

# Le thème transversal Informatique et Vivant

Marie-Dominique Devignes  
*Equipe Orpailleur*

« Développer des modèles et algorithmes pour la biologie et la santé »



1

## Plan de l'exposé

- I. Présentation du thème
- II. Equipes concernées et interactions entre elles
- III. Résultats : bilan des 4 dernières années
- IV. Prospective



2

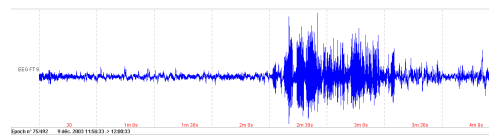
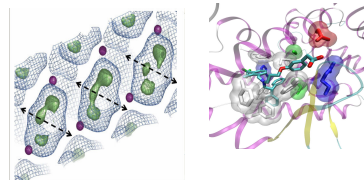
# I. Présentation du thème Informatique et Vivant



3

## Informatique et Vivant : quelles relations ?

- L'informatique modélise le vivant
  - **Modéliser pour simuler et comprendre**
    - Biomolécules et leurs interactions
    - Systèmes, organismes et populations
  - **Modéliser pour maîtriser**
    - Applications en santé
- Le vivant défie l'informatique
  - **Masses de données**
  - **Données complexes (structurées, temporelles, ...)**
  - **Propriétés émergentes**
- Autres disciplines impliquées
  - **Mathématiques**
  - **Physique, chimie**
  - **Etc.**



4

## Informatique et Vivant : historique au Loria

- 2000-2004 : Bioinformatique et Applications à la Génomique
  - *Alexander Bockmayr et l'équipe-projet ModBio*
  - *Autres équipes : Adage, L&D puis Orpailleur*
- 2004-2007 : D'un PRST à l'autre : projet Modélisation des Biomolécules et de leurs Interactions
  - *MD Devignes et l'équipe-projet Orpailleur*
  - *Autres équipes : Adagio, Alice, ABC, Algorille*
- 2008 : Axe transversal Informatique et Vivant
  - *Regrouper les activités ayant comme domaine d'étude ou d'application le vivant*
  - *Bioinformatique + Neuro-informatique + Comportement + Santé ...*



5

## Des domaines d'application variés

- Biomolécules et interactions
  - *Conception de nouveaux médicaments*
  - *Catalyseurs biologiques, biotechnologies*
  - *Mécanismes du cancer*
  - *Interactions plantes – microorganismes*
- Organismes vivants et comportements
  - *Perception visuelle, de la parole*
  - *Systèmes bio-inspirés (sociétés animales, réseaux de neurones physiologiques...)*
  - *Robotique*
- Santé et autonomie
  - *Imagerie médicale et radiochirurgie*
  - *Aide à la décision dans les protocoles thérapeutiques*
  - *Autonomie des personnes âgées ou handicapées*
  - *Plan Alzheimer*



6

## Objectifs scientifiques du thème :

Relever les « défis du vivant » en proposant ou en améliorant les modèles et algorithmes pour le vivant

- Masses de données :
  - *Analyser de façon performante des données volumineuses et hétérogènes issues des technologies avancées en génomique, neurosciences ou imagerie médicale.*
- Données complexes :
  - *Améliorer la modélisation du vivant par la prise en compte des connaissances du domaine et par des approches multi-modèles.*
  - *Coupler modélisation, simulation, visualisation*
- Emergence de propriétés dans les systèmes biologiques
  - *Modéliser l'émergence de propriétés nouvelles dans un système par rapport à l'ensemble des propriétés de ses composants.*



7

## Objectifs stratégiques du thème

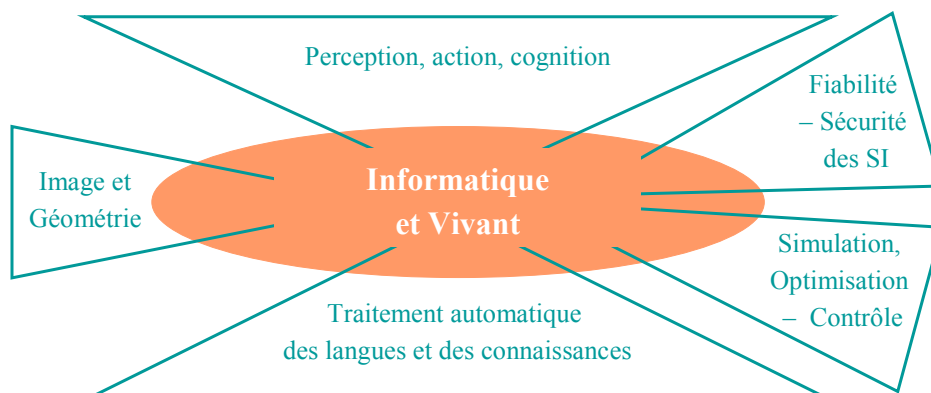
- Favoriser les interactions entre les différentes activités : repérer des approches et méthodologies communes ou complémentaires.
- Fournir un environnement stimulant à des thématiques minoritaires dans certaines équipes.
- Maintenir une plateforme de mise en commun des ressources et mutualisation des compétences.
- Proposer un point d'entrée structuré et informatif sur les activités du thème, notamment pour les biologistes.
- Favoriser le transfert et la valorisation



8

## II. Equipes concernées et interactions entre elles

### Le thème transversal et les axes prioritaires du LORIA



## Informatique et Vivant : Equipes concernées

Axe(s)	Equipe	Biomolécules et Interactions	Organismes et comportements	Santé et Autonomie
TALC	ORPAILLEUR	XX		X
PAC	ABC	X		
I&G	ADAGIO	X		
I&G	ALICE	x		
FSSI	ALGORILLE	x		
PAC	CORTEX		XX	X
PAC	MAIA		XX	X
PAC - TALC	PAROLE		X	X
SOC (INRIA)	CORIDA		X	
SOC (INRIA)	TOSCA		X	
I&G – PAC	MAGRIT			XX
SOC (INRIA)	MASAIE			X
FSSI - PAC	TRIO			x



11

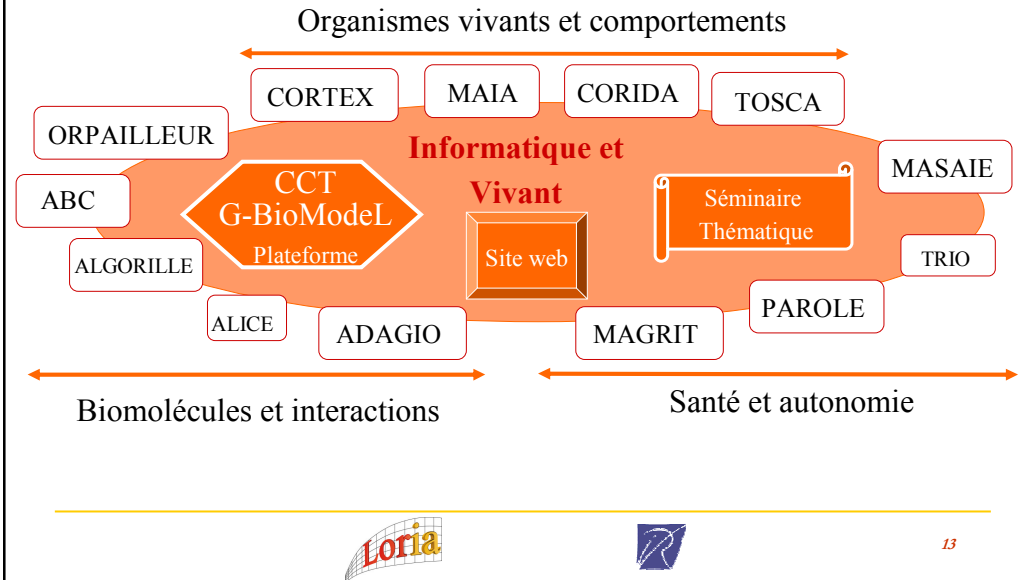
## Informatique et Vivant : diversité des approches

Axe(s)	Equipe	Approches / méthodes informatiques principales
TALC	ORPAILLEUR	Méthodes symboliques de fouille de données, ECBD, HMM
PAC	ABC	Apprentissage statistique
I&G	ADAGIO	Géométrie discrète, Algorithmique des mots
I&G	ALICE	Géométrie différentielle, Visualisation scientifique
FSSI	ALGORILLE	Calculs distribués sur la grille
PAC	CORTEX	Modèles connexionnistes neuromimétiques
PAC	MAIA	Systèmes multi-agents, modèles stochastiques
PAC - TALC	PAROLE	Modèles stochastiques, HMM, réseaux bayésiens
SOC (INRIA)	CORIDA	Equations aux dérivées partielles
SOC (INRIA)	TOSCA	Modèles stochastiques, Biostatistiques
I&G – PAC	MAGRIT	Réalités augmentée, fusion d'information
SOC (INRIA)	MASAIE	Automatique, systèmes dynamiques
FSSI - PAC	TRIO	Architectures temps réels distribuées, résistance aux fautes

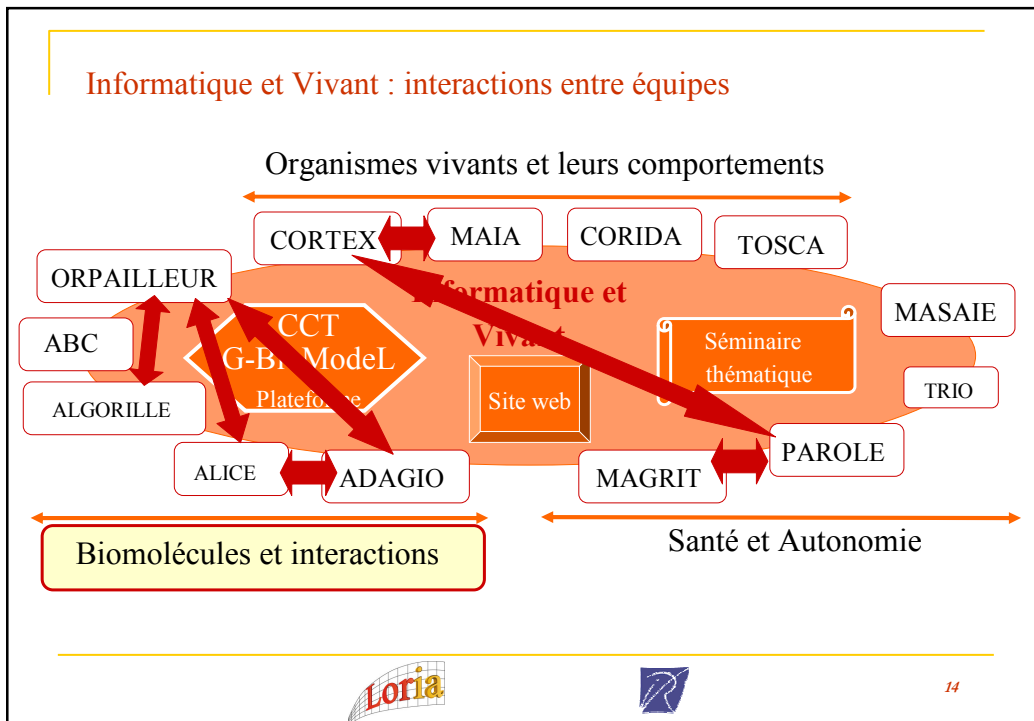


12

## Informatique et Vivant : dispositifs d'animation



## Informatique et Vivant : interactions entre équipes



# III. Résultats – Bilan des 4 dernières années

A. 2004-2005-2006 : Thème Bioinformatique et Applications à la Génomique  
du PRST Intelligence Logicielle

B. 2007: Thème Modélisation des Biomolécules et de leurs Interactions  
du PRST Modélisation, Information et Systèmes Numériques



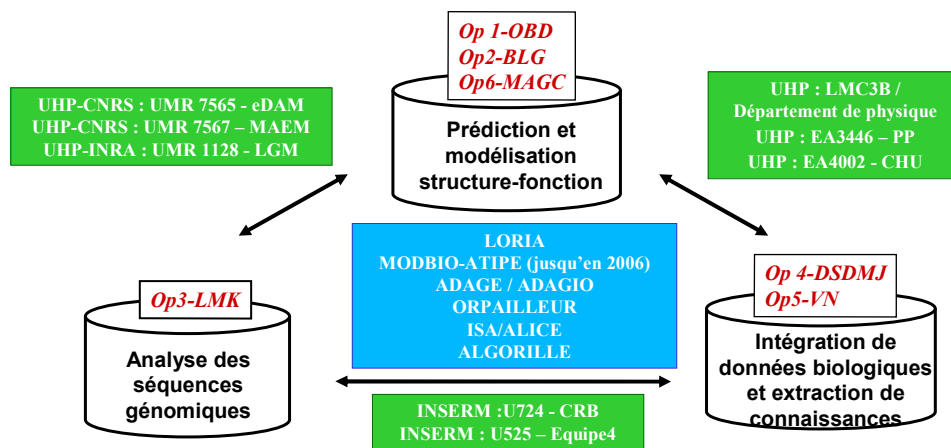
15

2004-2006 : Bioinformatique et Applications à la Génomique

2007 : Modélisation des Biomolécules et de leurs Interactions

## Bioinformatique et Applications à la Génomique

*Emergence de 3 pôles de compétences*

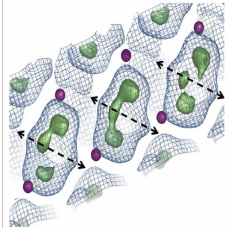


16



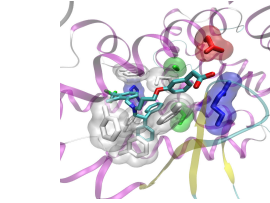
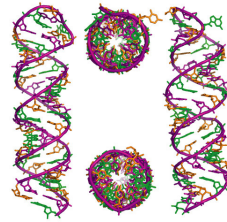
## 1. Résultats scientifiques majeurs (1/2)

### **Pôle 1 : Prédiction et modélisation des structures et des fonctions des macromolécules biologiques**



✓ Modèle 3D des ARN triloop (CUG)<sub>n</sub> (Leclerc et al., unpublished)

✓ Modélisation des enveloppes moléculaires : application de techniques d'optimisation pseudo-booléennes (Ourjountsev et al., *Acta Crystallographica D*, 2006, lectin SML2)



✓ Criblage virtuel des interactions protéines-ligands (Beautrait et al., *J. Mol Modelling*, 2008)

## 1. Résultats majeurs scientifiques et technologiques (2/2)

### **Pôle 2 : Analyse des séquences génomiques**

- ✓ Modélisation de l'épissage alternatif (Eveillard et al., 2004 ; Bockmayr et al., 2005)
- ✓ Identification des sites de fixation des facteurs de transcription par approche comparative (Touzain et al., 2005, 2008 article accepté dans *BMC Bioinformatics*)

### **Pôle 3 : Intégration de données biologiques et extraction de connaissances**

- ✓ Treillis de concepts et ontologies pour l'interrogation d'un annuaire de sources de données biologiques (Messai et al., 2005)
- ✓ Intégration de données génétiques et cliniques guidée par les connaissances et en vue de la fouille (Coulet et al., 2006a, 2006b +article sous presse dans *BMC Bioinformatics*)
- ✓ Analyse de cohorte et extraction de motifs rares ( Szathmary et al., 2006 – Prix du meilleur article d'EGC 2006)

## 2. Evolution des compétences

- Nouveaux domaines d'application pour les méthodes informatiques
  - *Programmation entière binaire <-> cristallographie (Modbio, LCM3B)*
  - *HMM <-> séquences biologiques (Orpailleur, LGM)*
  - *Fouille de données symboliques et numériques <-> Analyse de cohortes (Orpailleur, INSERM U525-équipe4)*
- Introduction de la dimension informatique dans la recherche en biologie
  - *Intégration de données hétérogènes récoltées sur le web dans des entrepôts personnalisés (Orpailleur, Laboratoire de Génétique Médicale, CRB)*
  - *Recherche de motifs pour identifier les gènes cibles des PPAR (Orpailleur, EA3446)*
  - *Approche comparative pour l'identification des sites de transcription (Adagio, LGM)*
- **Compétences interdisciplinaires**
  - *Modélisation statistique et formelle d'un système biologique : l'épissage alternatif (Modbio, MAEM)*
  - *Modélisation des interactions protéine-protéine (eDAM, Alice)*
  - *Modélisation des connaissances pour guider l'intégration et la fouille de données (Orpailleur)*



## 3. Valeur ajoutée par la coopération entre les équipes

- Publications communes
  - *Revue internationale : 11*
  - *Conférences (communications orales sélectionnées) : 18*
  - *Posters : 14*
- Contrats communs
  - *ACI Masse de données (2003) et IMPBio (2004) : 5*
  - *ANR : 1 (2006, ANR non thématique)*
  - *Divers : 2 (Décryphon, ARC)*
  - *Contrats internationaux : 4 (Israël, Pologne, Belgique, Luxembourg)*
- Logiciels
  - *BioRegistry, Xcollect*
  - *PromLoc, PpreFinder*
  - *SIGffRid*
  - *VSM-G*
- Thèses co-encadrées
  - *4 soutenues (2003, 2 en 2005, 2007)*
  - *2 à venir (2008)*
- Recrutements
  - *3 MCU, 2 CR, 1PU (dont LORIA : 1MCU, 1CR)*
  - *Rattachement d'1 DR en 2007 au LORIA*



### Le thème MBI du CPER 2007-2013

- ❑ Objectifs : développer des approches à base de connaissance pour la modélisation des biomolécules et de leurs interactions
- ❑ Aborder sous cet angle la modélisation des systèmes biologiques
- ❑ Créer un Centre de Compétences et de Transfert en Génomique et BioModélisation : G-BioModel

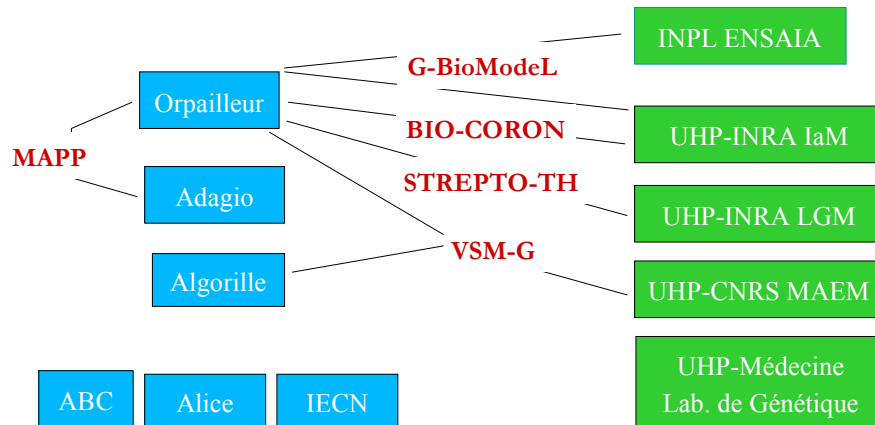


### Participants au projet MBI

- ❑ **LORIA**
  - Orpailleur (présent dans toutes les opérations, environ 10 personnes)
  - Adagio, ABC, Alice, Algorille (environ 10 personnes)
- ❑ **IECN : Institut de Mathématiques Elie Cartan de Nancy**
  - UMR 7502 : CNRS-INRIA-Nancy Université (environ 2 personnes)
- ❑ **MAEM : Maturation des ARN et Enzymologie Moléculaire**
  - UMR 7567 : CNRS-Nancy Université (3 personnes)
- ❑ **ENSAIA : Laboratoire Biocatalyse-Bioprocédés**
  - Nancy Université, INPL, (3 personnes)
- ❑ **IaM : Interactions Arbres-Microorganismes**
  - UMR 1136 : INRA-Nancy Université (4 personnes)
- ❑ **LGM : Laboratoire de Génétique et Microbiologie**
  - UMR 1128 : INRA-Nancy Université (3 personnes)
- ❑ **Faculté de Médecine**
  - Laboratoire de Génétique (2 personnes)



MBI : Opérations et Interactions entre les équipes



Bilan 2007 (1/4) : Opération STREPTO-TH

« Transfert Horizontal chez les Streptocoques »

- Objectifs de l'opération
  - Mettre en évidence par des méthodes sans a priori (HMM2) des hétérogénéités dans les séquences des génomes entièrement séquencés de Streptocoques.
  - Caractériser les régions atypiques (contenu en gènes, origines possibles, etc.).
  - Valider la méthode
- Thèse en cours
  - Catherine Eng
- Résultats 2007 : Analyse de régions atypiques
  - Post-traitements des résultats de l'analyse par HMM.
  - Article en cours de soumission dans BMC Bioinformatics concernant l'application des HMM2 au génome de *Streptomyces coelicolor* A3(2) (cf précédent CPER)

## Bilan 2007 (2/4) : Opération MAPP

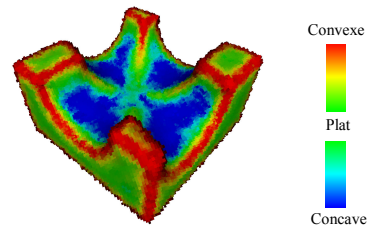
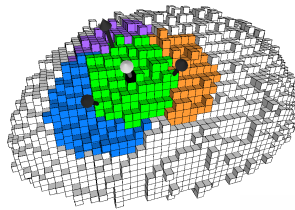
### « Modélisation des Assemblages Protéine-Protéine »

- Objectifs
  - Identifier rapidement les modes d'assemblage les plus pertinents entre plusieurs unités protéiques
  - Manipuler interactivement et analyser les interfaces protéine-protéine ainsi créées.
- Thèses en cours :
  - *Laurent Provot, Mathieu Chavent*
- Résultats
  - Propositions pour améliorer avec les méthodes de la géométrie discrète la recherche d'informations géométriques sur les surfaces des protéines.
  - Geometric Feature Estimators for Noisy Discrete Surfaces, article accepté à la conférence internationale DGC I 2008 (Lyon 16-18 avril 2008).

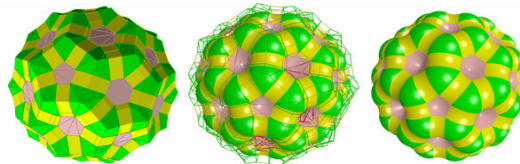


## Opération MAPP : illustrations

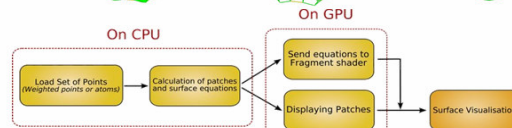
- 1) Calculs de « patches » à la surface de l'objet : utilisation d'un paramètre d'épaisseur pour une prise en compte du bruit



- 2) Extraction de caractéristiques géométriques à partir des patches (estimateur d'aire, vecteur normal, discrimination des zones plates, convexes, concaves).



- 3) Visualisation haute-performance des surfaces par « lancer de rayon » (MetaMol : collaboration avec ALICE)



## Bilan 2007 (3/4) : Opération VSM-G

### « Criblage Virtuel des Interactions Protéine-Ligand »

#### ■ Objectifs

- Améliorer les techniques de criblage virtuel (des interactions protéine-ligand), en particulier par l'introduction de contraintes sur les ligands, exprimées à partir de résultats d'ECBD, et par l'augmentation des performances de calculs sur grille ou clusters (Orpailleur - Algorille)
- Améliorer les techniques de docking spécialisées dans le domaines des complexes protéine-protéine et protéine - acide nucléique (Orpailleur-MAEM)

#### ■ Thèses

- *Alexandre Beautrait soutenue le 15 janvier 2008*
- *Léo Ghemtio (1ère année)*

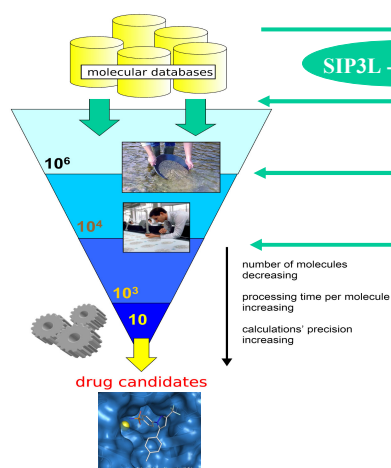
#### ■ Résultats

- Modèle structural de ligand inhibiteurs pour le récepteur c-Met
- Modélisation 3D d'un ARN (CUG)<sub>17</sub> et recherche de petites molécules capables de s'y fixer.



## Opération VSM-G : illustrations

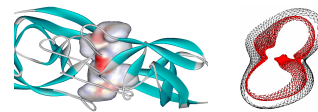
VSM-G un entonnoir de criblage



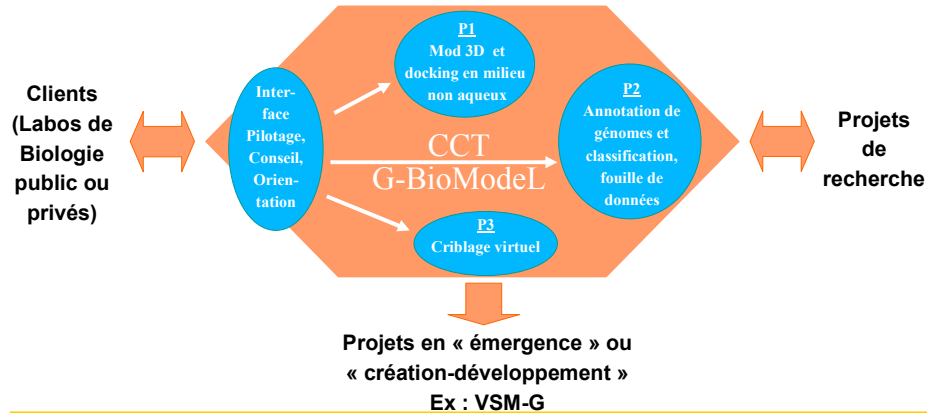
Fouille de données et extraction de connaissances sur les cibles (protéines) et leurs ligands en vue d'exprimer des contraintes sur les différentes étapes du criblage

SHEF - MSSH2 : Docking géométrique fondé sur les coefficients des fonctions harmoniques sphériques

Ligand	<i>deflation</i>
Protéine	<i>inflation</i>



## Bilan 2007 (4/4) : Opération CCT G-BioModel Centre de Compétences et de Transfert pour la Génomique et la Modélisation des Biomolécules en Lorraine



## CCT G-BioModel : Etat des lieux

- 2007-2009 : 3 projets pilotes pour explorer les modalités de fonctionnement
  - P1 : Docking en solvant non aqueux (ENSAIA-LBB)
  - P2 : Annotation des génomes de champignons (UHP-INRA Iam)
  - P3 : Criblage virtuel de ligands de la protéine Met par le logiciel VSM-G (UHP-IBDML)
- Equipe de pilotage
  - Aspects Scientifiques : MD Devignes, B Maigret, M Smail-Tabbone
  - Aspects Techniques : L Pierron (Ingénieur du SED à temps partiel)
  - Chef de Projet : M Souchet (en attente de poste)
- Equipement (OIP 2007 du CPER)
  - Serveur de données : 15 disques de 500 Go = 7,5 To
  - Serveur de calcul : 16 unités bi Quad = 128 processeurs, reliés en Infiniband
  - Serveur d'applications
- Personnel temporaire
  - 1 CDD (à prolonger) pour l'installation de logiciels pour la bioinformatique
  - Post-docs ou thésards financés par les opérations pilotes

## IV. Perspectives

Rappel des objectifs et des défis à relever :

**« Développer des modèles et algorithmes pour la biologie et la santé »**

- Masses de données :
  - *Analyser de façon performante des données volumineuses et hétérogènes issues des technologies avancées en génomique, neurosciences ou imagerie médicale.*
- Données complexes :
  - *Améliorer la modélisation du vivant par la prise en compte des connaissances du domaine et par des approches multi-modèles.*
  - *Coupler modélisation, simulation, visualisation*
- Emergence de propriétés dans les systèmes biologiques
  - *Modéliser l'émergence de propriétés nouvelles dans un système par rapport à l'ensemble des propriétés de ses composants.*



## Perspectives scientifiques (1/2)

- Biomolécules et interactions / Organismes et comportements
  - Au-delà de la diversité des domaines d'application, collaborer sur des voies d'approches communes pour des problèmes communs :
    - **Prise en compte des connaissances du domaine dans les modèles existants**
    - **Approches multi modèles**
    - **Couplage modélisation, simulation, visualisation**
    - **Modéliser l'émergence de propriétés originales dans les systèmes biologiques que ce soit au niveau des ensembles de biomolécules, de cellules ou d'organismes vivants**
    - **Gérer les changements d'échelle à l'intérieur d'un même modèle : modéliser l'abstraction des propriétés d'un système complexe**



33

## Perspectives scientifiques (2/2)

- Santé et Autonomie : poursuivre les avancées récentes
  - Aide au diagnostic et à la décision
    - **Ex : Analyse d'EEG, troubles de la vigilance (CORTEX)**
    - **Ex : Kasimir : protocoles thérapeutiques (ORPAILLEUR – Centre Alexis Vautrin)**
  - Autonomie, maintien à domicile, handicap
    - **Ex : ParaChute (MAIA)**
    - **Ex : Tête parlante (PAROLE)**
    - **Ex : Réseaux de capteurs sans fil (TRIO)**
  - Imagerie médicale
    - **Ex : Simulation en temps réel de la mise en place d'un coil intra-vasculaire pour le traitement des anévrismes, radiochirurgie stéréotactique (MAGRIT-Département de neuro-radiologie du CHU Nancy)**
    - **Ex : repérage de structures 3D lors d'interventions chirurgicales (MAGRIT-Département de radiologie et d'urologie du CHU Nancy)**

*Lien fort avec le transfert et la valorisation  
Responsable-coordonateur de ces activités (Laurent Bougrain, Cortex)*



34

## Animation scientifique du thème I&V

- Séminaire thématique Informatique et Vivant
  - Echange d'information et lieu de formation pour les équipes impliquées
    - *Etat d'avancement des projets*
    - *Mise au point sur des approches communes à plusieurs équipes*
    - *Collaborations et réponses conjointes à des appels d'offres*
  - Invitations des conférenciers extérieurs
    - *Transfrontaliers (Sarrebruck, Luxembourg)*
    - *Et au-delà...*
- Chaires de professeurs invités ou chaires d'excellence européenne
- Organisation de workshop ou conférences
  - Dans la suite de
    - *NeuroComp octobre 2006 (CORTEX)*
    - *WebDim4LS (workshop international, en conjonction avec WISE 2007, ORPAILLEUR)*



35

## Visibilité du thème I&V (1/2)

- Etendre le site web bioinfo.loria.fr aux 3 axes du thème
  - Système de gestion de contenu interactif
    - *Contributions autonomes des projets*
  - Point d'entrée centralisé pour l'extérieur
    - *Accueil des utilisateurs/clients des plateformes*
    - *Vitrine des projets en cours*
  - Annonces des événements
    - *Séminaires*
    - *Thèses*
    - *Evénements extérieurs en lien avec le thème*



36

## Visibilité du thème I&V (2/2)

- Plateformes ouvertes sur l'extérieur, orientées transfert de compétences
  - Ex : CCT G-BioModel
    - **Trouver un mode de fonctionnement pérenne aux plateformes**
    - **Réservoir de données**
    - **Catalyseur de collaborations interdisciplinaires**
  
- Projets de simulation - visualisation
  - **Environnement immersif**
  - **Robotique**



37

## Conclusion

- Informatique et Vivant : thème "prolifique"
  - Nouvelles équipes (MASAIE, projet issus de TOSCA, à suivre...)
  - Représentation dans le thème Bio de l'INRIA (MASAIE, CORTEX, ...)
  - Nécessité de trouver des convergences locales
  
- Atouts du LORIA
  - Ouverture à l'interdisciplinarité
    - **Contacts Fac des Sciences et Fac de Médecine**
  - Projets pilotés par les informaticiens
    - **Souligné dans le cas du PRST MBI**
  - Transfert et Valorisation
    - **RIV, applications en Santé et pour le médicament**



38