

CONTEXTE

Après le séquençage, maintenant accompli, de la majorité des génomes des organismes procaryotes et eucaryotes les plus utilisés en laboratoire, les biologistes sont maintenant face à un nouveau défi: comprendre, à grande échelle, la fonction des composants codés par le génome et notamment des protéines. Sachant qu'aucune protéine n'assure seule sa fonction et que la vie de chaque organisme dépend de dizaines de milliers d'interactions protéine-protéine (IPPs) spécifiques, comment accéder à la "notice de montage" de la cellule qui doit décrire en premier lieu les interactions et/ou le mode d'assemblage spécifique des protéines entre elles au niveau moléculaire ? Nous privilégions 3 niveaux d'étude des interactions dans le cadre du projet EIDIPP : le niveau moléculaire, le niveau des domaines protéiques et celui du réseau d'interaction.

OBJECTIFS PRINCIPAUX

Etude des interactions protéine-protéine au niveau de la Protein Data Bank (équipe 3: Joel Janin, LEBS)

- extraire des données PDB des paramètres discriminant les deux types d'interactions (spécifiques et contacts cristallins) et identifier de façon fiable les interactions spécifiques sur une base structurale
- améliorer les performances des méthodes d'amarage (docking) pour prédire la structure de complexes protéine-protéine à partir de celle de leurs composantes

Rôle des domaines protéiques dans les interactions (équipe 4: Daniel Kahn, LIPM)

- exploiter la combinatoire des domaines que l'on peut détecter dans les paires de protéines en interaction, afin d'identifier de manière non supervisée les couples de domaines sur-représentés
- calculer des probabilités d'interaction entre protéines, sur la base de leurs arrangements en domaines

Etude du réseau des protéines en interaction chez plusieurs organismes modèles (équipe 1: Bernard Jacq, LGPD)

- développement et automatisme de la méthode Prodistin de comparaison fonctionnelle entre protéines à partir des réseaux IPPs
- analyse et comparaison de réseaux d'IPP et d'arbres de classification pour différentes espèces

Outils informatiques et mathématiques pour l'étude des interactions protéine-protéine (équipe 2: Alain Guénoche, LIM et équipe 5: Irène Charon, ENST)

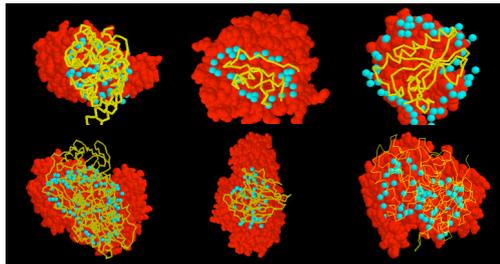
- Recherche de classes denses dans les graphes
- Etude de la robustesse de méthodes de distance
- Métaheuristiques des cliques maximales

RESULTATS (1)

- 11 articles publiés, acceptés ou soumis, 4 sont en préparation.
- poursuite de l'expérience internationale CAPRI (Critical Assessment of PRedicted Interactions) qui teste la capacité des algorithmes d'amarage (docking) à prédire la structure de complexes protéine-protéine à partir de celle des composantes séparées.
- étude de l'interaction protéine-protéine dans les assemblages multi-composantes.
- développement d'un modèle probabiliste exprimant la probabilité d'observer une paire de domaines sous une hypothèse de non-interaction.
- estimation des fréquences de co-occurrence de domaines de ProDom dans une liste multi-espèce de paires de protéines en interaction.

ILLUSTRATION (1)

contacts cristallins 1A7H, 1FEH complexes 1PPF, 1DHK dimères 1KPF, 1BMD



Les interfaces cristallines contiennent comparativement plus d'eau à l'intérieur que les interfaces dimériques où l'eau forme généralement une couronne ceinturant le cœur du patch.

- On peut observer 1 ou plusieurs couronnes de molécules d'eau s'expliquant par le caractère hydrophobe marqué du cœur des interfaces dimériques.
- Pour les complexes, la distribution est intermédiaire entre celle des grands contacts cristallins «humides» et celles des interfaces dimériques «sèches».
- Cependant la distribution des molécules d'eau dans ces 2 catégories d'interface ne permet pas une discrimination entre interfaces spécifiques et non-spécifiques.

LABORATOIRES IMPLIQUES

- Equipe **Bernard JACQ**, Laboratoire de Génétique et Physiologie du Développement (LGPD), MARSEILLE
- Equipe **Alain GUÉNOCHE**, Institut de Mathématiques de Luminy (IML), MARSEILLE
- Equipe **Joël JANIN**, Laboratoire d'Enzymologie et Biochimie Structurales (LEBS), GIF/YVETTE
- Equipe **Daniel KAHN**, Laboratoire des Interactions Plantes-Microorganismes (LIPM), CASTANET-TOLOSAN
- Equipe **Irène CHARON**, Département Informatique et réseaux, ENST,

MOTS-CLES

Interactions protéine-protéine, Protein data bank, interface de contact, domaine, ProDom, réseau d'interactions, classification

PUBLICATIONS PRINCIPALES

Brun C, Herrmann C, Guénoche A. Clustering proteins from interaction networks for the prediction of cellular functions. BMC Bioinformatics. 2004; 5(1):95.

Baudot A, Jacq B, Brun C. A scale of functional divergence for yeast duplicated genes revealed from analysis of the protein-protein interaction network. Genome Biol. 2004; 5(10):R76.

Guénoche A, Clustering by vertex density in a Graph Meeting of the International Federation of the Classification Societies, IFCS'2004, in Classification, Clustering and Data Mining, D. Banks et al. (Eds.), Springer, 15-23.

Janin J. Assessing predictions of protein-protein interaction: the CAPRI experiment. Protein Sci. (2005) 14(2):278-83.

Rodier F, Bahadur RP, Chakrabarti P, Janin J Hydration of protein-protein interfaces. Proteins (2005) 60(1):36-45.

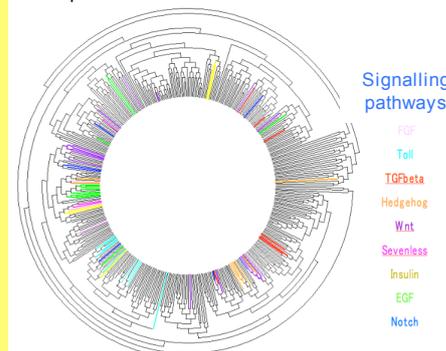
Charon I, Denœud L, Guénoche A, Hudry O. Comparing partitions by element transfers soumis pour publication au Journal of Classification

RESULTATS (2)

- développement de Prodistin Web Site, pour l'analyse de réseaux en ligne.
- étude de l'évolution de la fonction de protéines, produits de gènes paralogues issus d'une duplication ancestrale du génome, via Prodistin.
- investigation de voies de signalisation développementales de drosophile.
- construction progressive de classes de sommets de graphe : calcul d'une valeur de densité, fonction du nombre d'arêtes dans le voisinage d'un sommet.
- calcul d'une distance entre deux partitions constitués par des sommets de classes denses.
- comparaison des différents indices de distance trouvés dans la littérature. Evaluation de leur qualité et comparaison de leurs comportements.

ILLUSTRATION (2)

Utilisation de la méthode Prodistin de classification fonctionnelle pour l'étude de voies de signalisation chez la drosophile



A partir de 2500 interactions (IPPs), un arbre de classification a été obtenu, contenant 480 protéines. Le mapping de protéines appartenant à 9 voies de signalisation a été réalisé, montrant la proximité fonctionnelle de certaines voies. Par ailleurs, aucune voie ne se regroupe en une seule classe fonctionnelle, mais plutôt en sous-classes.