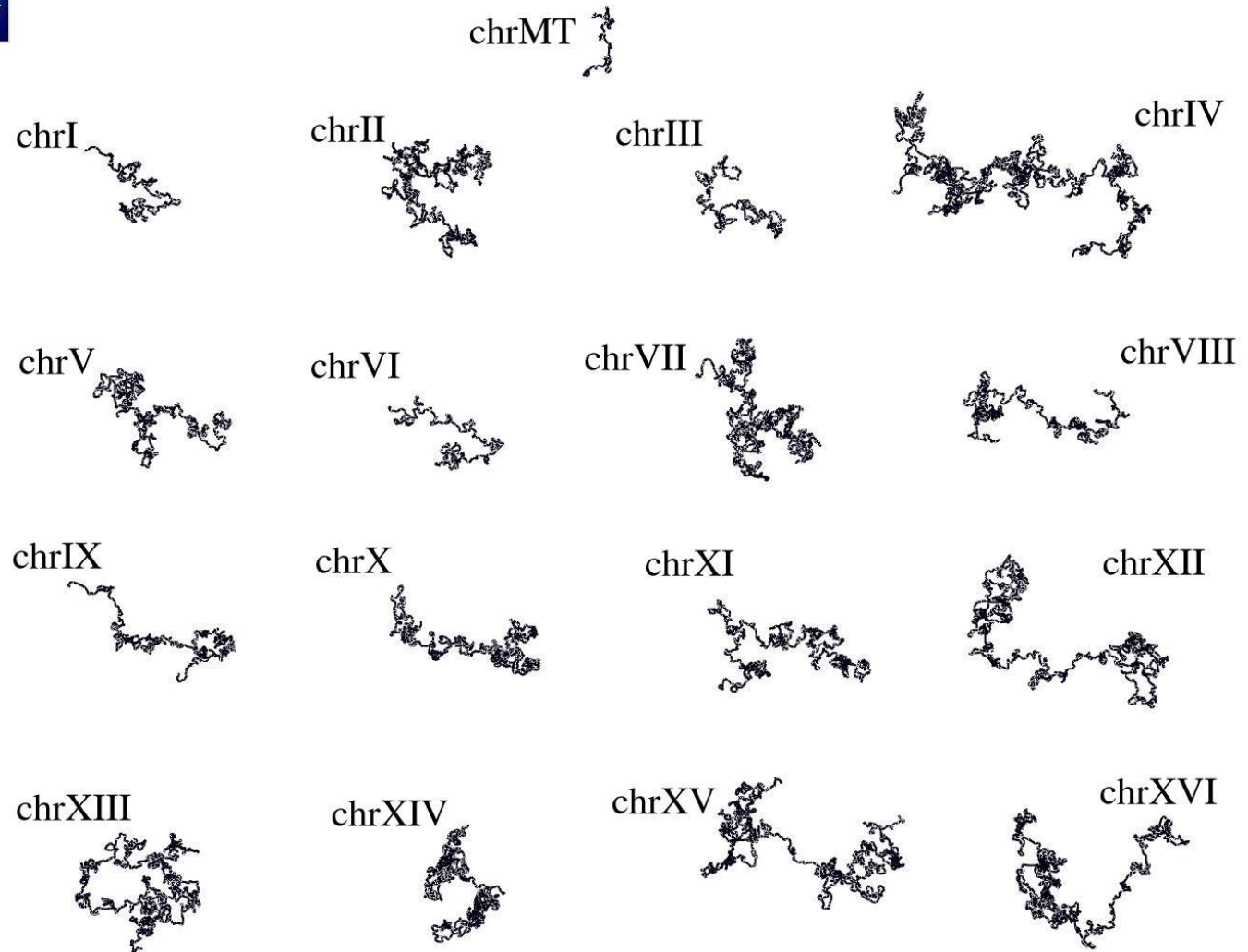
Front Perspective

Visualisation de l'ADN *nucléique*



Visualisation de l'ADN

S. cerevisiae



Structures de données *stockage*

- **Temps interactif** (> 15 ips)
- **Chromosome 5 Mpb**
 - $5.10^6 \times (4+3) \times 4 = 140.10^6$ opi = $2,1.10^9$ ops
- **Pré-calcul et stockage des coordonnées 3d**
 - Position : point 3d
 - Orientation : quaternion
- **5 Mpb $\rightarrow 5.10^6 \times (4 \times 8 + 4 \times 4) = 240$ Mo**

Structures de données

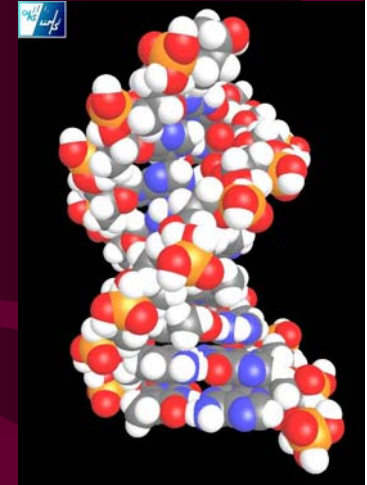
filtrage

- **Temps interactif** (> 15 ips)
- **Chromosome 5 Mpb**
 - 5 M segments/i = 75 M segments/s
- **Filtrage de l'information**
 - Statique uniforme
 - Dynamique uniforme
 - Dynamique non-uniforme



Front Perspective

Structures de données *optimisation*



- **Temps interactif** (> 15 ips)
- **Représentation nucléique**
 - 70 sphères/pb = $70 \times 500 = 35000$ triangles/pb
 - 20 pb $\rightarrow 700 \cdot 10^3$ triangles/f = $10,5 \cdot 10^6$ triangles/s
- **Détection des faces cachées**
 - $(70 \times 500)^2 \approx 1 \cdot 10^9$ calculs de distance
 - Structure physico-chimique

L'ADN *in virtuo* *en environnement immersif*

- **Espace de visualisation plus grand**
- **Meilleure perception profondeur**
 - Stéréoscopie
- **Interaction plus intuitive**
 - Navigation (EVI3d [Bourdot&Touraine, 02])
 - Interaction mains-libres
 - Souris 3d
 - Désignation, sélection
- **Incitation à l'exploration intuitive des données**

L'ADN *in virtuo* Muse



L'ADN in virtuo
navigation mains-libres (Evi3d)



Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique

Outils d'analyse

Applications biologiques

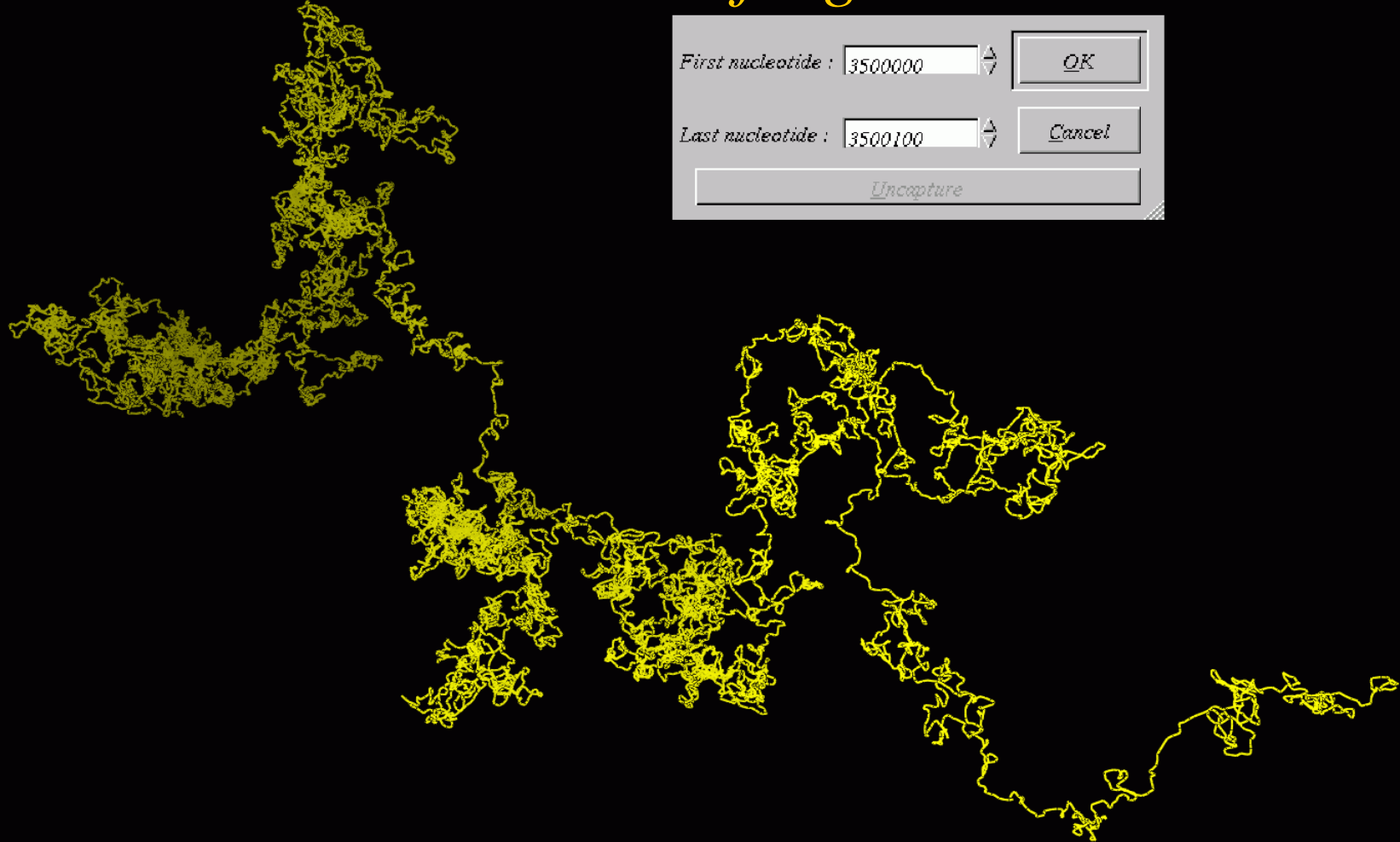
Démonstration *ADN-Viewer*

Architecture des génomes

- **Chromosome**
 - ...intergénique-gène-intergénique...
- **Eucaryotes**
 - Plusieurs chromosomes
 - Noyau contenant le génome
 - Gène : introns-exons
- **Procaryotes**
 - Pas de noyau
 - chromosome souvent « circulaire »
 - Bactéries : pas d'introns

Exploration des génomes

sélection d'un fragment

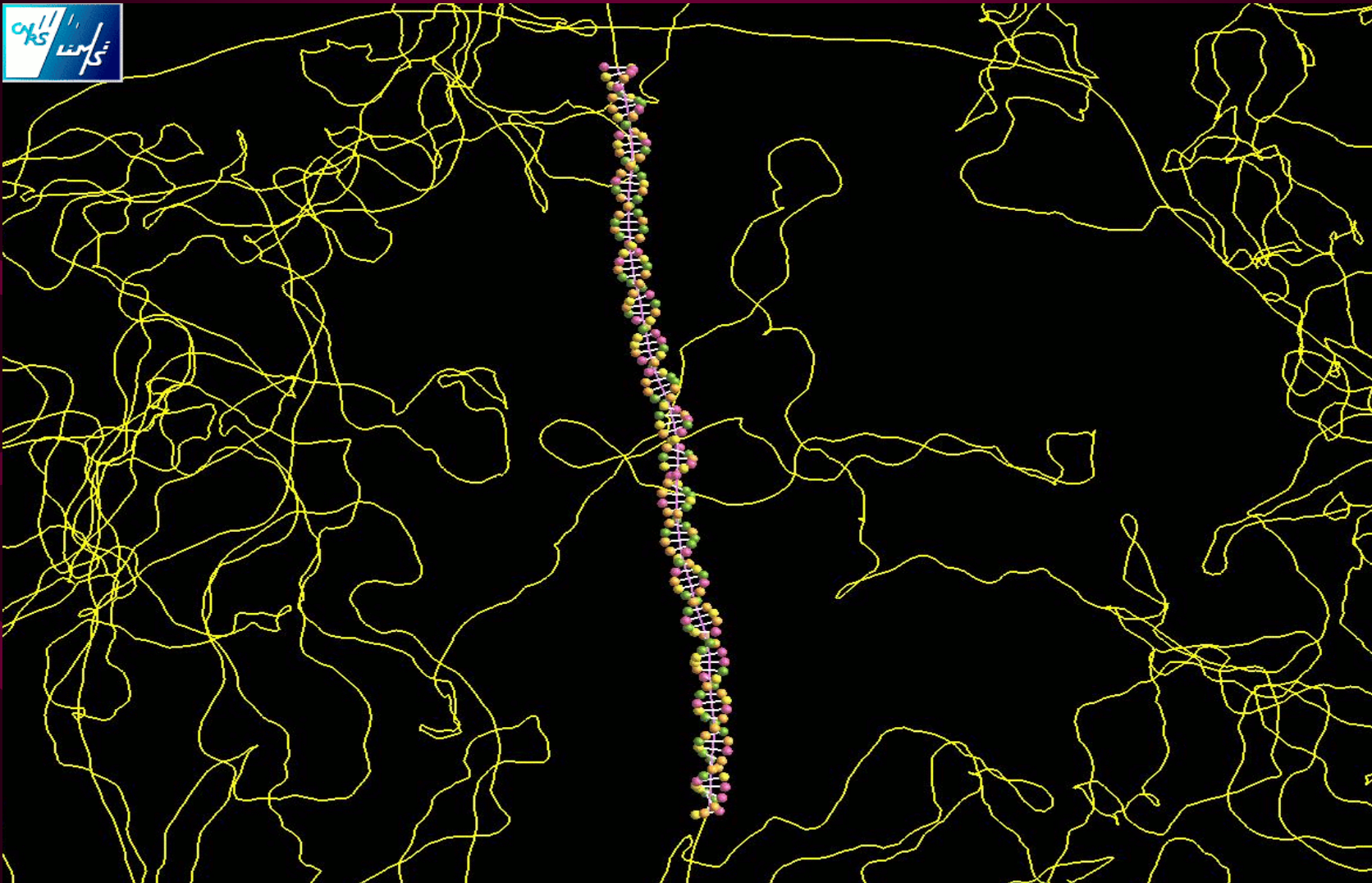


Front Perspective

Exploration des génomes

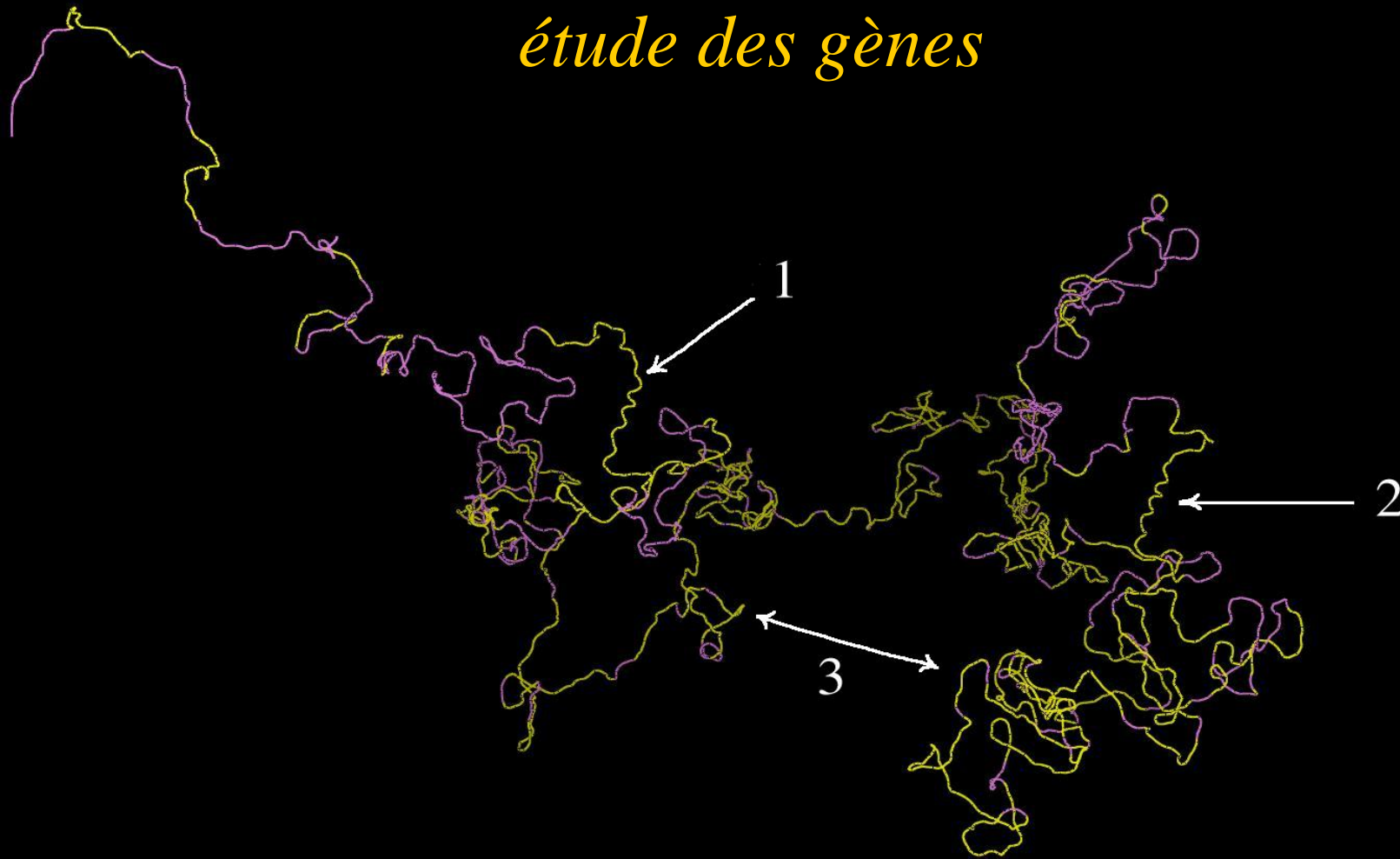
sélection d'un fragment

[Shneiderman, 96]



Exploration des génomes

étude des gènes

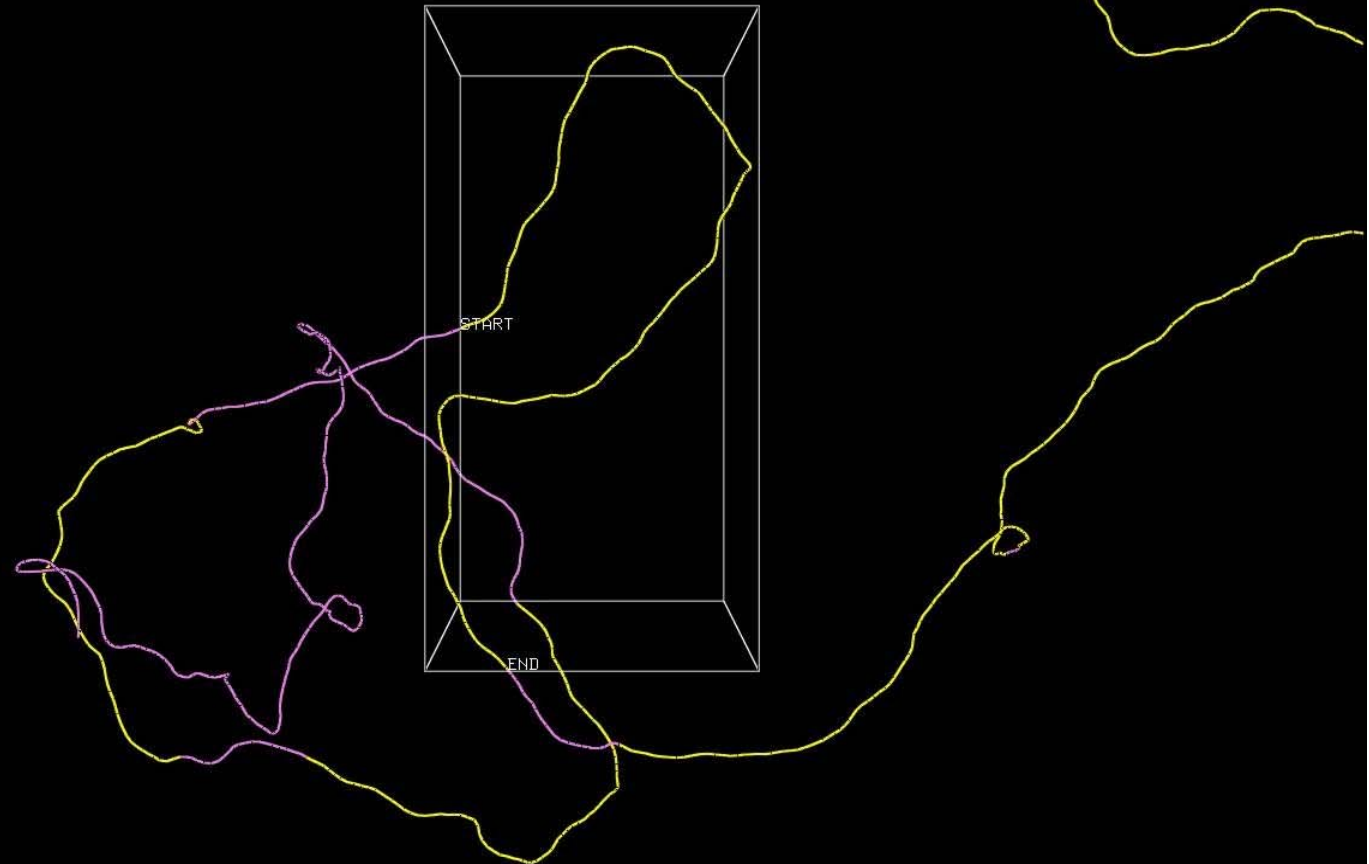


Name	Size
Data Base	Ur
Acanthamoeba castellanii mitochondri...	41
Acipenser transmontanus mitochondria...	16
-ATP 6	68
-ATP 8	16
-COI	15
-COII	69
-COIII	78
-cytb	11
-ND1	97
-ND2	10
-ND3	34
-ND4	13
-ND4L	29
-ND5	18
-ND6	52
-tRNA-Glu	70
-tRNA-Pro	70
Acropora tenuis mitochondrion comple...	18
Adiantum capillus veneris chloroplast...	15
Aeropyrum pernix, Aeropyrum pernix O...	16
Agrobacterium tumefaciens_str_C58_C...	28
Agrobacterium tumefaciens_str_C58_C...	20
Agrobacterium tumefaciens_str_C58_C...	54
Agrobacterium tumefaciens_str_C58_U...	28
Agrobacterium tumefaciens_str_C58_U...	20
Agrobacterium tumefaciens_str_C58_U...	54
Albinaria coerulea complete mithocho...	14
Alepocephalus tenebrosus mitochondri...	16
Alligator mississippiensis mitochondria...	16
Alligator sinensis mitochondrion comp...	16
Alloctytus niger mitochondrial DNA c...	16
Allomyces macrogynus mitochondrial ...	57
Amborella trichopoda chloroplast geno...	16
Ancylostoma duodenale complete mito...	13
Andrias davidianus complete mitochon...	16
Anguilla japonica mitochondrial DNA ...	16
Anomalops katoptron mitochondrial D...	15
Anomalopteryx didiformis mitochondri...	16
Anopheles gambiae complete mitoch...	15



Exploration des génomes

exploration par le contenu



Front Perspective

Acipenser transmontanus mitochondrial DNA complet, Acipenser transmontanus mitochondrial DNA complet

distance = 12210.3

Downloading :

Computing :

Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique

Outils d'analyse

Applications biologiques

Démonstration *ADN-Viewer*

Recherche de motifs 3d

intérêts

- **Intérêts de la recherche de motifs**
 - Caractérisation, découverte de fonction...
- **Intérêts de la recherche de motifs 3d**
 - $F(\text{texte}) \rightarrow 3d$ non bijective
 - But : détecter des motifs 3d proches
- **Principe**
 - Trouver toutes les positions d'un motif M dans un texte T
- **Problèmes**
 - Espace fini \rightarrow espace infini
 - Espace discret \rightarrow espace continu

Recherche de motifs 3d

contraintes

- **Comparaison des points 3d 2 à 2**
 - Dépendante du repère (translations, rotations)
- **Comparaison des angles 2 à 2**
 - Angle trop faible entre 3 nucléotides
 - Trouver une fenêtre de longueur pertinente
 - Information topologique

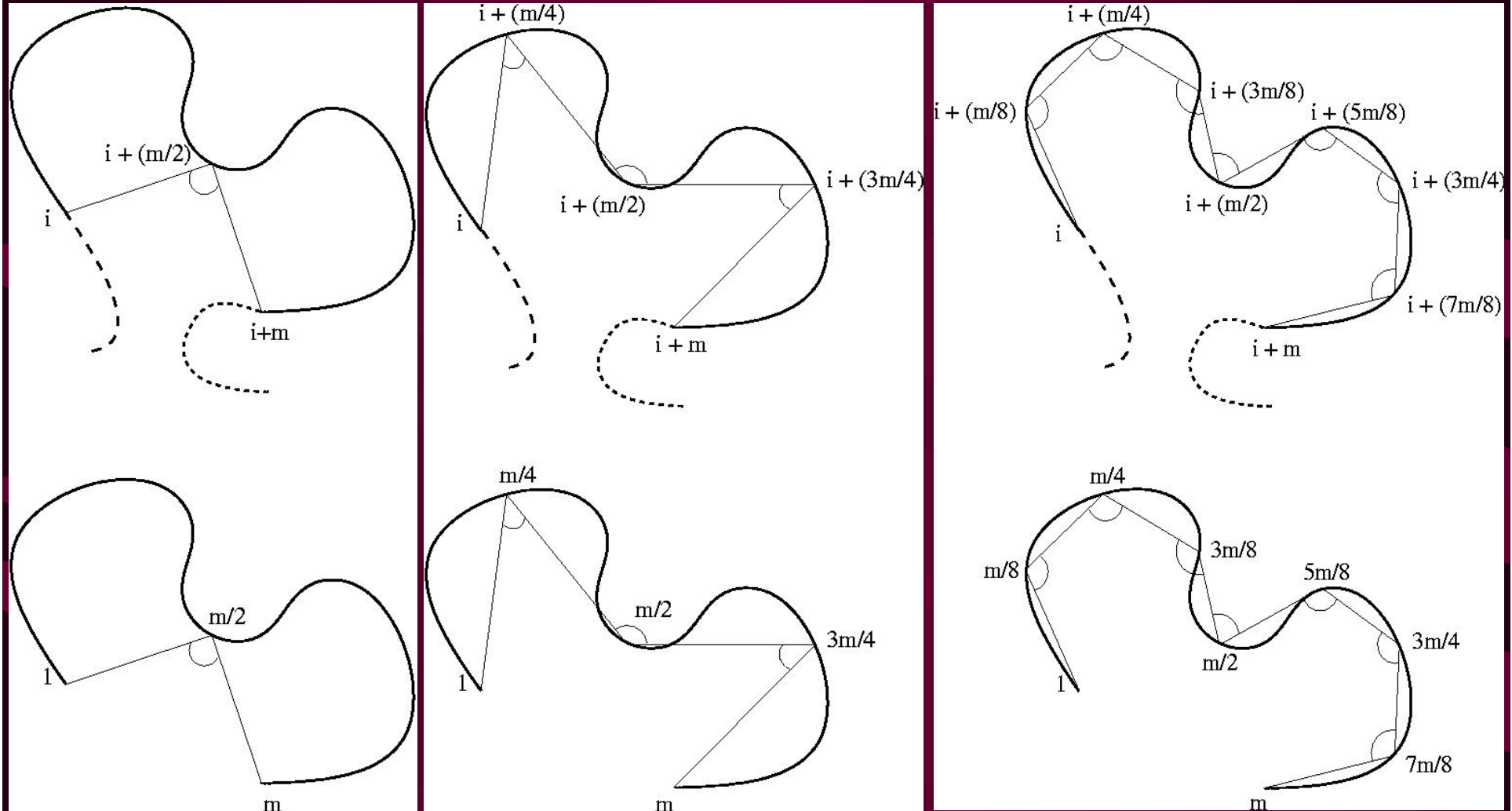
Recherche de motifs 3d

approche globale

- **Découpage vectoriel par dichotomie**
 - Angle formé entre le premier, le dernier et le nucléotide médian
 - Si les angles diffèrent \Rightarrow pas de reconnaissance
 - Si les angles sont proches \Rightarrow réitération au niveau inférieur
- **Qualité de la reconnaissance**
 - Valeur ε , niveau du découpage
- **Filtrage des positions**
- **Détection double sens**

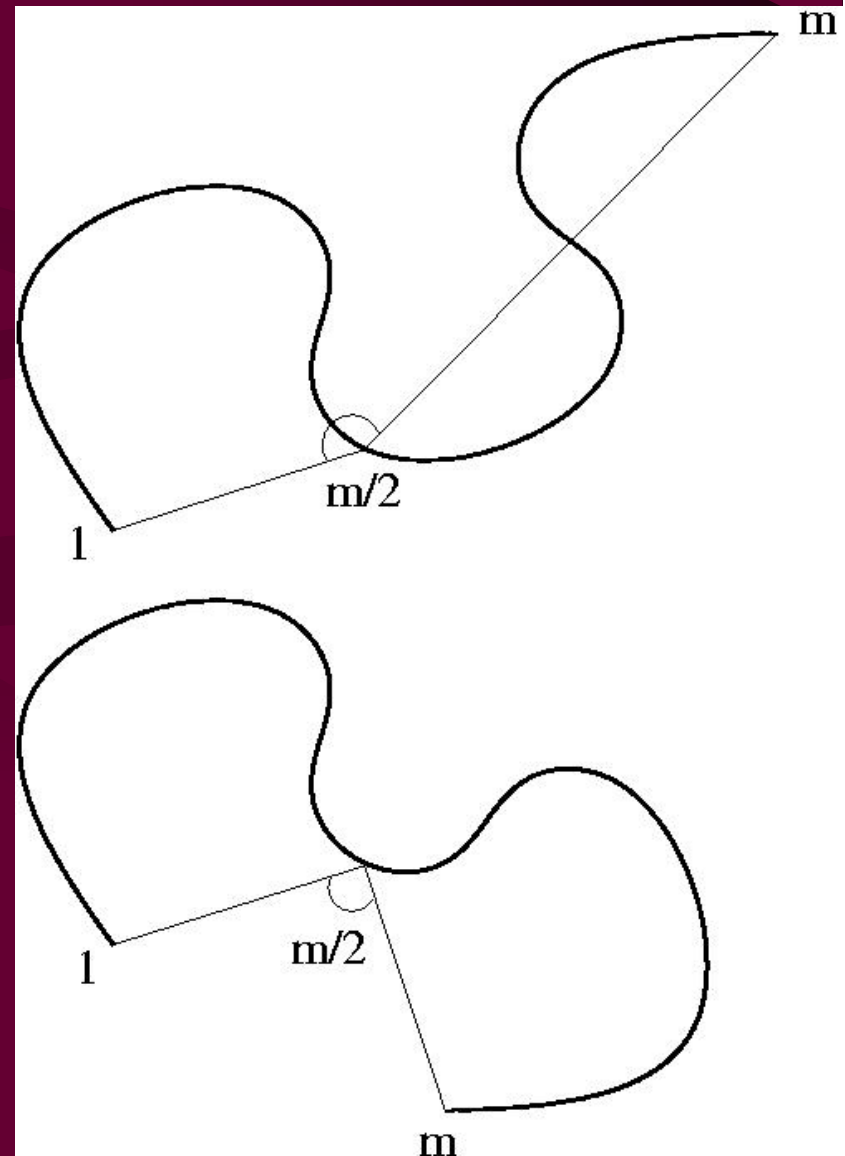
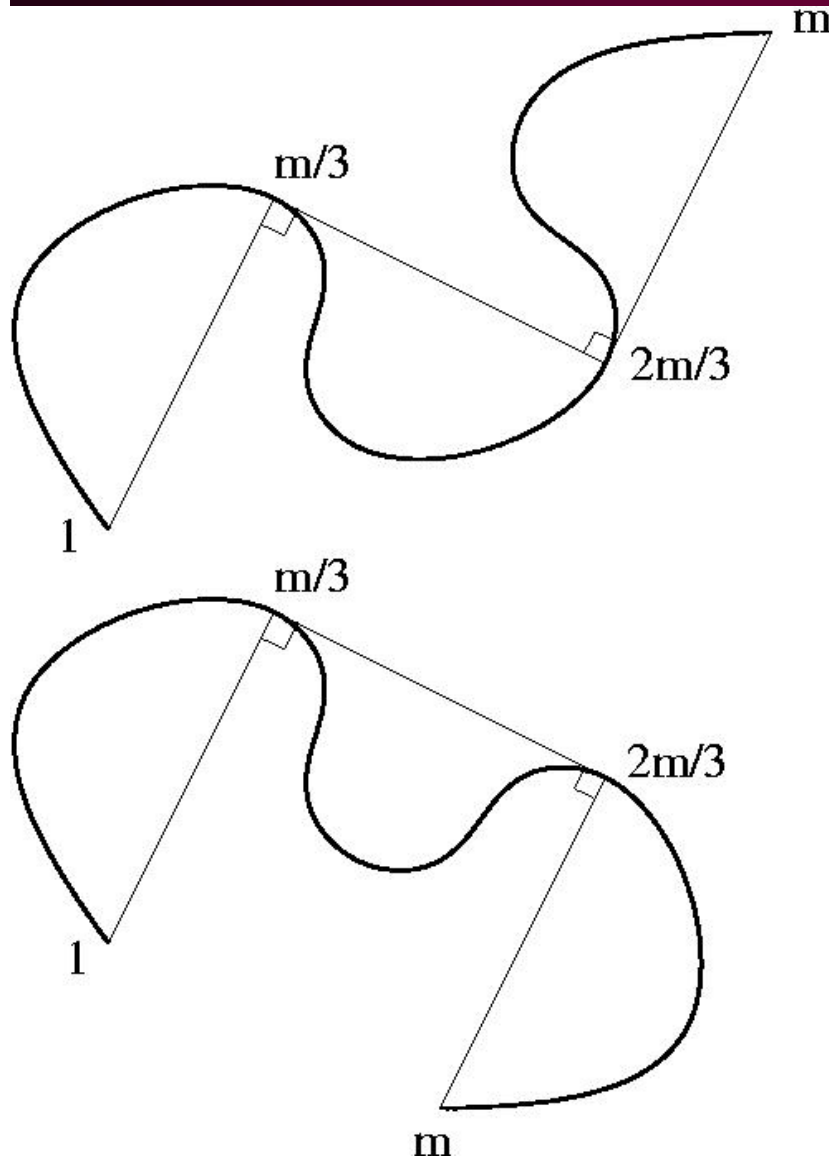
Recherche de motifs 3d

approche globale



Recherche de motifs 3d

naïve vs. globale



Recherche de motifs 3d

approche globale

- **Découpage vectoriel par dichotomie**
 - Angle formé entre le premier, le dernier et le nucléotide médian
 - Si les angles diffèrent \Rightarrow pas de reconnaissance
 - Si les angles sont proches \Rightarrow réitération au niveau inférieur
- **Qualité de la reconnaissance**
 - Valeur ε , niveau du découpage
- **Filtrage des positions**
- **Détection double sens**



Recherche de motifs 3d

résultats

Front Perspective



Front Perspective

Recherche de motifs 3d

résultats



Affinement du modèle de conformation 3d

- **Modèle à faibles contraintes**
 - Local
 - ADN nu
 - Non spécifique
 - Plasmide, chromosome bactérien
- **Fermeture des séquences**
- **Intégration d'un modèle global**
 - Distances spatiales entre paires de fragments d'un chromosome [Dekker et al., 02]

Affinement du modèle de conformation 3d

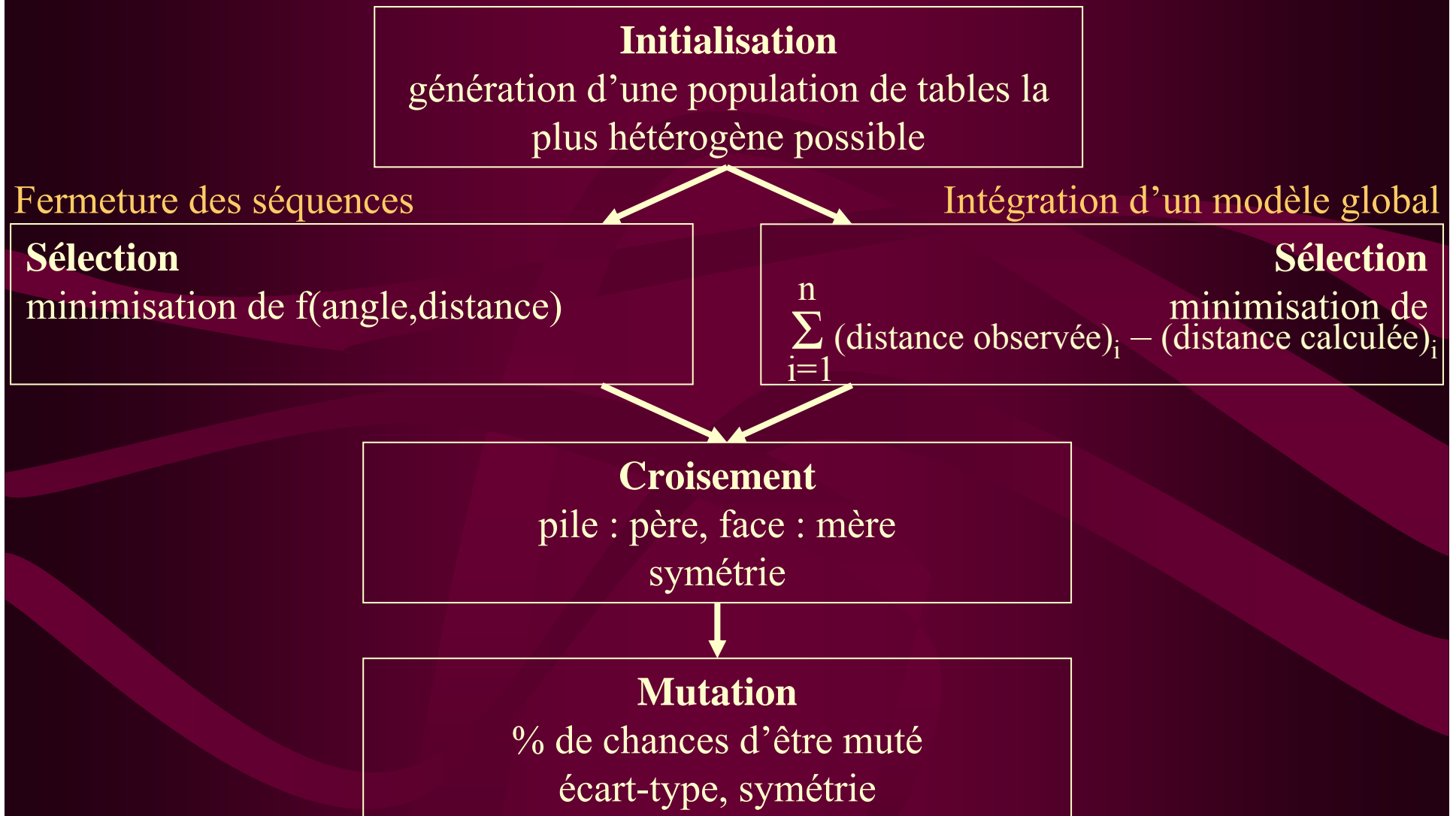
- **Modification du modèle local**
 - Table de rotations
 - Écart-type
 - Combien de valeurs ?
 - De combien on fait varier ?
 - Problème combinatoire
- **Algorithme génétique**
 - Élagage de l'arbre des solutions
 - 1 solution = 1 table = 1 individu
 - Évaluation

Affinement du modèle de conformation 3d

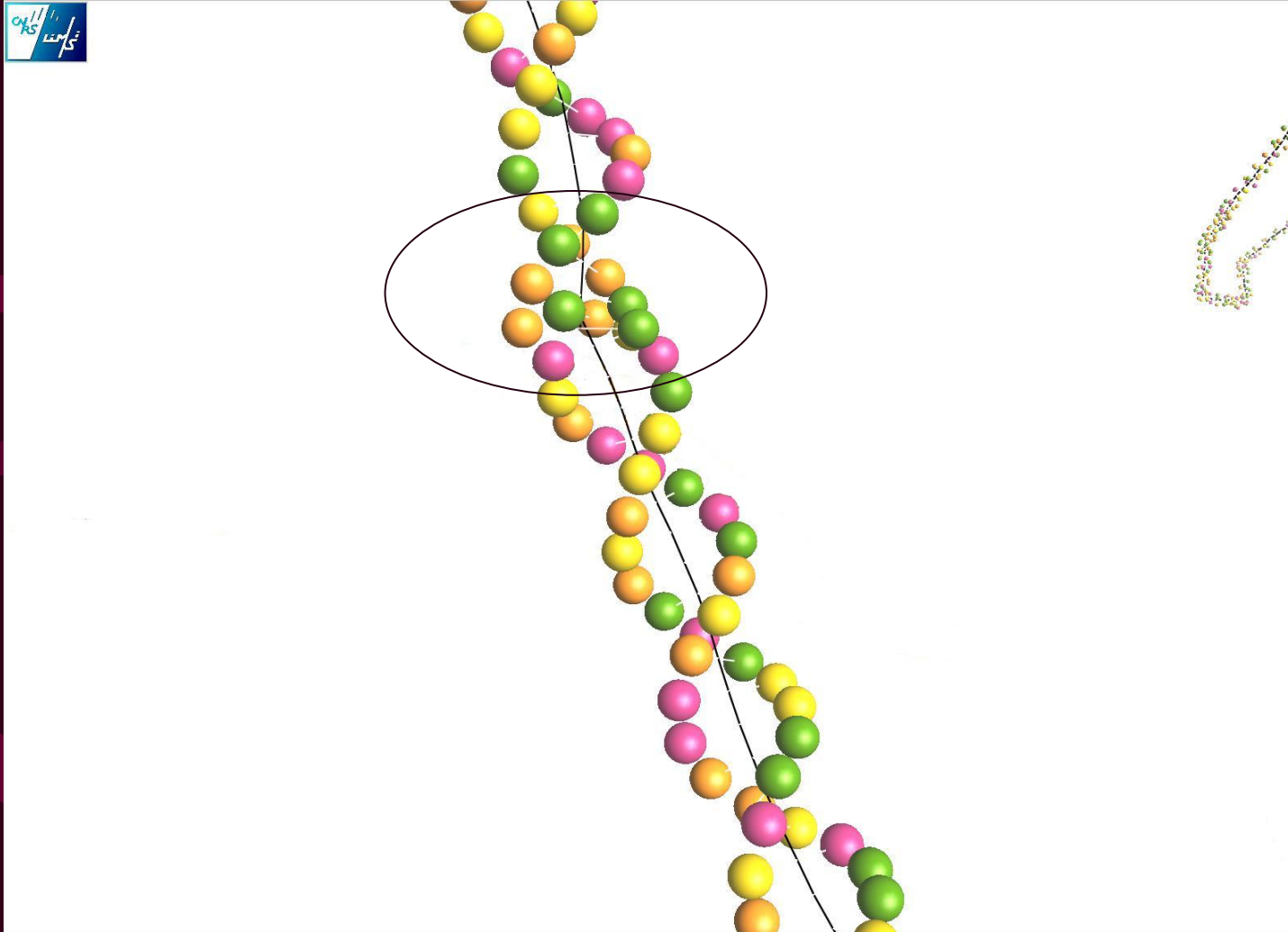
Principe d'un algorithme génétique

- **Population**
 - ensemble d'individus
- **Sélection**
 - Élimination de certains individus
- **Croisement** (explorer)
 - génération d'individu(s) enfant(s) à partir d'individus parents
- **Mutation** (améliorer la solution existante)
 - modification de tout ou partie d'un individu

Affinement du modèle de conformation 3d



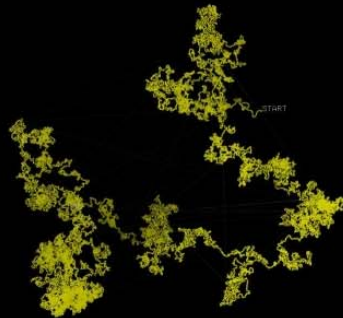
Affinement du modèle de conformation 3d *résultats - fermeture*



Affinement du modèle de

conformation 3d

résultats – modèle global



Front Perspective

Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique

Outils d'analyse

Applications biologiques

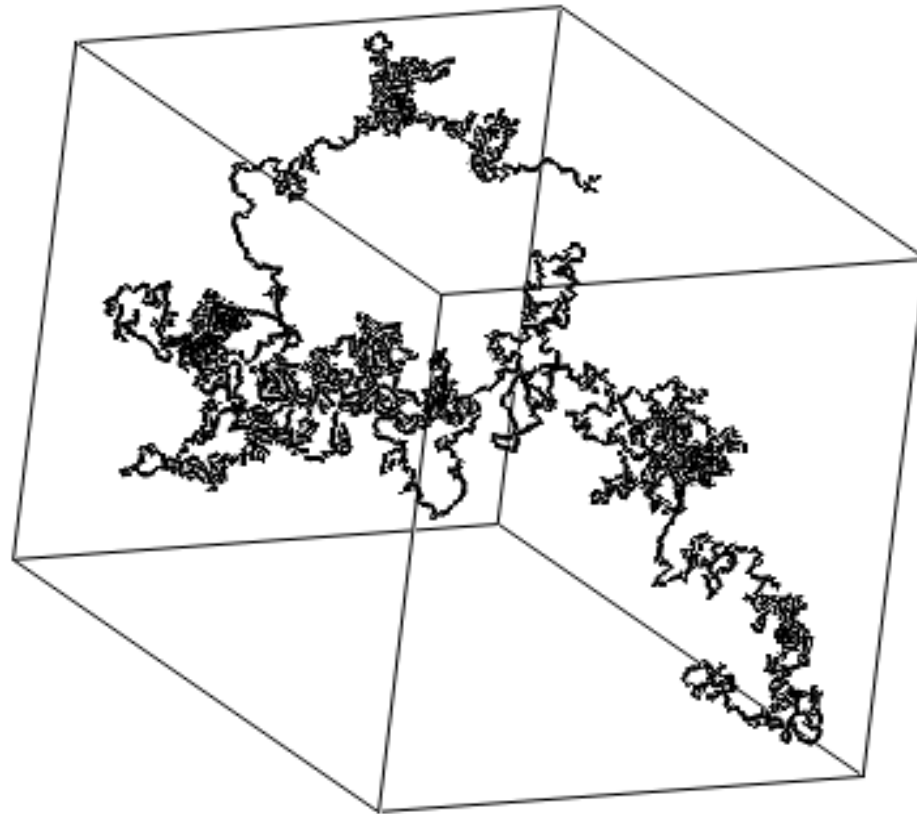
Démonstration *ADN-Viewer*

Étude pluri-génomique de la compaction de l'ADN (IGM)

- **Auto-protection de la molécule d'ADN**
- **Fixation de certaines protéines**
- **Eucaryotes : *histones***
- **Mécanismes de fixation liés à des zones de courbure intrinsèque de l'ADN ?**
- **Données : 35 génomes**
 - 5 eucaryotes, 8 archées, 22 bactéries
 - MIPS, EMBL

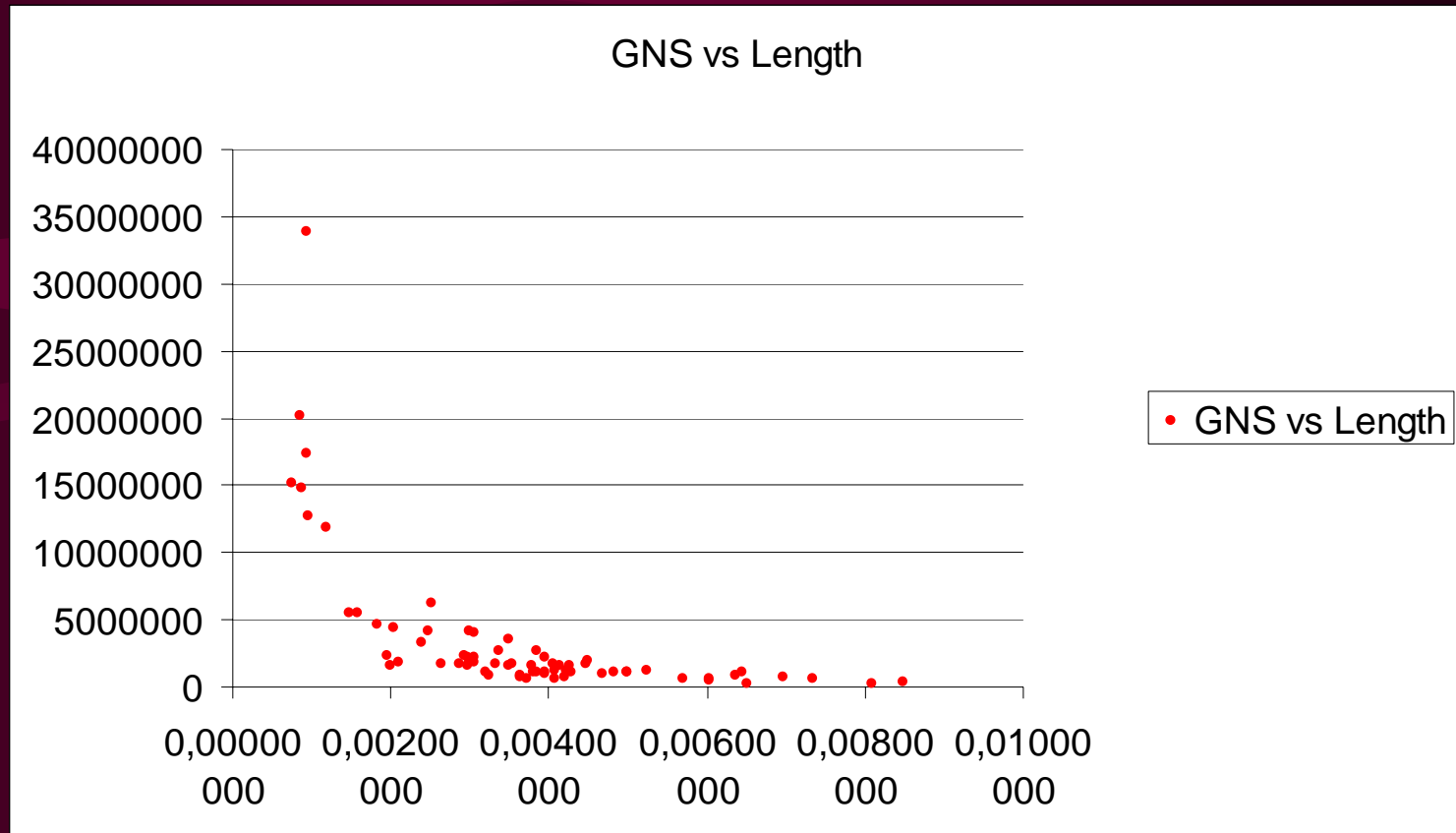
Étude pluri-génomique de la compaction de l'ADN

Étendue
globale



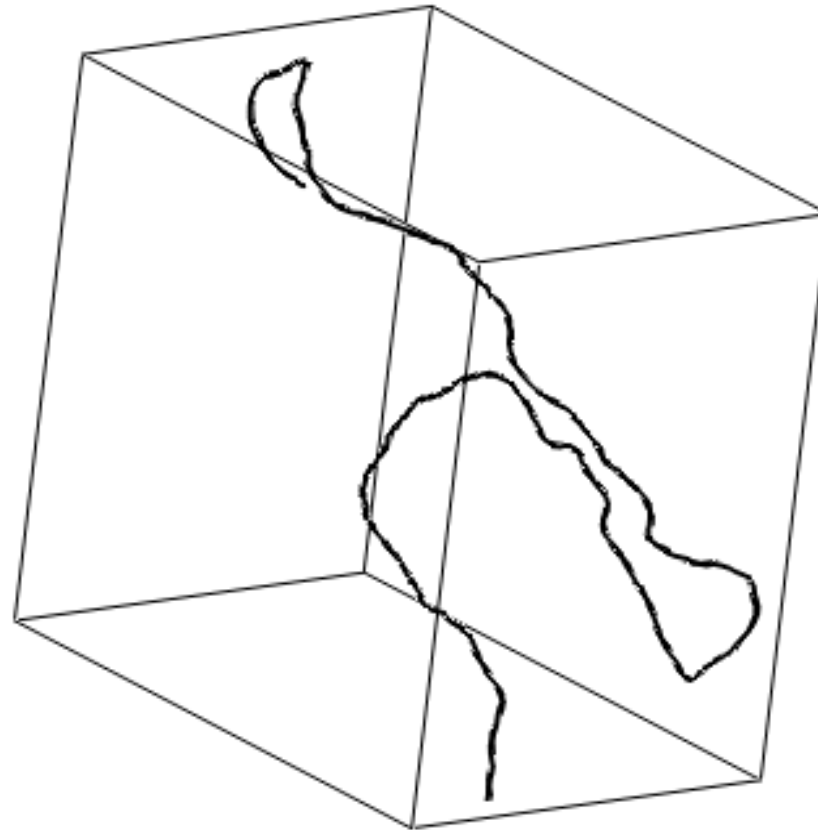
Étude pluri-génomique de la compaction de l'ADN

Étendue globale



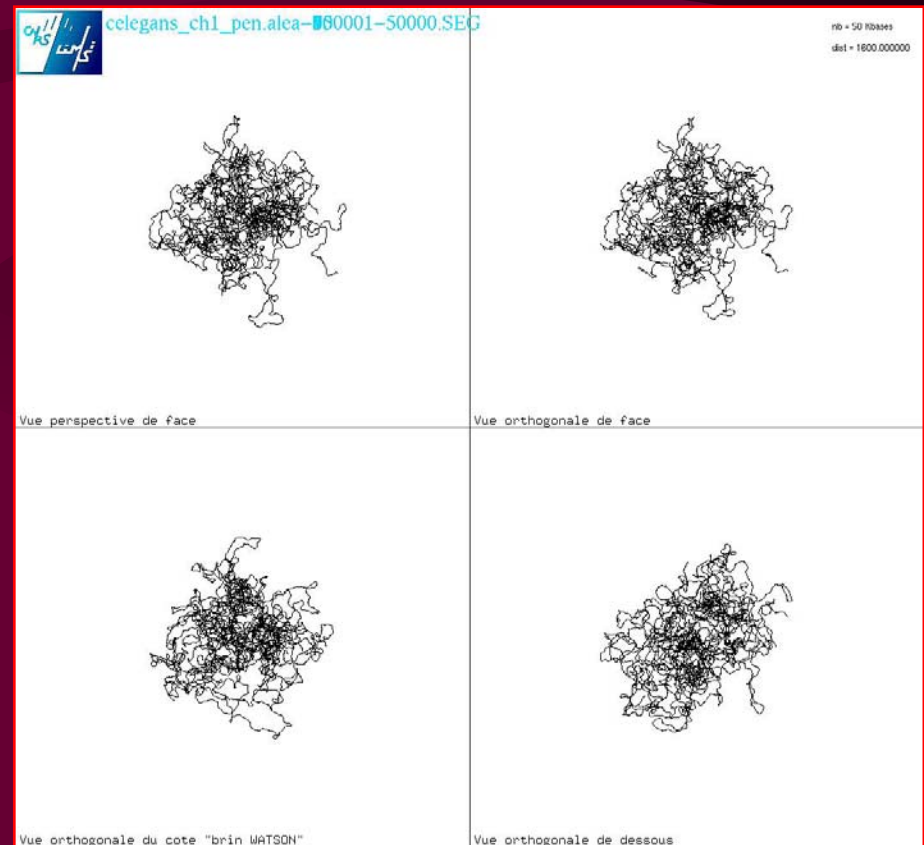
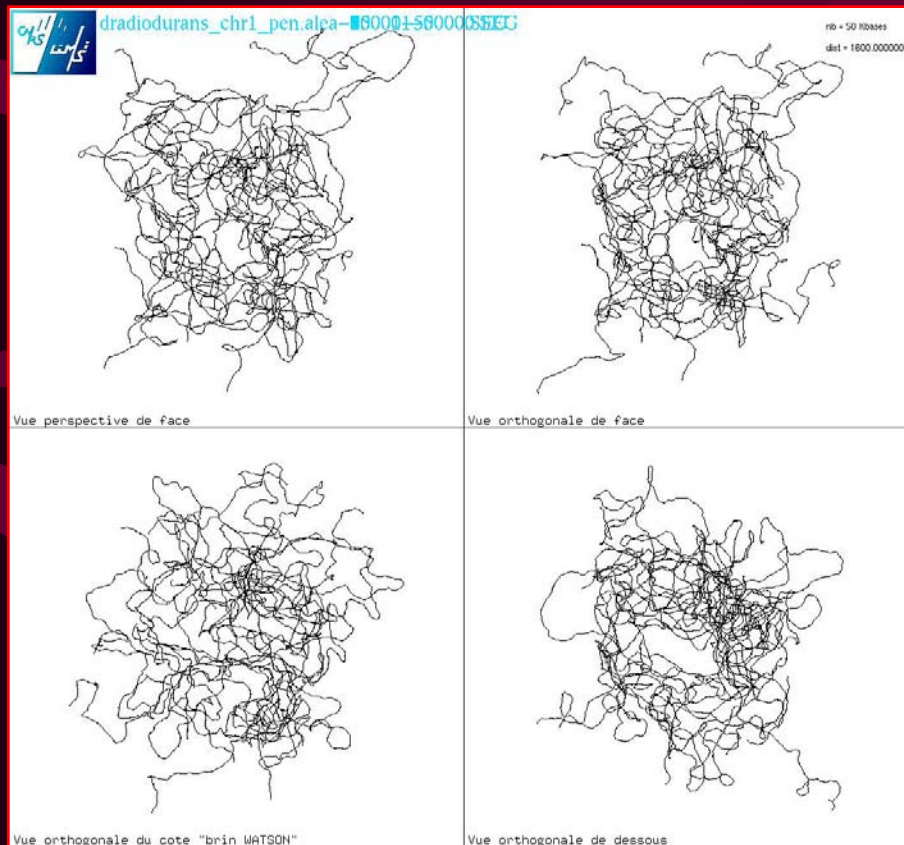
Étude pluri-génomique de la compaction de l'ADN

Étendue locale



Étude pluri-génomique de la compaction de l'ADN

Étendue locale



Étude des sites d'épissage chez *A. thaliana* (IBP)

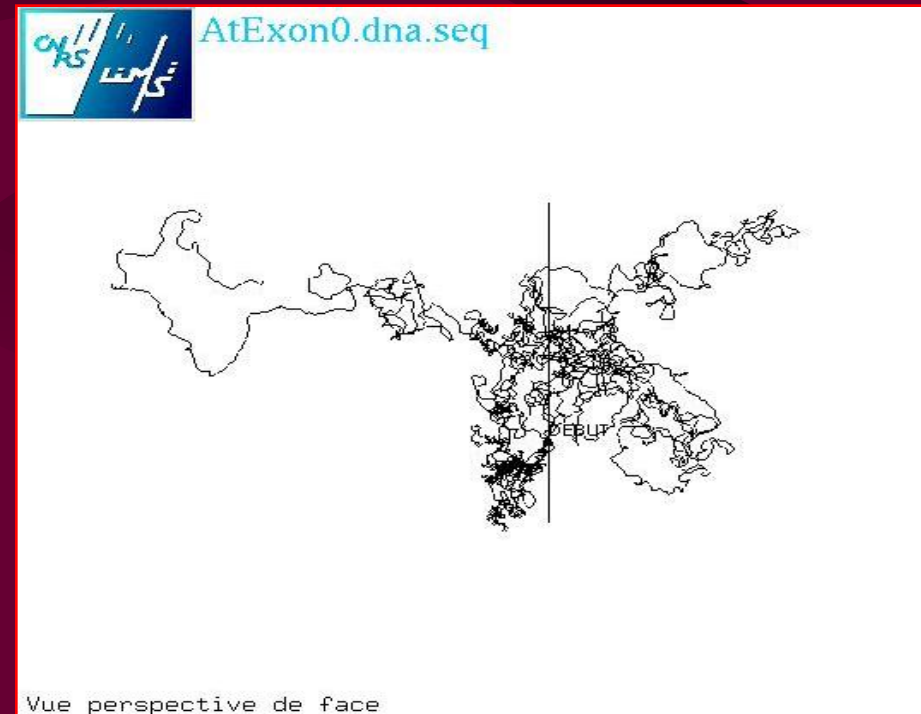
- **Prédiction textuelle**
 - 60 %
- **Améliorer par critères géométriques**
 - Introns vs. Exons
 - Sites d'épissage
- **Données**
 - EID, PlantGene

Étude des sites d'épissage chez *A. thaliana* *introns vs. exons*

Introns



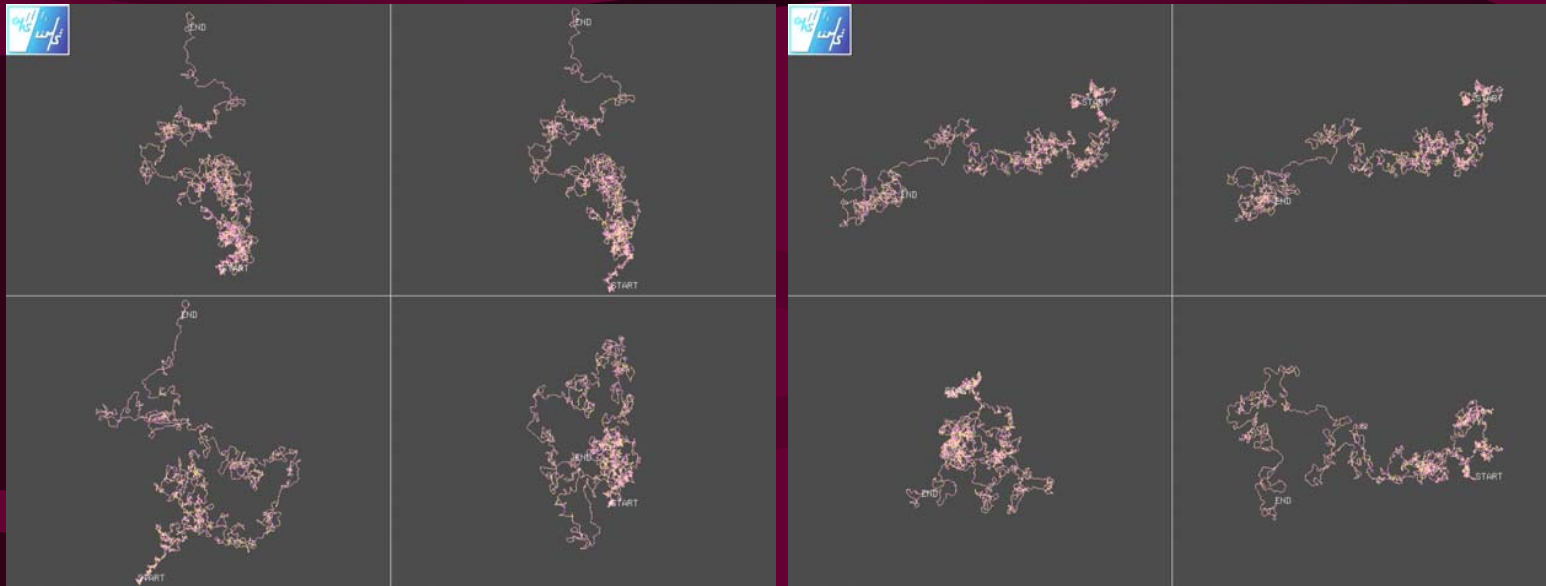
Exons



Introns – Exons : compaction différente

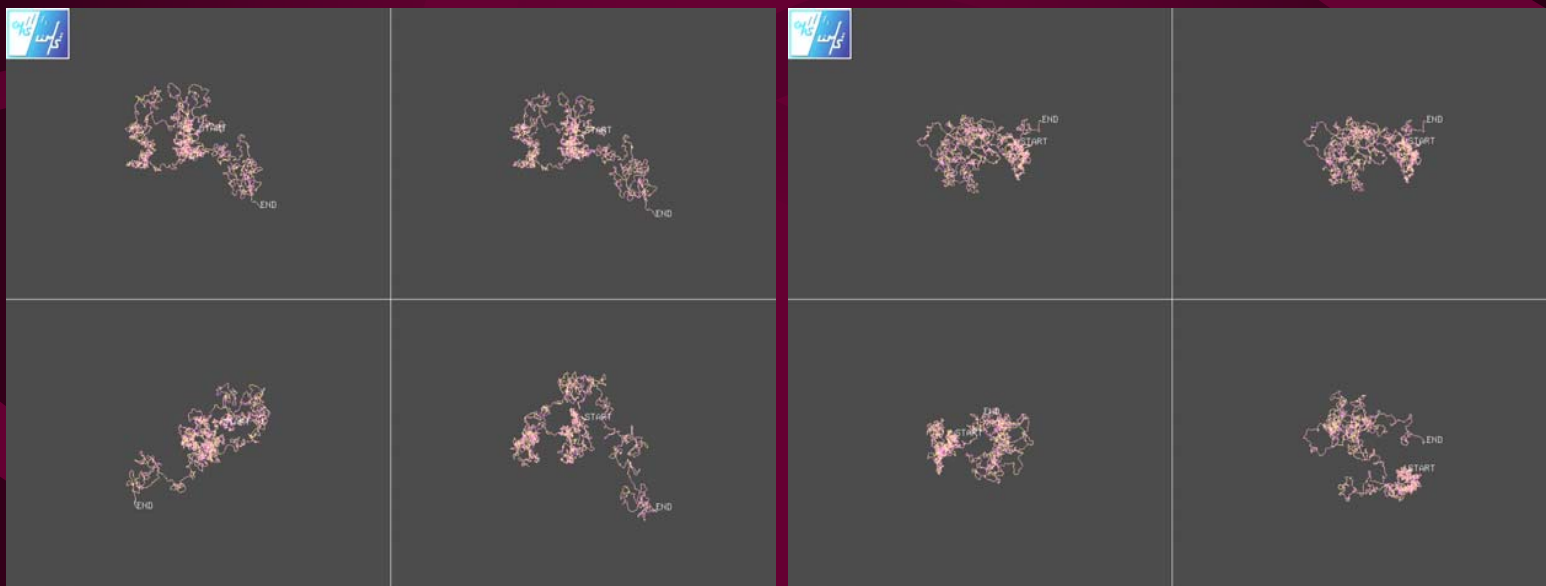
Exons

55% AT
1268510 bp



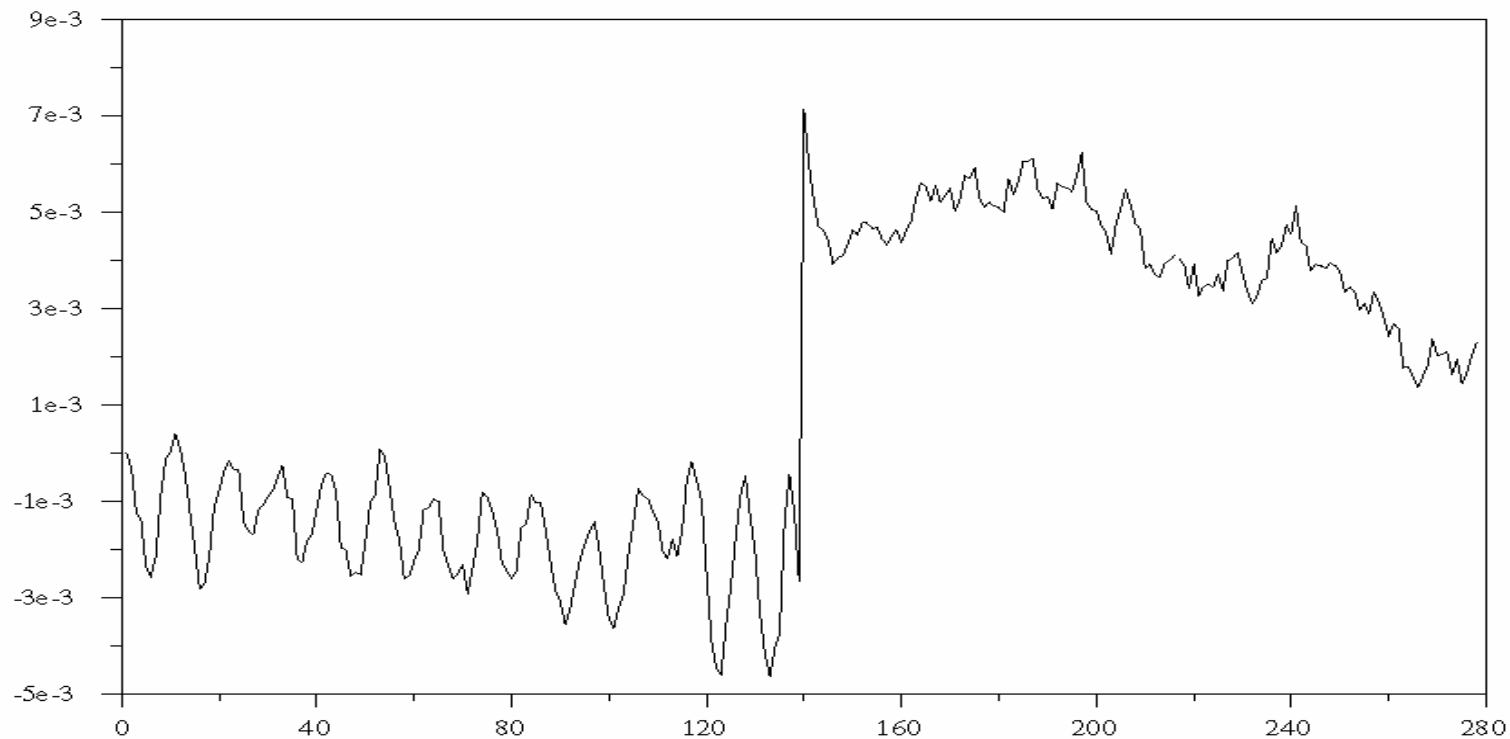
Introns

67% AT
127293 bp



Étude des sites d'épissage chez *A. thaliana* *sites d'épissage*

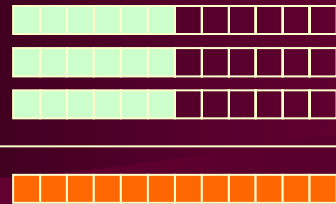
At accepteurs, delta-mean X coordinate, 280 bp



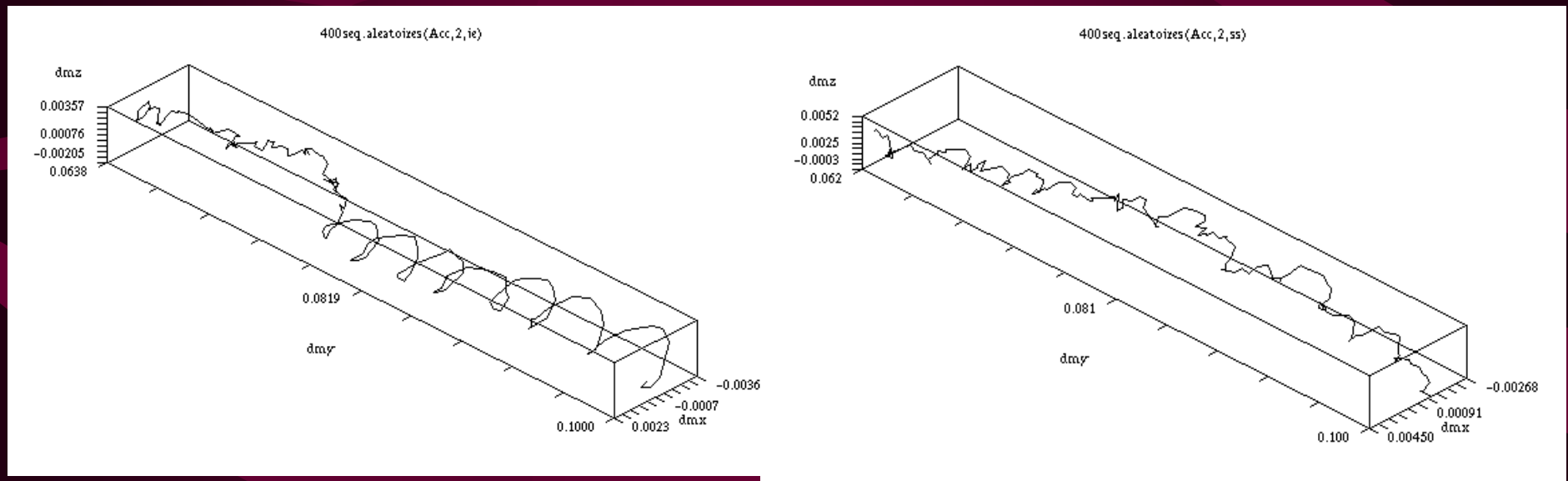
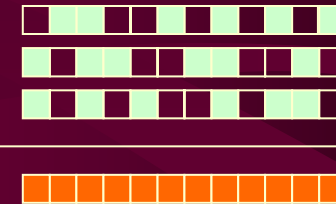
Séquences aléatoires (1)

Analyse linguistique des séquences utilisées, puis GenRGenS

avec distinction intron-exon



sans distinction intron-exon



Conclusion : effet de la composition en nucléotides de la séquence

Conclusions

Compaction des introns par rapport aux exons

Hélice très visible pour les introns (périodicité de $\sim 10,5$ bp)

Suite de l'étude

Effet de moyenne \Rightarrow signaux individuels

Analyse en contexte de gène et de séquences génomiques (intergénique ?)

\Rightarrow reconnaissance des sites d'épissage (système à apprentissage)

Questions ouvertes

Périodicité de 10,2 bp (Fourier, ACF) : associée aux nucléosomes ?

Intron avec faible %AT : signaux d'épissage fort ?

Rôle dans l'expression : introns courts \Leftrightarrow forte expression du gène ?

Applications

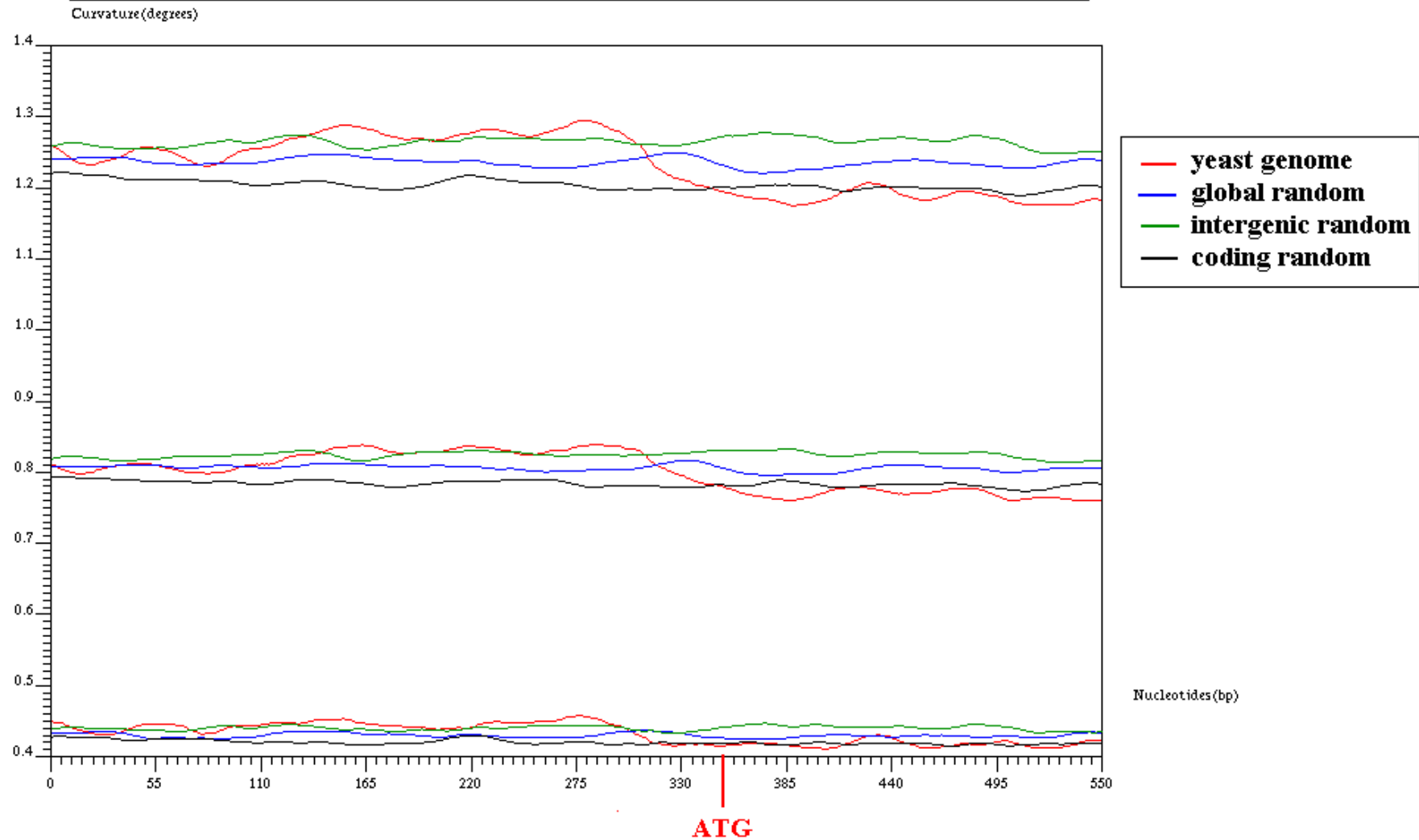
Courbure de zones régulatrices

chez *S. cerevesiae*

IGM Orsay

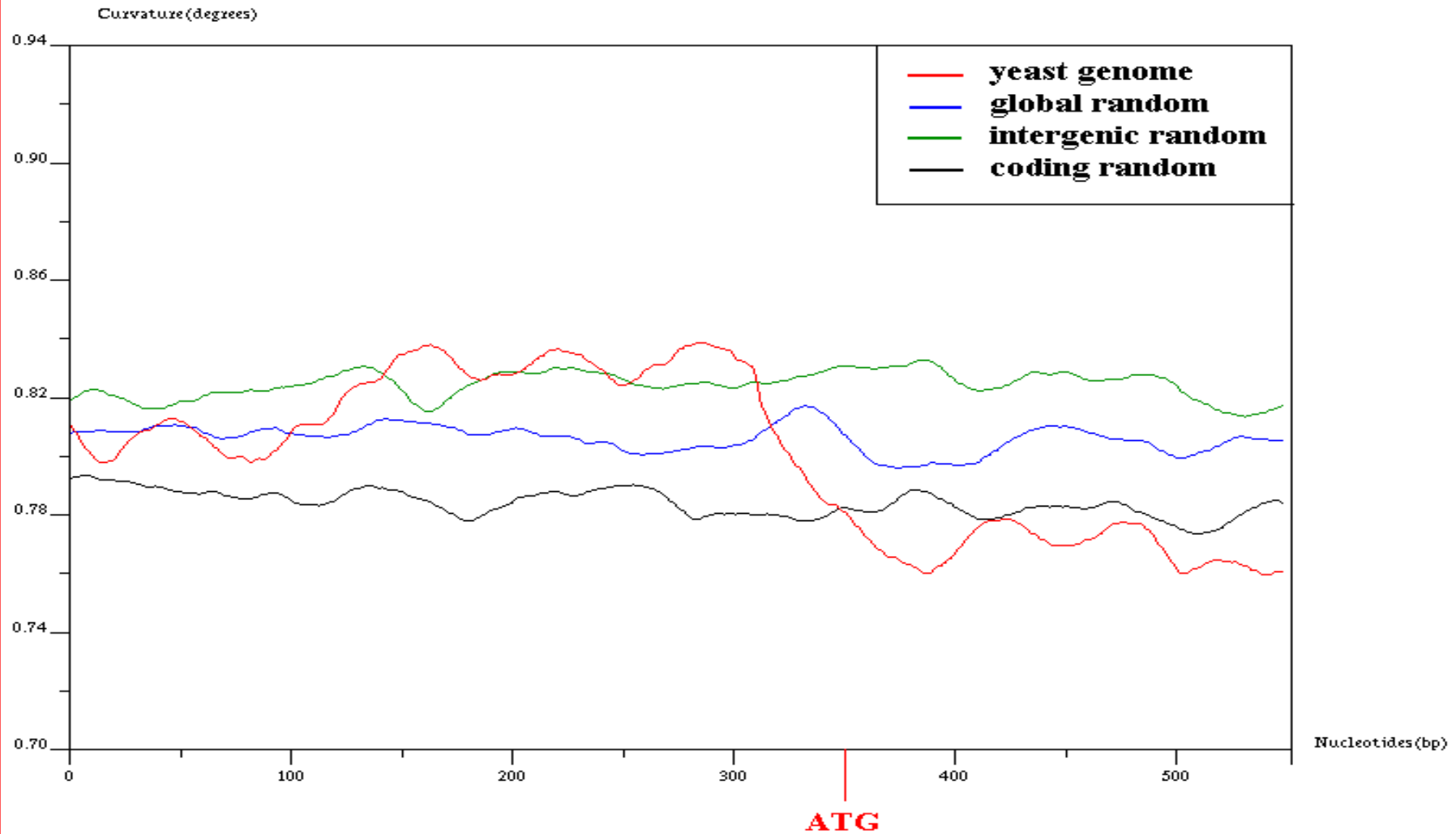
Etude génomique de zones intergéniques

Comparison between mean curvature of the yeast genome and random sequences with standard deviations



Etude génomique de zones intergéniques

Comparaison of mean curvatures for yeast genome and random sequences



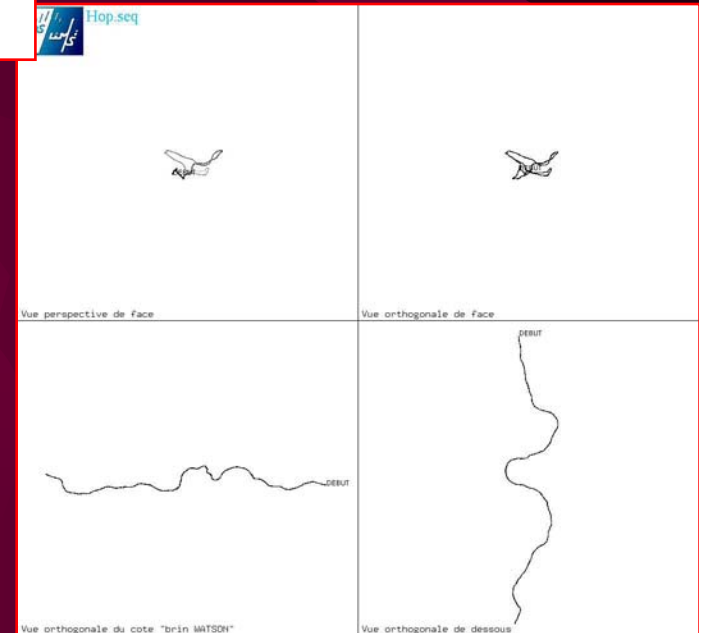
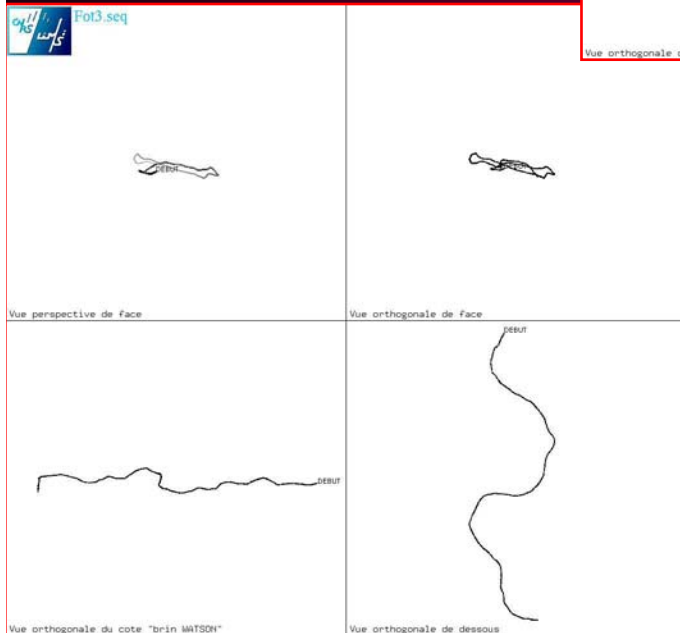
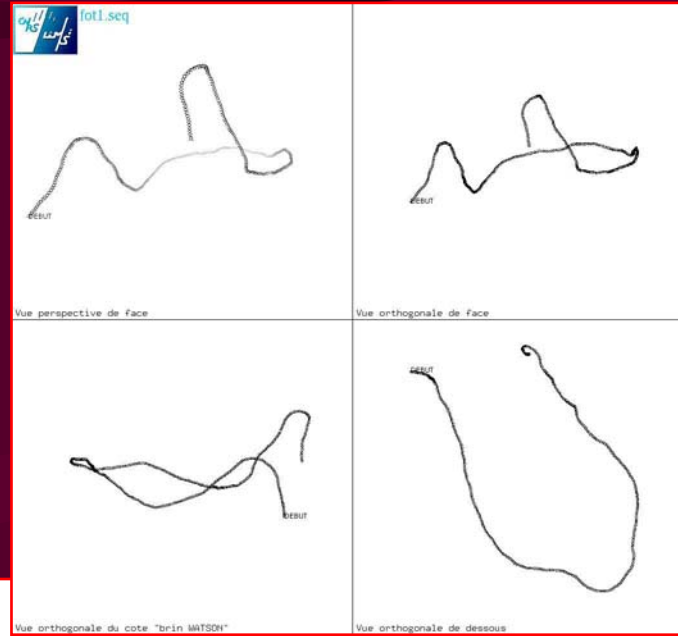
Applications

Structure de sites d'insertion et transposons

chez *F. oxysporum* et *M. grisea*

IGM Orsay

Exemples de transposons



Organisation globale de la transcription chez *S. cerevisiae*

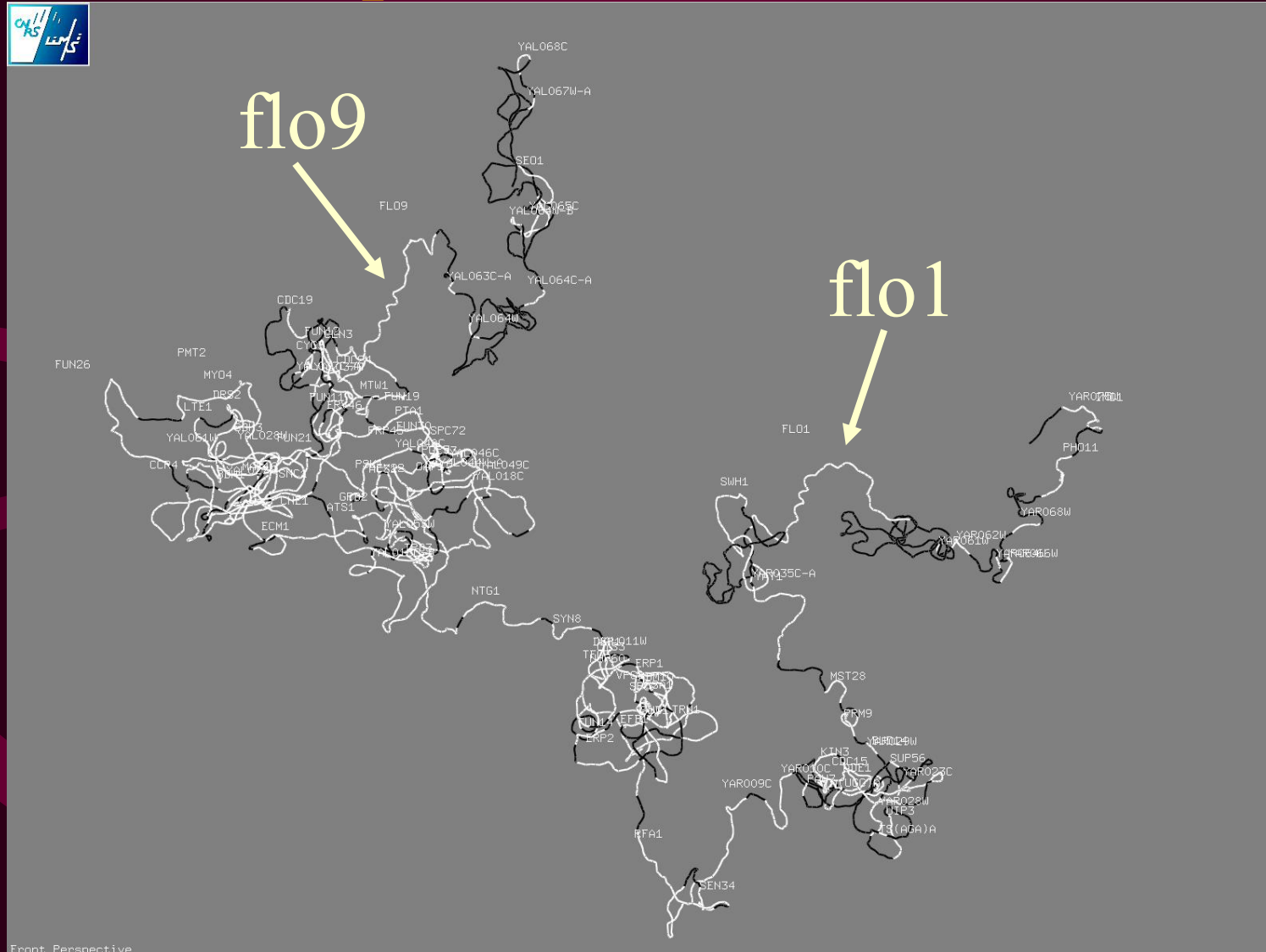
(LRI, Prog. Épigenomique)

- **2 gènes éloignés dans la séquence textuelle peuvent être spatialement proches**
- **Co-régulation**
- **Structure 3d de l'ADN nu joue-t-elle un rôle dans le rapprochement spatial des zones de co-régulation ?**
- **Données**
 - 106 régulateurs
 - ChIP

Organisation globale de la transcription chez *S. cerevisiae*

- **Distance linéaire vs. distance spatiale**
- **Degré d'accessibilité des gènes**
 - Modèle d'ADN nu statique
 - 2 gènes périphériques vs. 2 gènes proches dans une zone ultra-compacte

Organisation globale de la transcription chez *S. cerevisiae*



Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique

Outils d'analyse

Applications biologiques

Démonstration *ADN-Viewer*

Plan

Problématiques

Etat de l'art et concepts de base

Modélisation et visualisation 3d du génome

Exploration du contenu génomique


Outils d'analyse

Applications biologiques

Démonstration *ADN-Viewer*

Contributions croisées

- **Informatique**
 - Gestion de grosses masses de données
 - Calcul, stockage, rendu graphique
 - Environnement bureau/immersif
 - Algorithmes 3d
- **Génomique**
 - Représentation pertinente des données
 - Traitement et analyse des structures 3d
 - Validation et intégration des modèles



Bioinformatique structurale des protéines

Protéomique : importance du repliement

The 3D structure significantly determines the function.



Deux Problématiques complémentaires pour le repliement

1. Protein Folding Prediction ---

Given a **protein sequence**, determine the **3D folding** of the sequence.

2. Protein Sequence Design ---

Given a **3D structure**, determine the **fittest protein sequence** for the structure, i.e., one that has the smallest energy among all possible sequences when folded into the structure.

Complexité des problèmes en repliement

- **Protein Folding Prediction ---**
Given a protein sequence, determine the 3D folding of the sequence.

NP-hard under various models.

- **Protein Sequence Design ---**
Given a 3D structure, determine the fittest protein sequence for the structure, i.e., one that has the smallest energy among all possible sequences when folded into the structure.

Solvable in polynomial time (constraints).

Exemple de protéine

Proteome: all proteins encoded within a genome

half millions distinct proteins (temporal, spatial, modifications)

~30,000 human genes

mRNA and protein expressions may not correlate

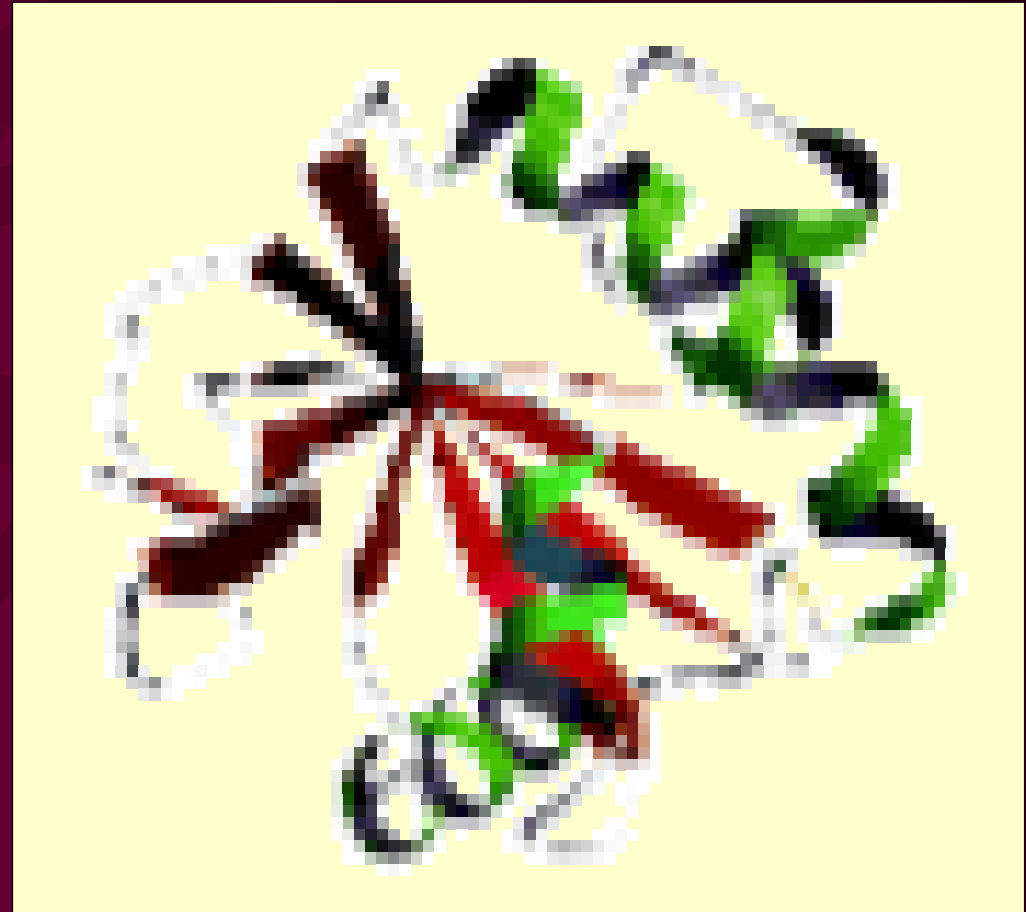
Technologies:

2D gel electrophoresis (images)

mass spectrometry (3d)

yeast two-hybrid system

protein chips (images)



Protein-protein association kinetics and protein docking

Carlos J. Camacho et Sandor Vajda

Plan

- Introduction
- Les méthodes de docking
- La cinétique de liaison
- Une nouvelle approche
- Conclusion

Introduction

Plan

Introduction

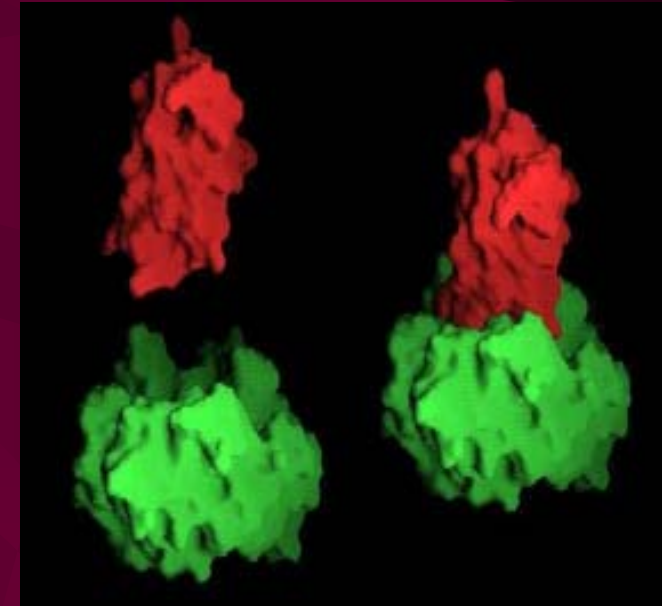
Les méthodes de docking

La cinétique de liaison

Une nouvelle approche

Conclusion

- Interaction protéinique & Biologie
- Interaction protéinique & Informatique
- Un peu de vocabulaire :
 - Le docking
 - La cinétique de liaison
 - Les méthodes de docking



Les méthodes rigides

Plan

Introduction

Les méthodes de docking

La cinétique de liaison

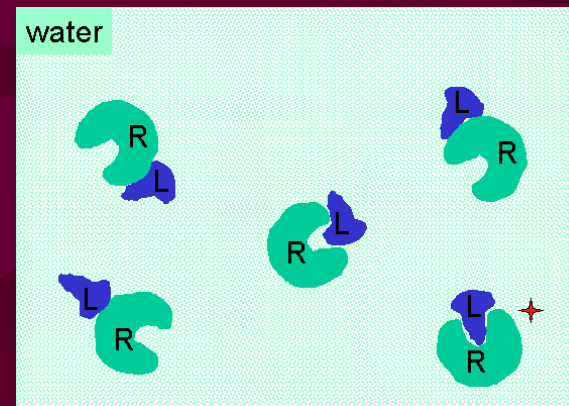
Une nouvelle approche

Conclusion

- L'amarrage se déroule

en deux étapes :

- Searching
- Scoring



- Hypothèse : pas de changements de conformation
- Le problème des protéines cristallisées séparément

Les filtres

Plan

Introduction

Les méthodes de docking

La cinétique de liaison

Une nouvelle approche

Conclusion

- Des méthodes de discrimination :
 - Statistiques
 - Énergie Électrostatique
 - Énergie de Van der Waals
- Un filtrage total impossible

La cinétique de liaison

Plan

Introduction
Les méthodes de docking
La cinétique de liaison
Une nouvelle approche
Conclusion

- Mise en valeur de certains paramètres :
 - L'interaction ligand protéine
 - Les forces électrostatiques
 - Les forces de "desolvation"
- Lien entre la cinétique de liaison et le docking

Une nouvelle approche

Plan

Introduction
Les méthodes de docking
La cinétique de liaison
Une nouvelle approche
Conclusion

- Utilisation des connaissances sur la cinétique de liaison
- Une méthode de docking flexible en deux parties :
 - Recognition
 - Refinement
- Méthodes rigides VS Méthodes flexibles

Un exemple

Plan

Introduction
Les méthodes de docking
La cinétique de liaison
Une nouvelle approche
Conclusion

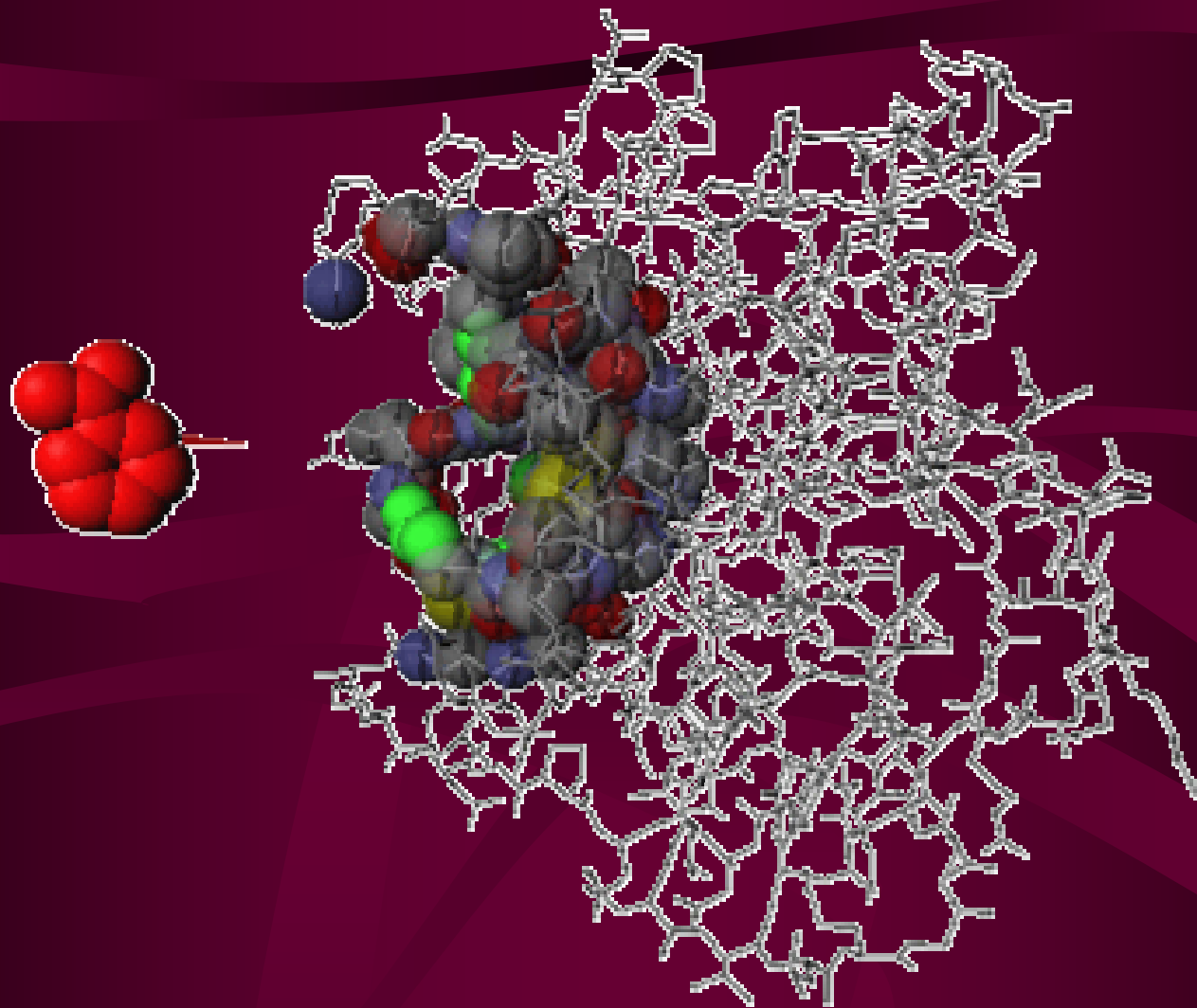
- Utilisation de l'interaction ligand protéine
- Le Ligand caractérise la zone d'interaction
- Recherche du ligand par :
 - Analyse de la structure moléculaire
 - Étude des forces électrostatiques

Conclusion

Plan

Introduction
Les méthodes de docking
La cinétique de liaison
Une nouvelle approche
Conclusion

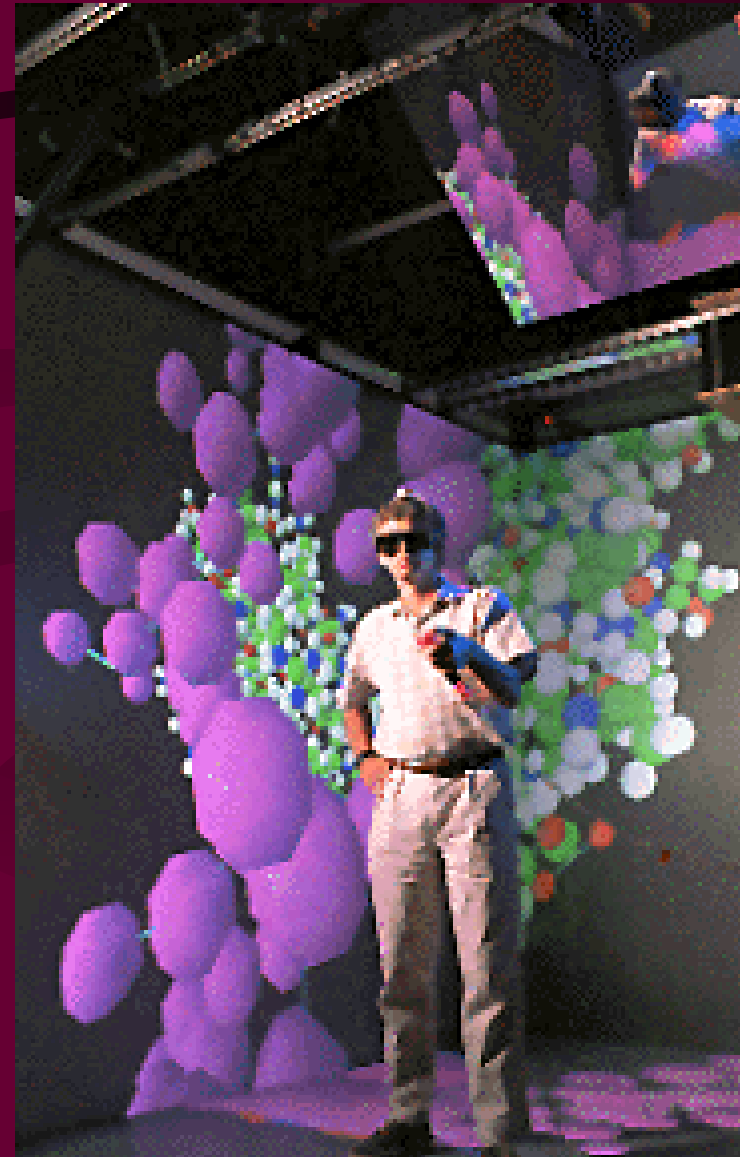
- Des méthodes de docking prometteuses et en cours de validation
- Un secteur de recherche nouveau (juin 2001 conférence *Modeling of Protein Interactions in Genomes*) et en plein expansion



HotDock, visualisation, teaching and communication of Molecular Docking.
group "Visualisation in Chemistry"

AG Domik (Visualisation, Computer Science)

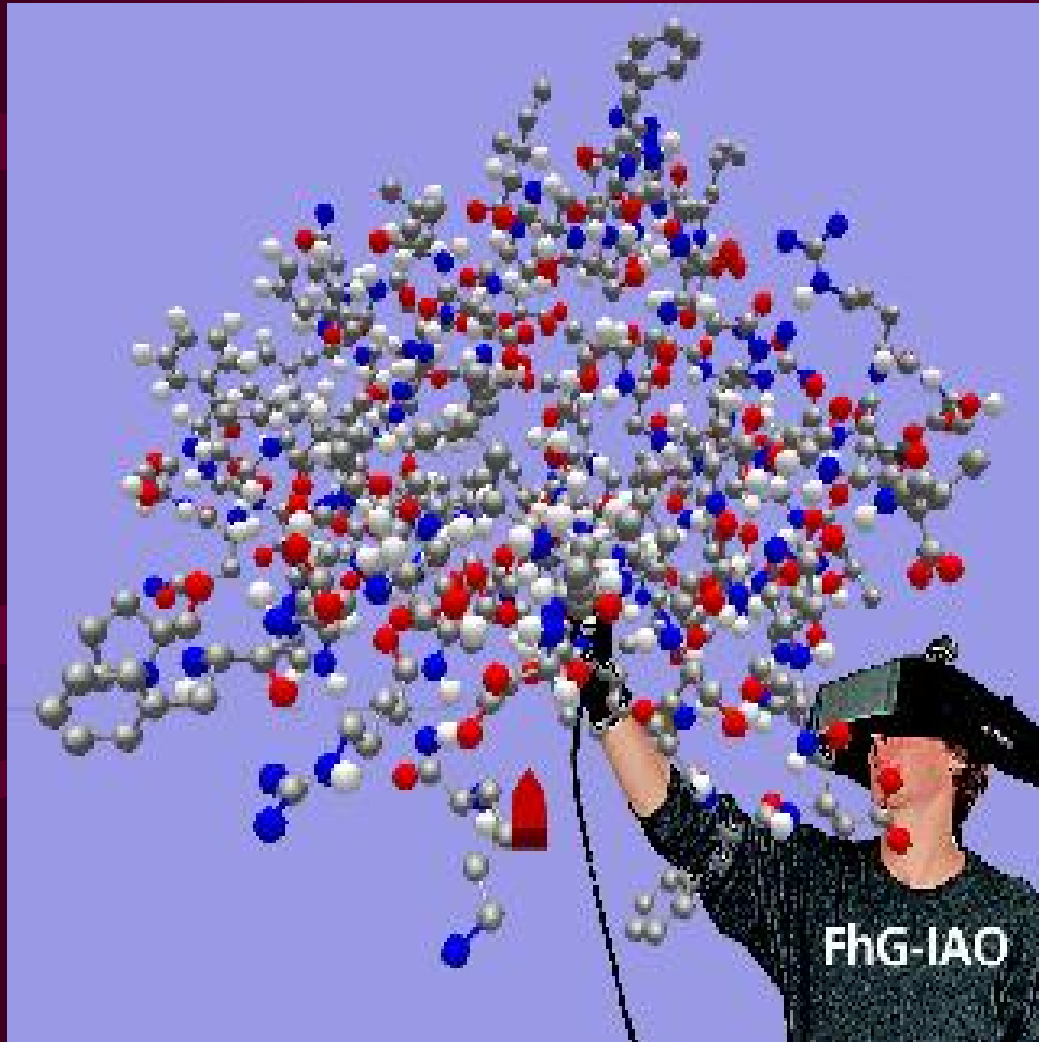
Dr. Fels (organic chemistry), University of Paderborn




STALK: A Mixed-Reality System
Center for Mechanistic Biology and Biotechnology
Mathematics and Computer Science Division,

UP-XI Master Rech - TIG, 2004-2005

LIMSI-CNRS, Université Paris-Sud XI



HIMM: High Immersive Molecular Modelling
VIS-Lab, Stuttgart , Fraunhofer-Gesellschaft IAO , 1996



Fouille visuelle de données
factuelles et textuelles
en biologie moléculaire

Données factuelles – Données textuelles

- **Relier les séquences d'ADN aux publications pertinentes**
- **Relier un article aux éléments biologiques concernés**
- **Unification des données et de leur représentations**
- **Problématiques dues à l'exploration des données**



Immersive Graph-based Visualization and Exploration of Biological Data

Introduction

- **Biological data**

- Huge (*Decryphon 39G*)
- Heterogeneous

- **Visualization & exploration**

- visualization and exploration paradigm
- manage various biological data
- well-adapted to biologist's requirements

- **Using virtual reality to manage huge data**

- Bigger visualization space

- 8m²

- Immersive visualization

- Stereoscopy (3D and relief)

- Intuitive interaction

- 3D Mouse
- Voice
- Gesture

- **Using biological databanks specificities to manage heterogeneous data**

- Focused and centered on a specific biological object of interest

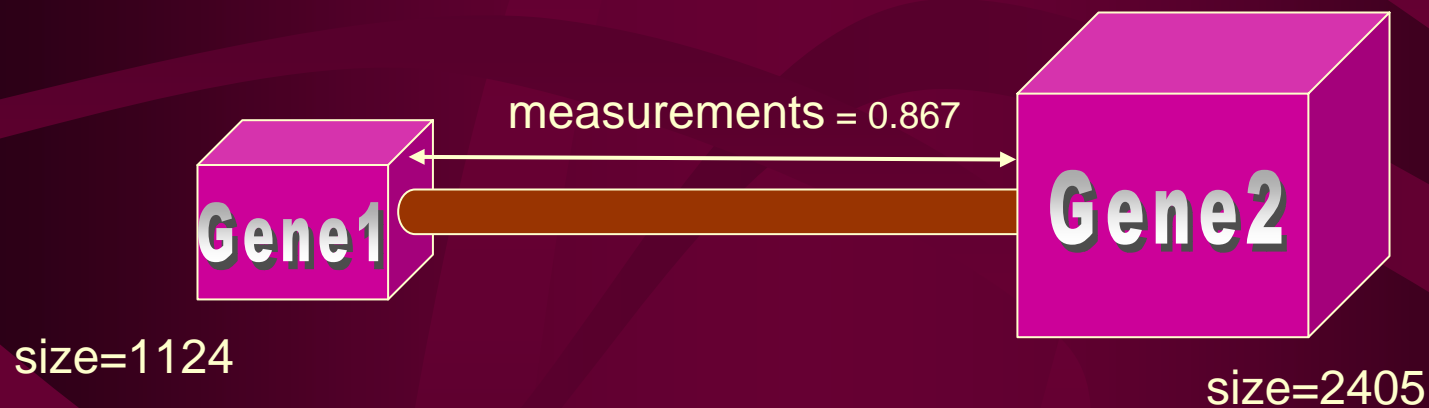
- Protein sequences (*SwissProt, Decryphon...*)
- DNA sequences (*GenBank, EMBL...*)
- Biological terms (*Pubmed*)

- With an attributes list

- Linked by numerical measurements

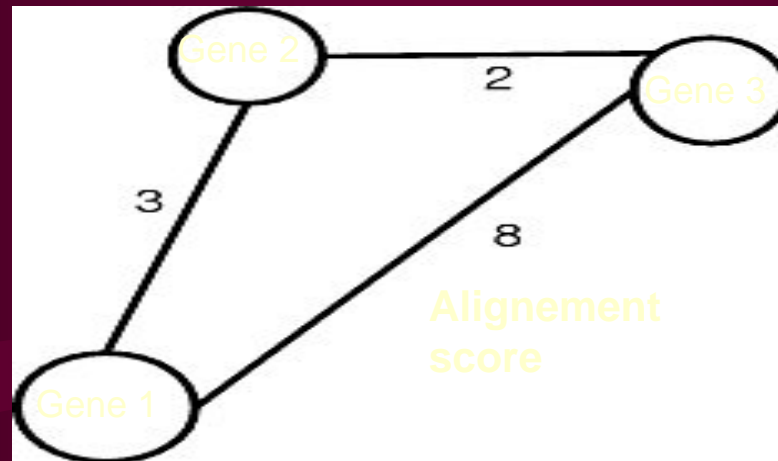
- Sequence alignment score (*BLAST, ZScore...*)
- Term co-occurrences

- **XML Description language for biological data**
 - multi-valued objects (symbolic or numerical)
 - binary multi-valued relationships (symbolic or numerical)
- **3D Graph-based visualization**

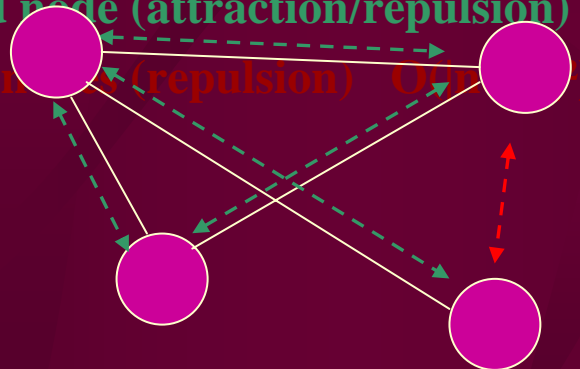


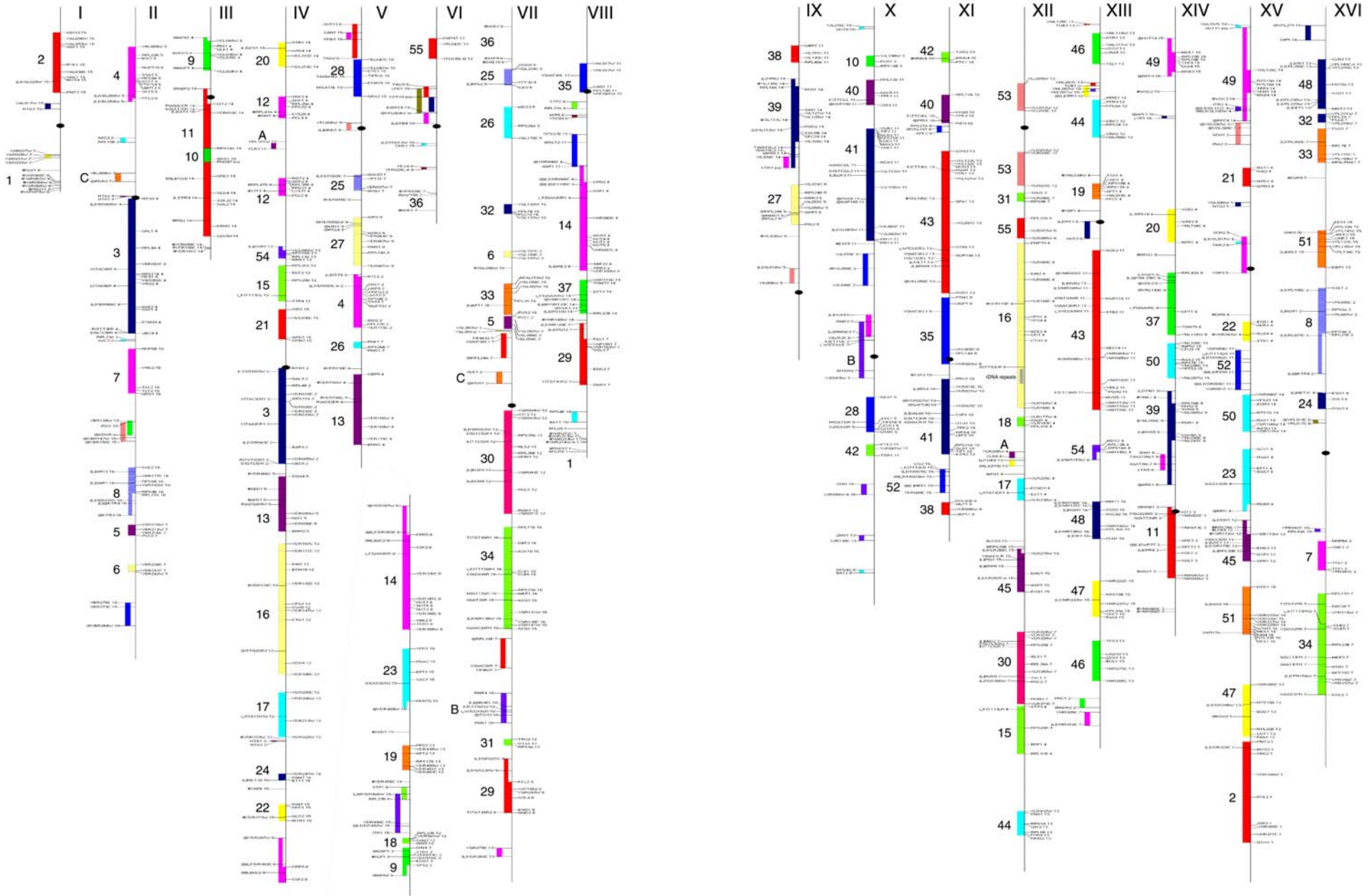
Graph-based representation problems

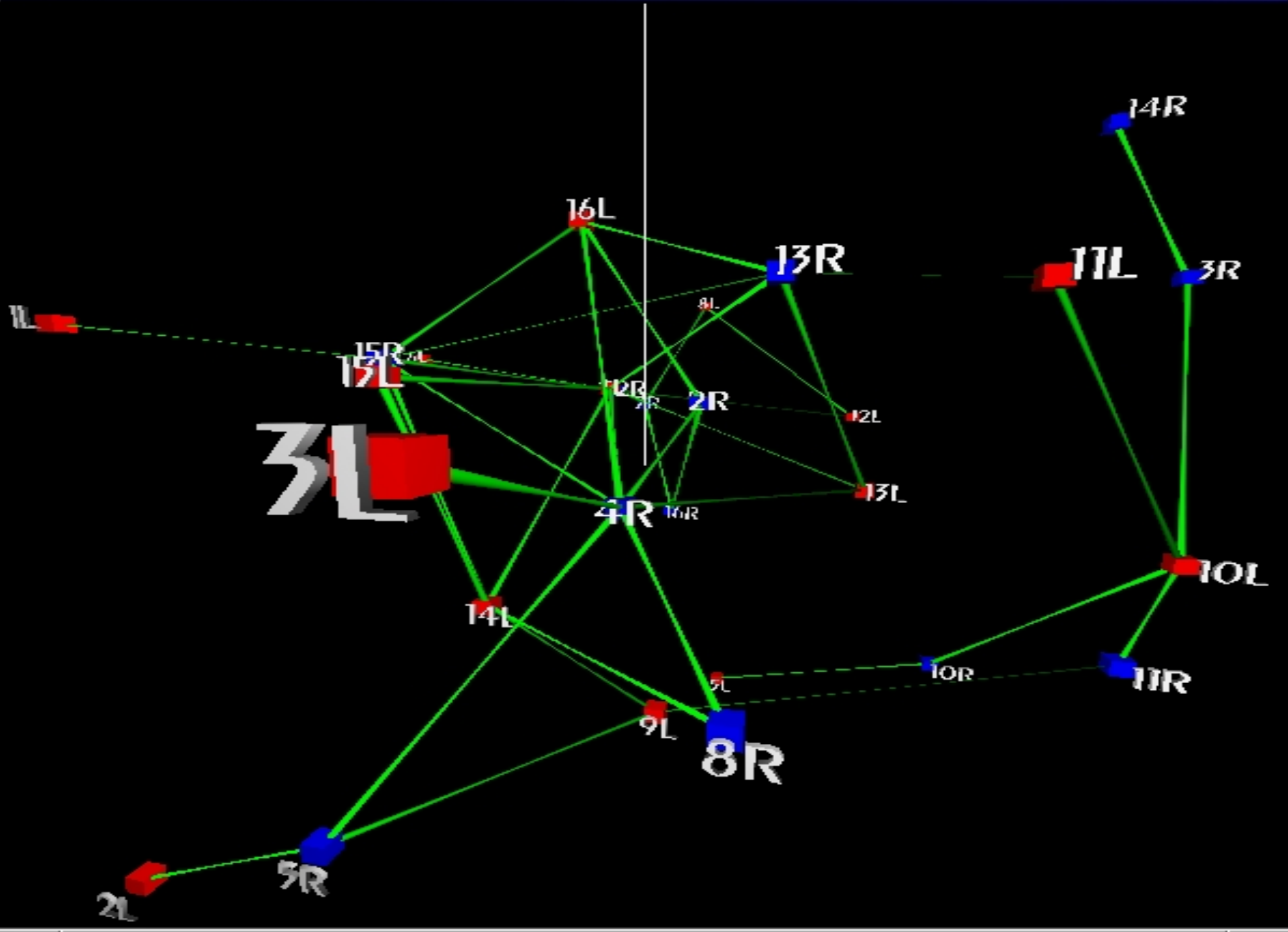
- Insolvable geometrical constraints



- Solution : Force-directed placement algorithm (Eades)
 - Force between linked node (attraction/repulsion) $O(|\text{edge}|)$
 - Force between other nodes (repulsion) $O(|E|)$







Decrypthon: a tool for comparative proteomics

IBM, CNRS, Genomining

- **Goals**
 - A database of all pairwise similarities between proteins.
 - An annotation resource
- **Applications**
 - Identification of conserved functional domains
 - Reconstruction of families of homologous proteins
 - Metabolic pathways reconstruction

The Decrypthon task

Data:

560000 proteins extracted from various well known public databases.

Protein comparison algorithm:

Optimal local sequence alignment (Smith and Waterman, 1981).

The computation process

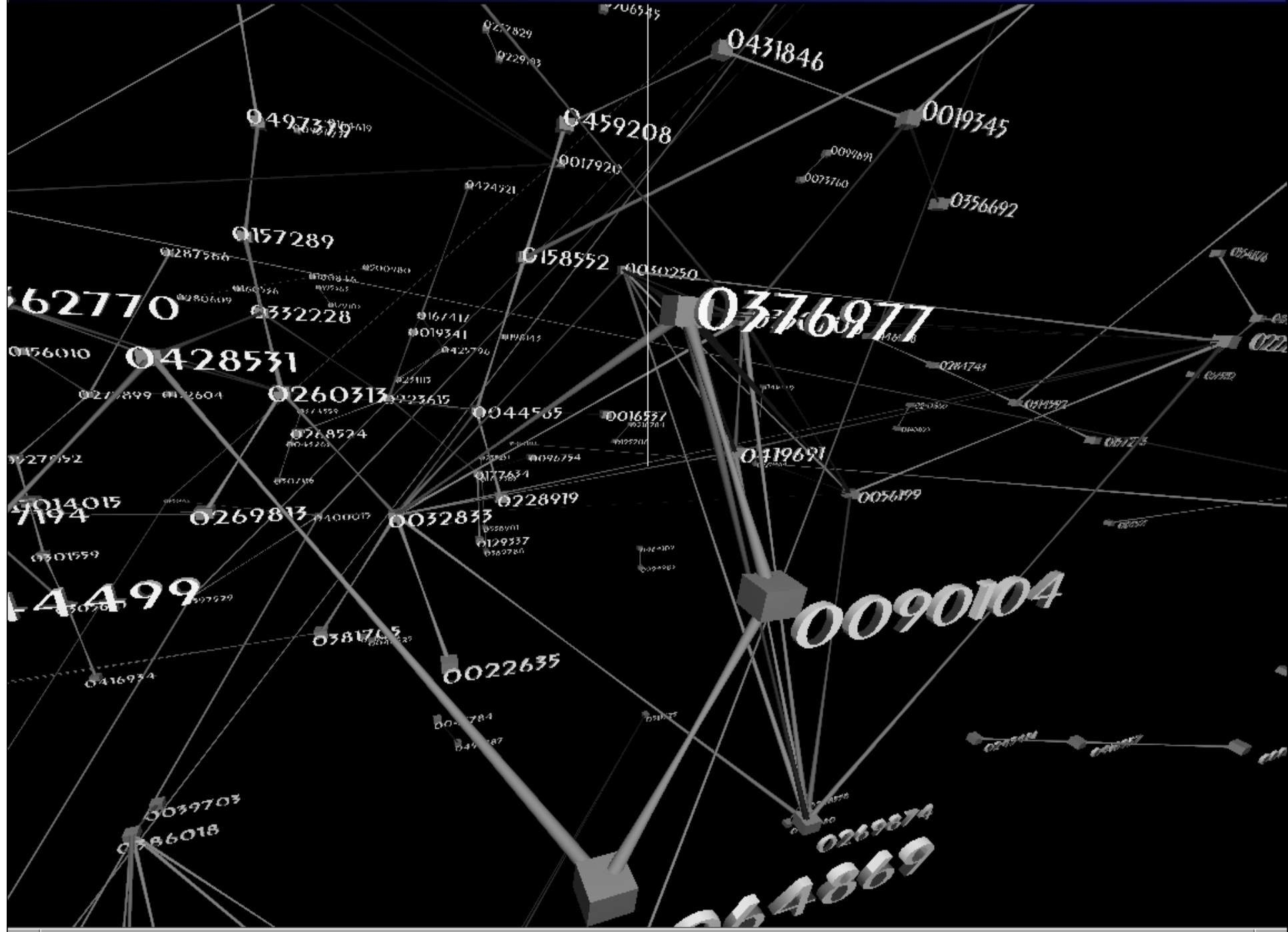
560,000 sequences

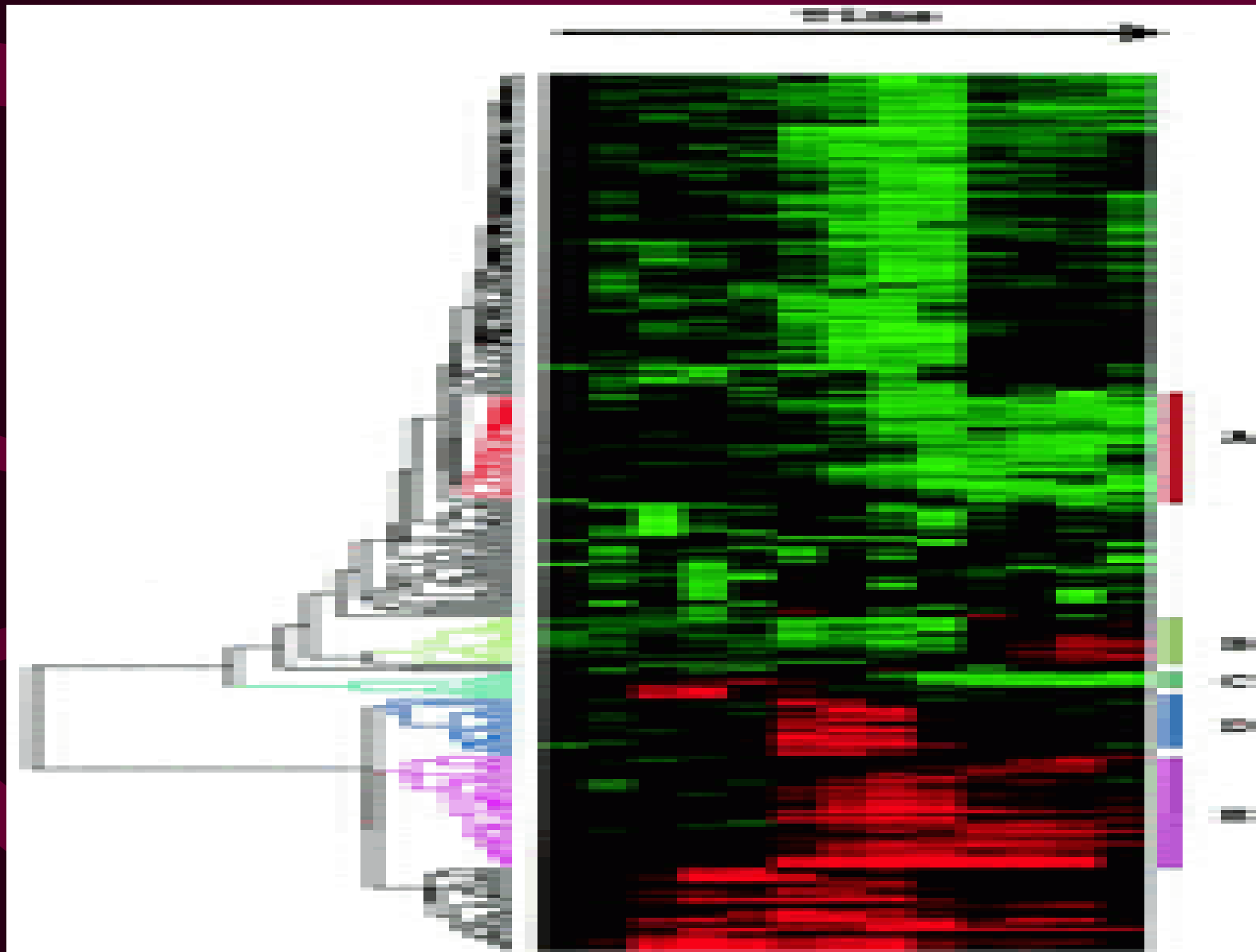


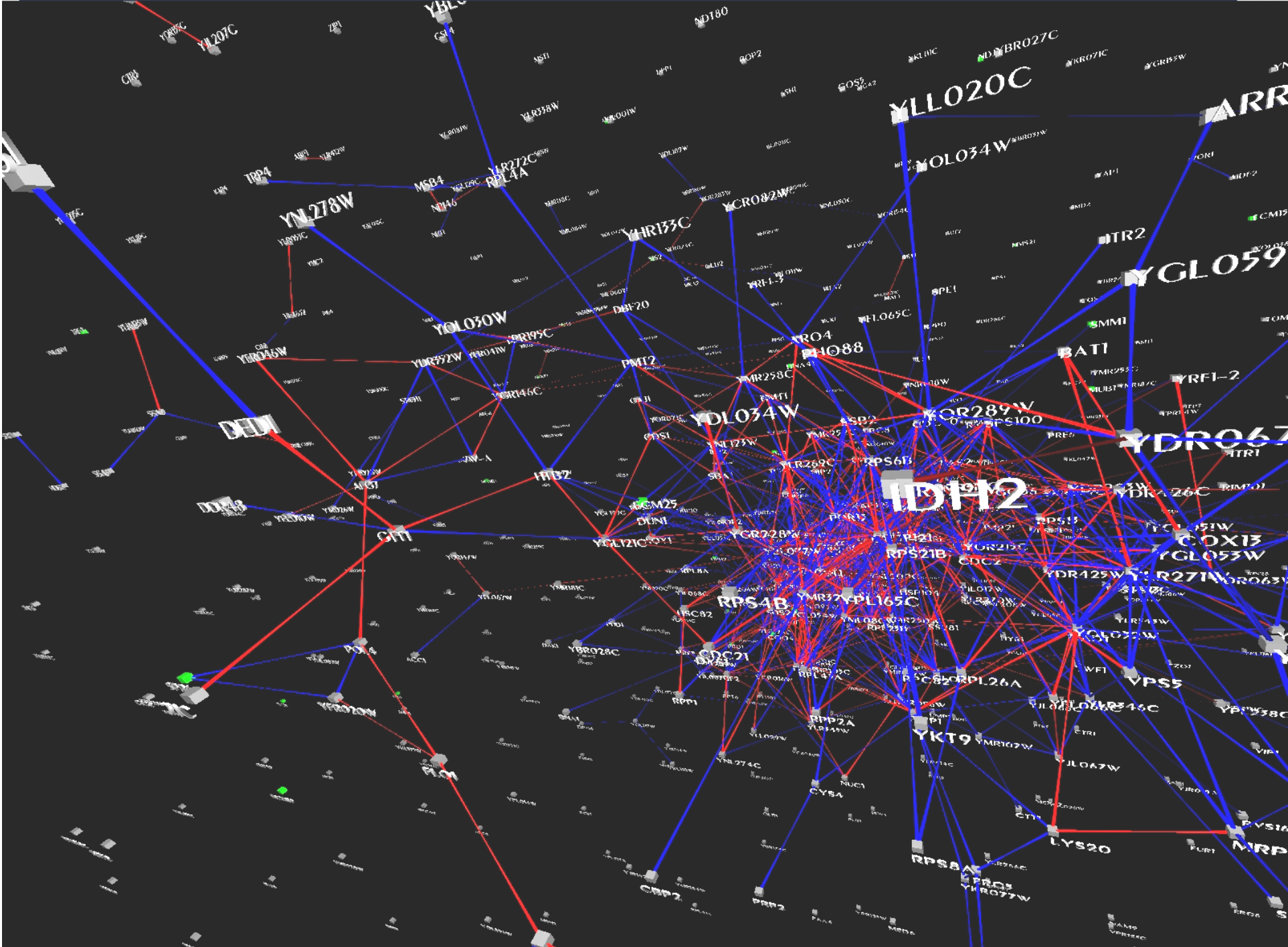
560,000 sequences

The results have been released

- 295,000,000 reported results
- Repository :
<http://infobiogen.fr/services/decryphon/>
 - (17,000 files ~ 39 Go)
- The Decryphon Search Engine
 - As access paradigm

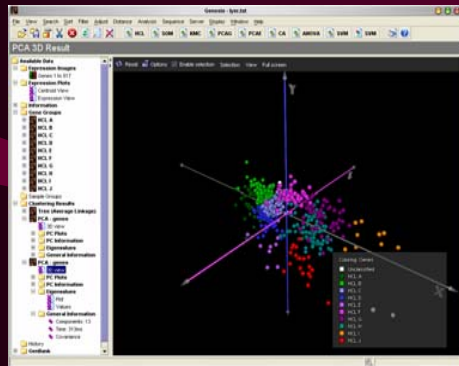






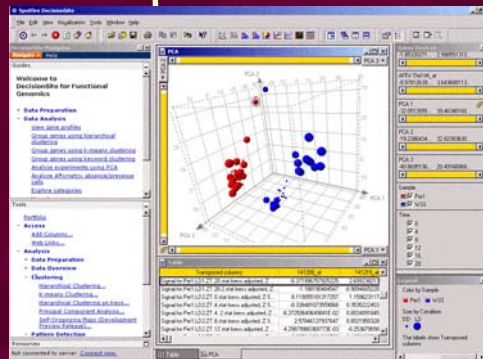
Analyse et visualisation de données

Genesis



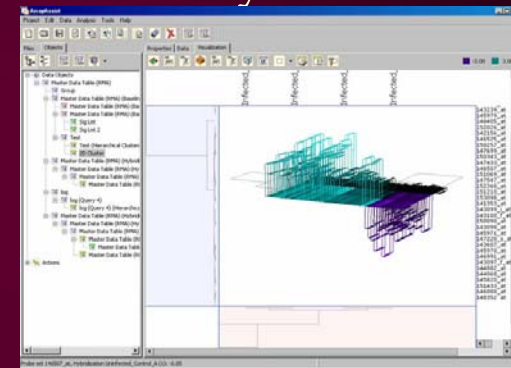
TUG

Spotfire



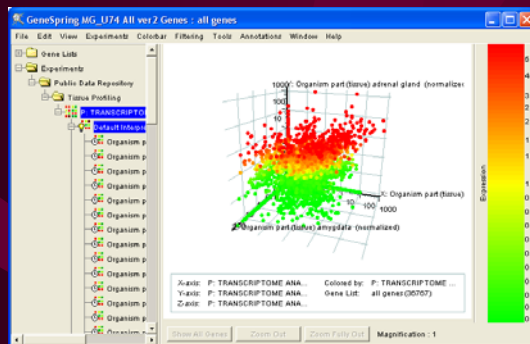
BimCore

ArrayAssist



Stratagene

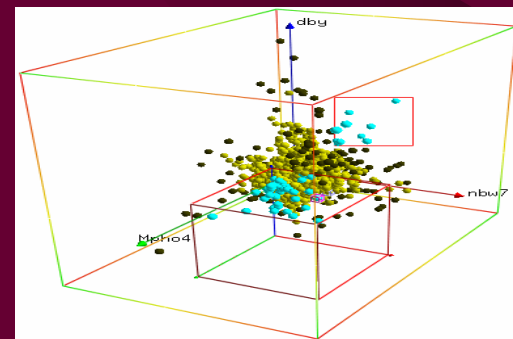
GeneSpring



SiliconGenetics

UP-XI Master Rech - TIG, 2004-2005

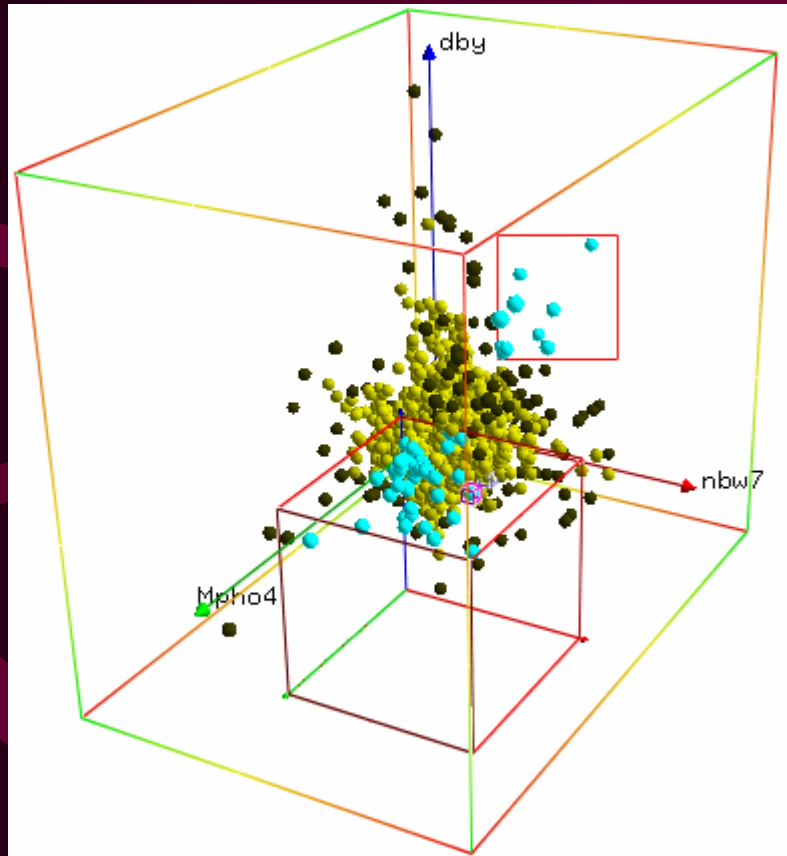
Genebox




UC davis


LIMSI-CNRS, Université Paris-Sud XI

Genebox : Fonctionnement



Controls

 Rotate GeneBox

 Zoom

Translate GeneBox +

SelectionBox +

RangeSelection +

Draw all Genes

Draw Selected Genes

Hold Selection

Intersection

Search

#Gene 0

#SelectedGene 0

Near plane 1.0

Far plane -1.0

Field of View 30.0

Quit

Significance Threshold 0.0

Significance granularity 0.15

#genes 6528

GeneSize 0.1

Show icons

PickGene

Remove similar genes

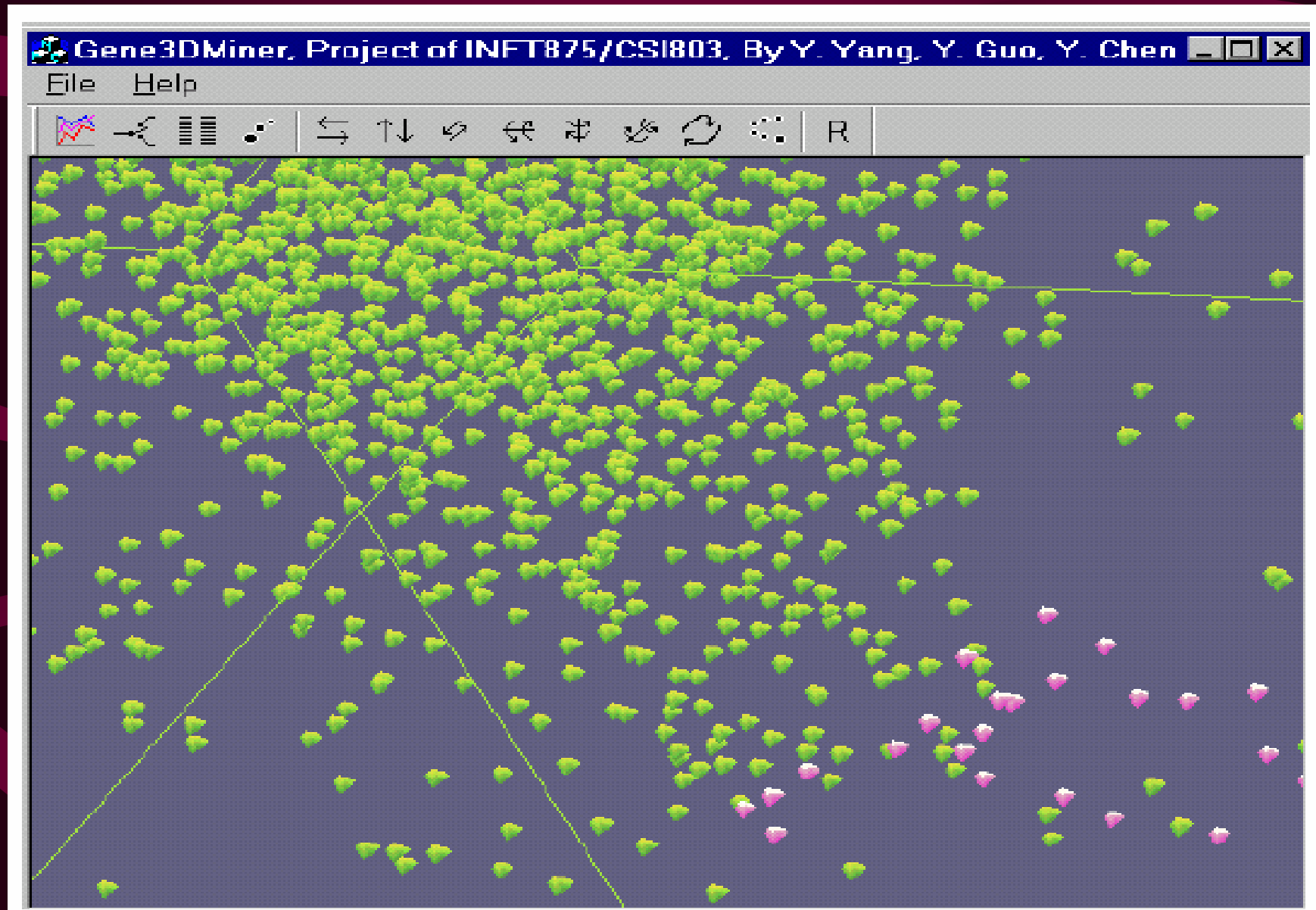
List selected genes

red-blue

Configure

Configure axes +

Mark Selection +



Gene3DMiner

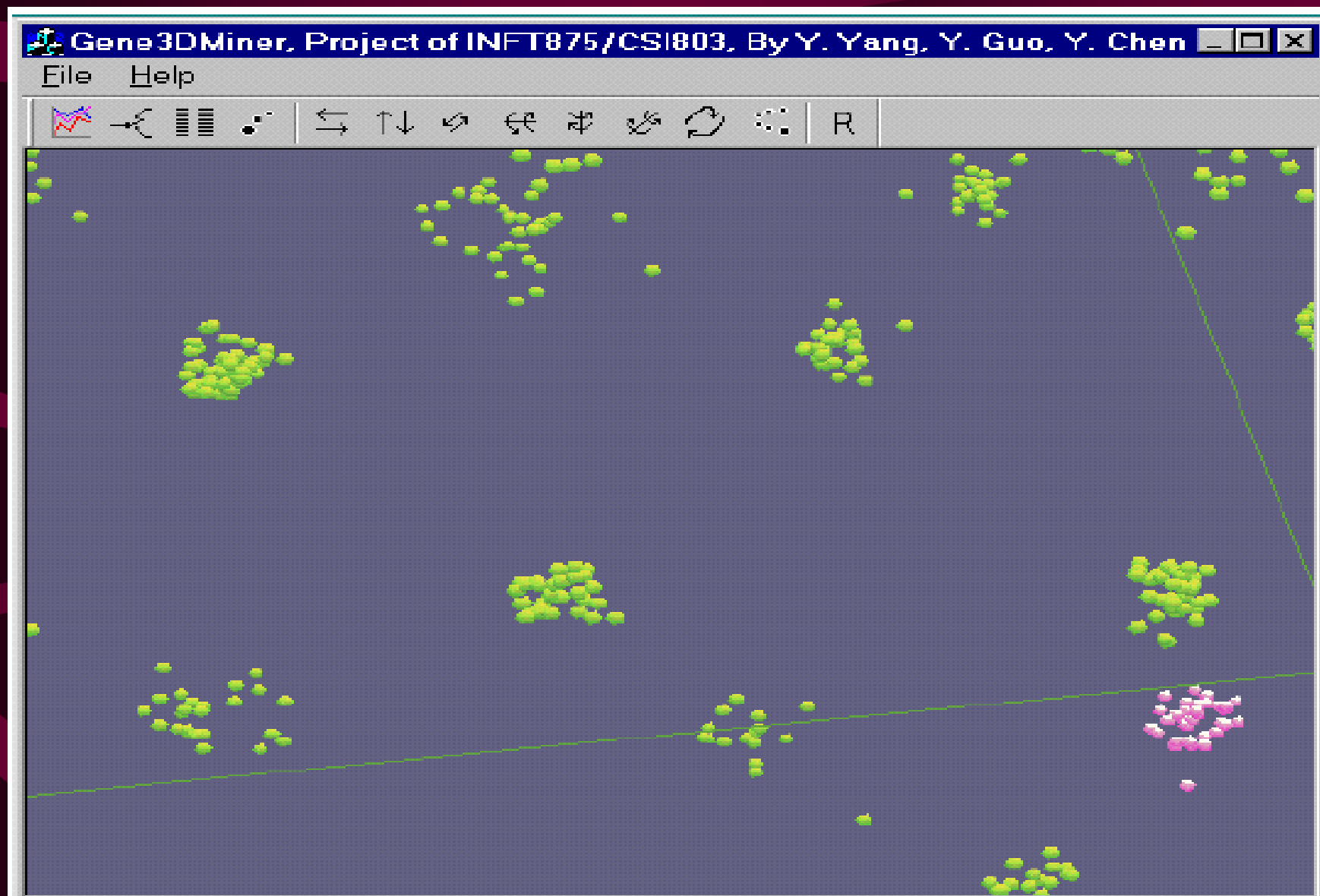
Un des objectifs des scientifiques est de grouper les n gènes selon les similarités de leurs données de temps-série.

Plusieurs algorithmes ont été élaborés dans ce but

- Espace 3D de dessin des gènes
- Nous récupérons les résultats de l'ACP en 3 dimensions afin de les afficher en 3D
 - Les membres de gènes appartenant au même groupe sont colorés.
- Nous pouvons voir les gènes sous différents angles.
- L'inconvénient, est que la scène paraît surpeuplée, et certains gènes se chevauchent.

b. Dessin des gènes divisé par groupe de gènes

- Division de l'espace 3d en m petits cubes, chaque cube affichant le contenu d'un groupe de gènes
 - Ceci permet de se débarrasser de l'inconvénient énoncé avant.
 - Le degré de chevauchement est significativement réduit, ce qui rend l'affichage plus clair.
 - Cet affichage donne aussi 6 degrés de liberté de manipulation dans l'espace 3d.





Problèmes de compréhension en fouille visuelle

- Avec objectif
- Descriptive
- Procédurale
- Interprétative
- Navigationnelle

Où suis-je?

Qu'est ce que je peux
faire avec cette
visualisation?



Pourquoi mesurer la compréhension visuelle

- Pour caractériser
 - Subjectif/objectif
- Pour évaluer
- Pour prédire
- Pour améliorer et affiner
- Pour augmenter la performance de l'utilisateur