

# Evolution in silico

(Projet de Recherche 2003)

## Evolution des protéines : simulation et contrôle

<http://biology.polytechnique.fr/biocomputing/impbio.html>



### CONTEXTE

L'évolution dirigée in silico, a permis le développement de nouveaux ligands et catalyseurs. En parallèle, le séquençage de centaines de génomes permet une vision moléculaire de l'évolution naturelle. Il devient alors pertinent de simuler des aspects de l'évolution qui ne sont pas directement accessibles par expérience. Suivant les travaux pionniers de Eigen, nous simulons l'évolution de protéines individuelles et d'ensembles de protéines en interaction. Nous utilisons également une évolution dirigée in silico pour développer de nouveaux enzymes.

### OBJECTIFS

Nous simulons l'évolution 'neutre' de protéines individuelles et d'ensembles de protéines en interaction. L'évolution neutre admet comme seule pression sélective le maintien de la structure native; sa dynamique prend la forme d'une marche au hasard sur un 'réseau neutre' de séquences viables. Nous analyserons la topologie et la complexité de ces réseaux; leurs échelles de temps importantes; leur robustesse par rapport à une augmentation soudaine de la taille du code génétique; le rôle des interactions protéine-protéine dans leur évolution. En parallèle, nous appliquerons l'évolution dirigée à la ré-ingénierie de la spécificité d'aminocyl-ARNt synthétases. Ces enzymes très anciens attachent un acide aminé spécifique sur un ARNt spécifique, établissant l'intégrité du code génétique. Cet exercice a donc un lien étroit avec l'évolution du code génétique. Les résultats sont vérifiés expérimentalement par les biochimistes du laboratoire. Ils seront mis en parallèle avec les résultats d'évolution simulée avec des codes génétiques de complexités variables.

### PROJET

Développer des codes pour la simulation d'évolution neutre de protéines et pour l'ingénierie de protéines par évolution dirigée. Les appliquer pour explorer les mécanismes de l'évolution moléculaire et pour obtenir de nouveaux enzymes.

### RESULTATS PRINCIPAUX

- Etude exhaustive de l'évolution de protéines modèles sur réseau;
- Etude exhaustive de l'évolution d'ensembles de protéines en interaction (modèles sur réseau) avec une pression sélective liée à l'association protéine-protéine;
- Simulations d'évolution de 3 protéines avec un modèle 3D hors réseau;
- Développement de fonctions d'énergie empiriques pour l'association protéine-protéine avec 6 codes génétiques différents de complexités croissantes [4];
- Développement d'une suite de codes pour l'évolution dirigée [5];
- Invention d'un modèle de solvant décomposable en paires pour l'évolution dirigée [1]
- Portage de tous les codes dans un environnement de calcul distribué
- Simulations détaillées de l'aspartyl-ARNt synthétase révèlent les bases moléculaires de la spécificité [2,3].

### CALENDRIER

2004 Septembre, démarrage du projet  
fin 2005 plateforme de calcul distribué déployée à l'Ecole Polytechnique  
Premiers articles soumis (PNAS, Biochemistry, J Phys Chem) ou en rédaction.

### LABORATOIRES IMPLIQUES

- Laboratoire de Biochimie, Ecole Polytechnique, 91128 Palaiseau
- Collaborations:
- Dépt de Physique, U. de Chypre
- Toronto Hospital, Canada

### MOTS CLES

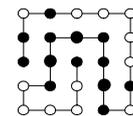
Modélisation moléculaire, évolution dirigée, algorithmes génétiques, calcul distribué, aminocyl-ARNt synthétases

### PUBLICATIONS PRINCIPALES

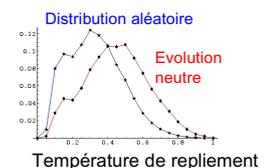
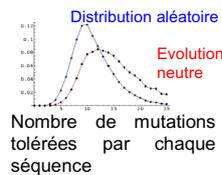
- [1] G Archontis, T Simonson, soumis, J Phys Chem B. A residue-pairwise Generalized Born solvent model suitable for protein design.
- [2] D Thompson, P Plateau, T Simonson, soumis, Proc Nat Acad Sci (USA). Free energy simulations reveal long-range electrostatic interactions and substrate-assisted specificity in an aminocyl-tRNA synthetase.
- [3] D Thompson, T Simonson, soumis, Biochemistry. Molecular dynamics simulations show that ATP and Mg<sup>2+</sup> contribute to substrate specificity in aspartyl-tRNA synthetase through long-range electrostatic interactions.
- [4] G Launay, R Mendez, S Wodak, T Simonson, en préparation. Simplified energy functions for protein fold recognition and identification of protein-protein complexes.
- [5] A Lopes, A Jaramillo, G Archontis, T Simonson, en préparation. Algorithms and solvent models for protein sidechain reconstruction calculations.

### Evolution neutre de protéines sur réseau

1081 conformations  
33.554.432 séquences  
**6.290.817 séquences viables**



Une protéine modèle: une chaîne sur réseau 2D



Il existe un unique état stationnaire stable pour l'évolution neutre par mutations ponctuelles. Dans cet état (courbes rouge ci-dessus), la population est fortement enrichie en séquences résistantes aux mutations (fortement "modifiables") et à la dénaturation thermique. On retrouve ce résultat avec un modèle 3D plus réaliste.

**Les Polytechniciens font de la Bioinformatique:**  
Calcul distribué sur les PC de volontaires à l'Ecole Polytechnique

