

## VI. **Simulation des mécanismes d'activation des canaux ioniques de potassium par la modélisation moléculaire**

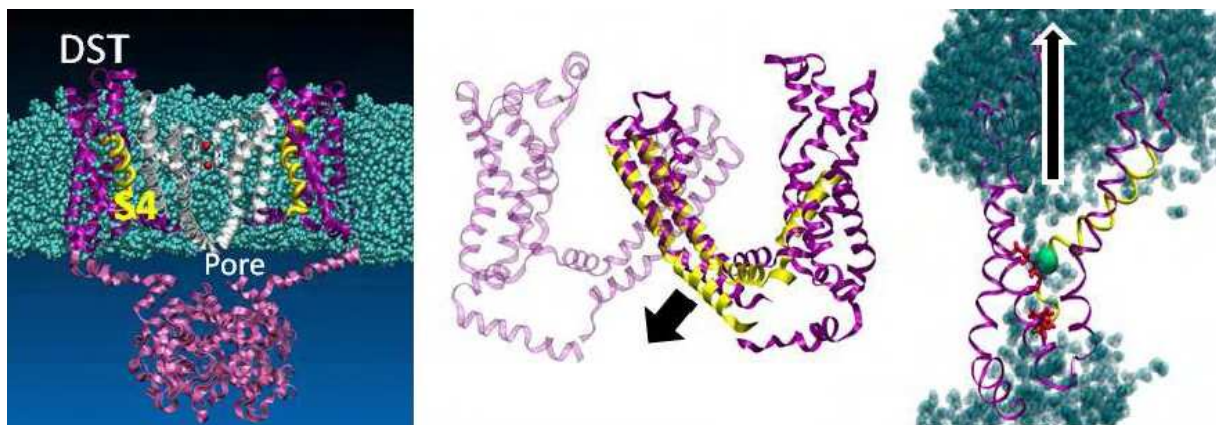
**Auteur :** Mounir Tarek (Equipe de Dynamiques des Assemblages Membranaires, UMR 7565 CNRS - Nancy Université)

**Informations :** <http://www.edam.uhp-nancy.fr/>

Notre groupe de recherche étudie par modélisation moléculaire plusieurs protéines qui jouent un rôle clef dans le fonctionnement de membranes biologiques. Parmi elles, les canaux ioniques sont des micro-générateurs de signaux électriques présents dans toutes les cellules et plus particulièrement dans les cellules neuronales, cardiaques ou musculaires. Les travaux menés visent à étudier l'activation de ces canaux afin de mieux caractériser leur fonction et de comprendre les mécanismes impliqués dans leur modulation ou dysfonctionnement.

Jusqu'à présent, les moyens de calcul nationaux ne permettaient pas d'appréhender ce type de simulation avec des temps de restitution raisonnables.

Les travaux menés cet été sur le supercalculateur JADE a permis, considérant un système de 300 000 atomes, de simuler la réponse d'un canal potassique mammifère soumis à des conditions physiologiques d'activation sur une échelle de temps inédite, dépassant la microseconde. Des détails moléculaires impliqués dans le fonctionnement du canal, jusque là inaccessibles expérimentalement, ont ainsi pu être révélés.



*Simulations de dynamique moléculaires du canal ionique sensible à la tension Kv1.2. A droite, représentation du système total (~300 000 atomes) comprenant le canal avec son domaine intracellulaire (rose), la membrane lipidique (bleu), et la solution intra et extra cellulaire (non représentée ici). Au centre, changements conformationnels observés après 800 nanosecondes de simulation, A gauche, les résultats correspondant à la conséquence de mutations. Les calculs montrent que ces mutations, incriminées dans des pathologies génétiques.*

**Ce grand défi nous aura aussi permis d'étudier par comparaison, l'effet de certaines mutations génétiques identifiées comme responsables d'épilepsies ou de paralysies, et déterminé les mécanismes moléculaires associés à ces pathologies lourdes.** Ces résultats originaux, mis en évidence *in-silico*, ouvrent de nouveaux champs d'investigation et nous permettent aujourd'hui de proposer des expériences complémentaires ciblées dans le domaine de l'excitabilité cellulaire.



Nancy-Université  
Université  
Henri Poincaré

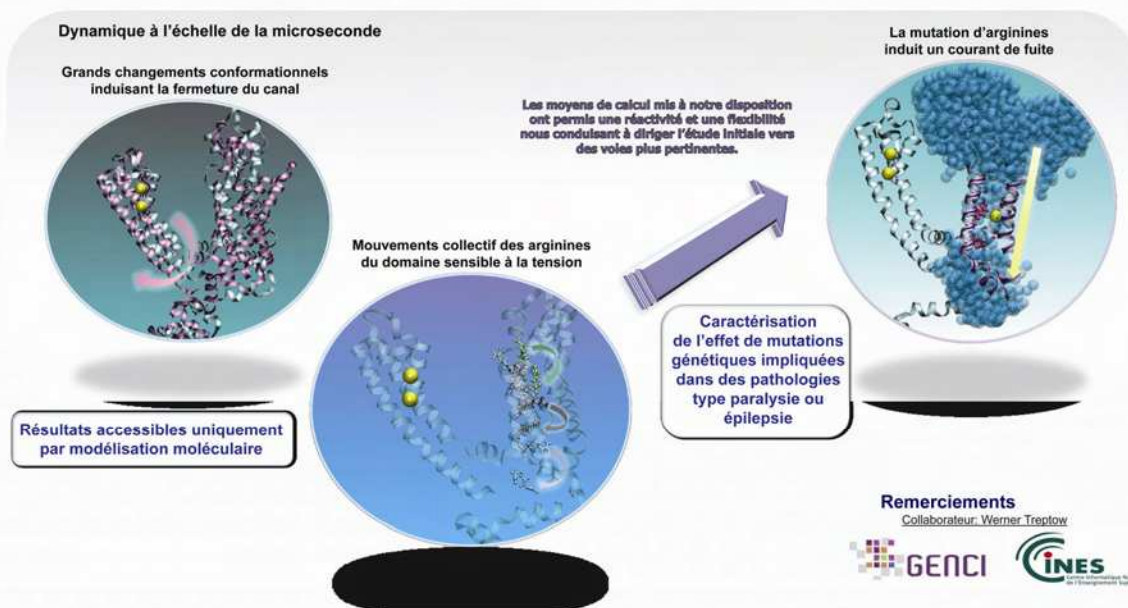
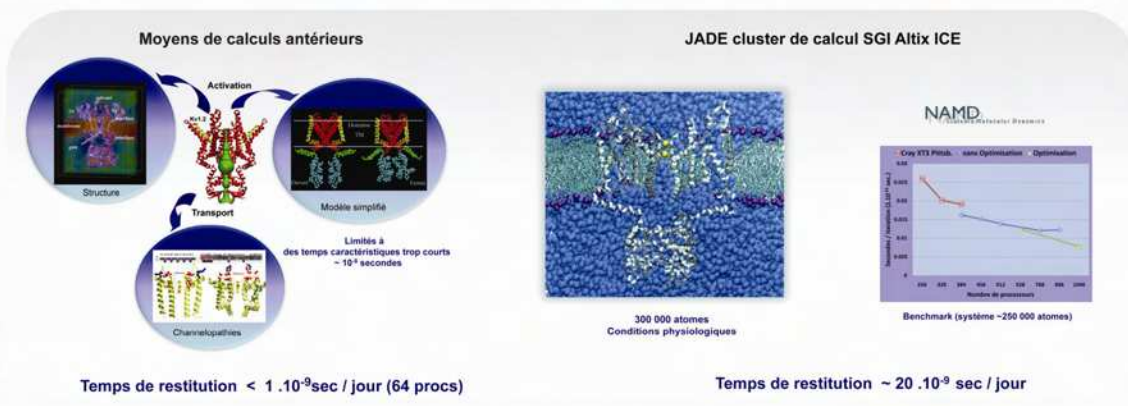
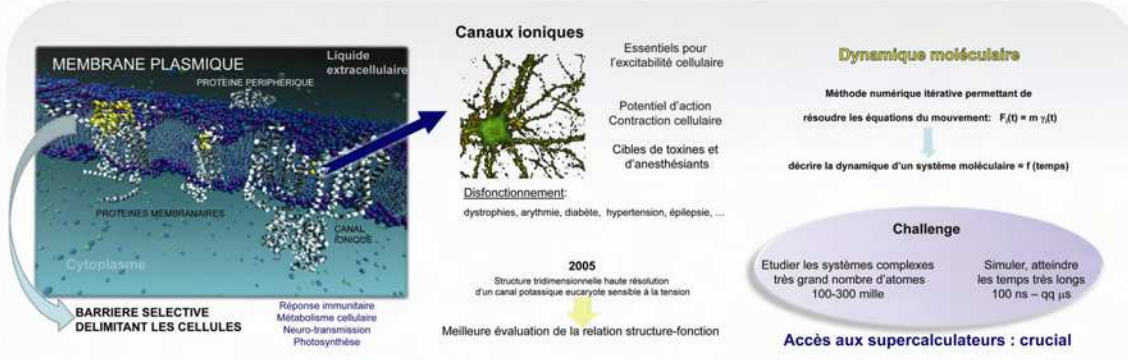
# Activation des canaux ioniques de potassium par modélisation moléculaire

Mounir Tarek

Equipe de Dynamique des Assemblages Membranaires - Unité Mixte de Recherches 7563 CNRS - Nancy Université

SYSMC  
UMR CNRS/UHP 7563

eDAM  
Equipe de Dynamique des  
Assemblages Membranaires



## VII. Prédiction de la structure des protéines

**Auteur :** Thomas Simonson (Département de Biologie, Ecole Polytechnique)

**Informations :** Proteins@Home <http://biology.polytechnique.fr/proteinsathome/>

Les projets internationaux de séquençage ont fait exploser l'information génomique disponible, changeant la façon de penser la biologie. Des centaines de génomes sont aujourd'hui complètement séquencés, dont le génome humain ( $3.10^9$  nucléotides), achevé en 2004.

En revanche, le nombre de structures 3D protéiques connues augmente beaucoup moins vite.

Pour contribuer à cette modélisation, nous étudions les structures de protéines existantes, et nous essayons de prédire quelles séquences d'acides aminés sont adaptées à ces structures: ainsi, nous créons des protéines homologues « virtuelles » *in silico*. La prédiction de structures à grande échelle que nous développons dans la cadre du projet ANR Proteus permettra d'avancer dans la génomique structurale, dans l'attribution des fonctions protéiques et l'annotation des génomes, et dans l'ingénierie de protéines avec des fonctions nouvelles.

Pour explorer l'espace des séquences adaptées à une structure 3D donnée, nous avons implémentée et testée une méthode dite « d'évolution dirigée », qui génère des séquences aléatoires puis teste leur adéquation, mimant un processus d'évolution naturelle.

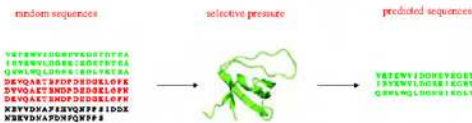
Les calculs se déroulent en deux phases: une première phase qui nécessite peu de mémoire et peu de communication entre processeurs; une seconde phase qui en nécessite davantage et qui rend l'utilisation de supercalculateurs nécessaire.

Nous avons porté la première phase sur une plateforme de calcul distribué, qui permet de déployer les calculs chez des internautes volontaires et de traiter rapidement de gros jeux de données.

Un ensemble de calculs de la seconde phase a été effectué au CINES pendant cette période de test de la machine JADE. Ces calculs correspondent à un ensemble de protéines-test, qui sert de benchmark pour notre méthode de modélisation et de prédiction.

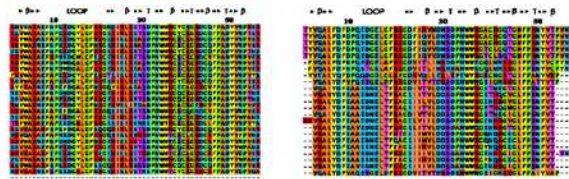
**GOAL OF THE PROJECT**

Protein structure prediction remains a major challenge. In the 80's, the inverse folding problem was formulated: Instead of predicting the 3D structure from the amino acid sequence, one considers a given backbone structure and predicts the amino acid sequences that fold into it. Computational protein design addresses this problem. The computed sequences can then be compared to natural sequences whose 3D structure is unknown, and help to characterize them.



**MAIN RESULTS**

We generated ~450,000 sequences for 25 SH3 domains. On average, the low energy sequences had a 33% identity with the native sequence; the highest identities were ~46%. The figure below shows high-identity designed sequences (left) for one SH3 protein, Gtb2, and compares them to experimental sequences (right).

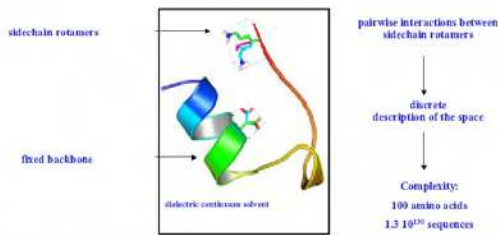


**THEORETICAL BACKGROUND**

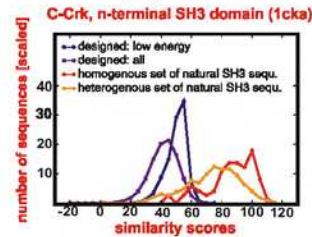
We consider a particular, experimental, protein backbone structure. We select amino acid sequences and sidechain conformations that stabilize this structure. Stability is characterized by the protein folding free energy:

$$\Delta G_{\text{folding}} = \Delta G_{\text{fold}} - \Delta G_{\text{unfolded}}$$

We thus require a model of both the folded and unfolded protein structures. In the unfolded state model, each amino acid sidechain interacts only with solvent and nearby backbone. In the folded state model, the backbone is fixed and sidechains occupy standard conformations, called rotamers.



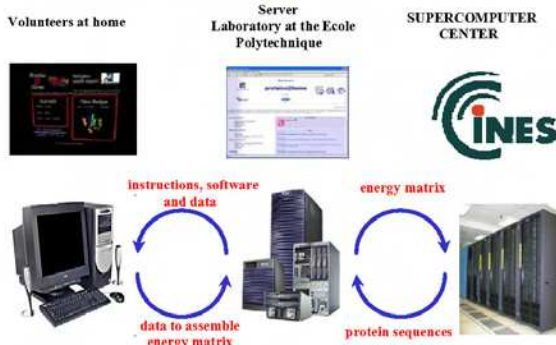
To evaluate the native-like character of the computed sequences, the Blossum62 similarity scores of natural and designed sequences were computed for selected SH3 proteins (figure below, C-Crk). Similarity scores were computed for the entire stock of *in silico* sequences (designed: all) and 10,000 low-energy sequences (designed: low energy). We compared with two sets of natural SH3 sequences. A more homogeneous set contained 73 and a more diverse set over 3,000 SH3 sequences (see below). The overall performance is comparable (not shown) to the best competing methods.



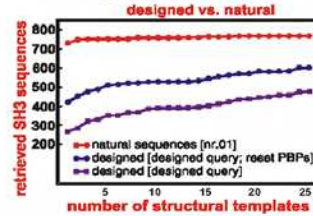
Fold-recognition using the computed sequences. We used the computed sequences to construct position-specific scoring matrices for sequence homologous searching with PSI-Blast. We compared the performance of PSSMs evolved from designed and natural sequences (see below). Retrieved sequences were examined using the Conserved Domain Database (CCD), the protein classification component of NCBI (data not shown). Preliminary results suggest that the computed sequences may have the potential to supplement existing sequence homologous searching tools.

**HIGH PERFORMANCE COMPUTING**

Computations of the energy matrices were done using large-scale volunteer computing, through our Proteins@Home project. Over 20,000 volunteers have participated, from over 100 countries. For a list of participants, see <http://biology.polytechnique.fr/proteinsathome>. Each volunteer donates time on her/his computer, by running our screensaver, shown above (left). This massive, international, volunteer resource was established with the help of the BOINC software: the Berkeley Open Infrastructure for Network Computing. The energy matrices were then translated into protein sequences via the CINES SUPERCOMPUTER JADE (10,600 processor intel(R), Xeon 54.72 a 3.06 Hz).



**Searching protein databases for SH3 proteins**



**PUBLICATIONS**

- 1.) Lopes, A. Aleksandrov, C. Bathelt, G. Archonta and T. Simonson. **Computational sidechain placement and protein mutagenesis with implicit solvent models.** *Proteins*, 67: 853-867, 2007.
- 2.) T. Simonson, D. Mignon, M. Schmidt am Busch, A. Lopes and C. Bathelt. **The inverse folding problem: structure prediction in the genomic era.** In *Distributed & Grid Computing - Science Made Transparent for Everyone. Principles, Applications and Supporting Communities*, page 000, Teknum Publishers, Berlin, 2008.
- 3.) M. Schmidt am Busch, A. Lopes, D. Mignon and T. Simonson. **Computational protein design: software implementation and parameter optimization.** *J Comp Chem*, 2008, 29 (7), 1092-1102.
- 4.) M. Schmidt am Busch, A. Lopes, N. Amara, C. Bathelt and T. Simonson. **Testing the CaChamb/Accessible Surface Area solvent model for protein stability, ligand binding, and protein design.** *BMC Bioinformatics* 2008, 9:148 (15 March 2008)
- 5.) M. Schmidt am Busch, D. Mignon and T. Simonson. **Computational protein design: a systematic study on the sequences space within the SH3 protein family**, submitted