

DESCRIPTION DU PROJET DE RECHERCHE

Titre : Robustesse des réseaux biologiques

Résumé Nous avons récemment développé une théorie mathématique de la robustesse des réseaux biologiques [1,8]. Selon cette théorie, la robustesse paramétrique des grands réseaux biologiques est la conséquence d'un phénomène de concentration de mesure dans des espaces de haute dimension [1,8]. La concentration résulte par projection de l'espace de phases en haute dimension et une variété inertielle de petite dimension. Cette projection permet en outre la réduction de modèle [1-10]. Nous mettons en évidence trois types de robustesse et nous concevons des testes statistiques pour leur identification [1,8]. Dans cette étude nous proposons la vérification expérimentale de la généralité de la robustesse par concentration, en appliquant nos testes de manière extensive à des modèles des différentes bases de données.

Liste de publications

- [1] Gorban AN, Radulescu O (2007). Dynamical robustness of biological networks with hierarchical distribution of time scales. *IET Systems Biology* **1**: 238–46.
- [2] O. Gorban A, Zinovyev A, Lilienbaum A (2008). Robust simplifications of multiscale biochemical networks. *BMC Systems Biology* **2**: 86.
- [3] Crudu A, Debussche A, Radulescu O (2009) Hybrid stochastic simplifications for multiscale gene networks. *BMC Systems Biology* **3**: 89.
- [4] Vakulenko S, Manu, Reinitz J, Radulescu O (2009). Size regulation in the segmentation of Drosophila: interacting interfaces between localized domains of gene expression ensure robust spatial patterning. *Physical Review Letters* **103**: 168102.
- [5] Manu, Surkova S, Spirov AV, Gursky VV, Janssens H, Kim AR, Radulescu O, Vanario-Alonso CE, Sharp DH, Samsonova M, Reinitz J (2009). Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. *PLoS Biology* **7**: e1000049.
- [6] Manu, Surkova S, Spirov AV, Gursky VV, Janssens H, Kim AR, Radulescu O, Vanario-Alonso CE, Sharp DH, Samsonova M, Reinitz J (2009). Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Comput Biology* **5**: e1000303.
- [7] Gorban AN, Radulescu O, Zinovyev A (2010). Asymptotology of chemical reaction networks. *Chemical Engineering Science* **65**: 2310-24.
- [8] O.Radulescu, A.N.Gorban, A.Zinovyev (2010). Parametric robustness in gene networks: reliable functioning with unreliable components. JOBIM, Montpellier 2010.
- [9] S.Vakulenko, O.Radulescu (2011, submitted). Flexible and robust patterning by centralized gene networks. Submitted to the special number "Watching the Daisies Grow: from biology to biomathematics and bioinformatics" of Fundamenta Informaticae.
- [10] V.Noel, D.Grigoriev, S.Vakulenko, O.Radulescu (2011). Tropical geometries and dynamics of biochemical networks. Applications to hybrid cell cycle models. SASB, Venise 2011.
- [11] von Neumann J. (1963). Probabilistic logics and the synthesis of reliable organisms from unreliable components. In: J. von Neumann Collected works vol.5. Oxford: Pergamon Press; 1963.
- [12] von Dassow G, Meir E, Munro EM, Odell GM (2000). The Segment Polarity Network is a Robust Developmental Module. *Nature*. 2000;406:188–192.
- [13] Gromov M. (1999). Metric structures for Riemannian and non-Riemannian spaces, Progr.Math. 152. Boston: Birkhauser; 1999.
- [14] Janssens H, Hou S, Jaeger J, Kim A-R, Myasnikova E, Sharp D, Reinitz J. Quantitative and predictive model of transcriptional control of the Drosophila melanogaster even skipped gene. *Nature genetics* 2006;38(10):1159-65.
- [15] Rizk A, Batt G, Fages F, Soliman S (2008). On a continuous degree of satisfaction of temporal logic specifications and applications to systems biology. *Lecture Notes in Computer Science* **5307**: 251-268.
- [16] Rizk A, Batt G, Fages F, Soliman S (2009). A general computational method for robustness analysis with applications to synthetic gene networks. *Bioinformatics* **25**: i169-78.
- [17] Kitano H (2007). Towards a theory of biological robustness. *Molecular Systems Biology* **3**: 137.
- [18] Chaves M, Albert R (2008). Studying the effect of cell division on expression patterns of the segment polarity genes. *Journal of the Royal Society - Interface* **5**: S71-84.

Présentation générale

La robustesse, définie comme la capacité d'un système de fonctionner de façon fiable avec des composants incertains ou s'adapter aux changements des conditions externes, représente une caractéristique commune des systèmes vivants. Les organismes les plus adaptés sont ceux qui résistent aux maladies, aux imperfections ou les défaillances des mécanismes régulateurs et qui peuvent assurer leur homéostasie dans des conditions diverses. Comprendre l'origine de la robustesse des réseaux biologiques est un des objectifs majeurs de la « biologie des systèmes », qui est une nouvelle approche de la physiologie des cellules et des organismes biologiques par des modèles dynamiques de réseaux moléculaires [17].

Nous étudions la robustesse des modèles de réseaux par rapport à des changements aléatoires des paramètres. Les premiers idées sur la robustesse peuvent être trouvées dans la discussion de von Neumann des réseaux d'automates [11]. Von Neumann remarque la relation intrinsèque entre l'aléatoire et la robustesse. En le citant "sans aléatoire, des situations peuvent surgir où les erreurs ont tendance à être amplifiées au lieu d'être compensées ; par exemple il est possible que la machine se rappelle ses erreurs et les perpétue ensuite". Nous mettons en évidence des mécanismes génériques de compensations d'erreur dans des grands réseaux de régulation moléculaires [1-10]. Nous distinguons entre trois types de robustesse paramétrique générique : concentration cube, simplex ainsi que systèmes robustes/fragiles [1,8]. Les deux premiers types sont reliés à des phénomènes de concentration de mesure dans des espaces à haute dimension [13]. Le dernier est caractéristique des systèmes dissipatifs à plusieurs échelles de temps, dont la dynamique possède des variétés inertielles de petite dimension. Ces systèmes sont contrôlables par un petit nombre de paramètres critiques, mais restent insensibles aux changements d'un grand nombre de paramètres non-critiques.

Nous avons proposé un scénario simple permettant de tester, pour un modèle donné, le type de robustesse [1]. Dans ce scénario, la variabilité d'une propriété du modèle (quantifiée par sa log-variance) est calculée suite à des variations aléatoires indépendantes des paramètres du modèle dans deux cas : i) tous les paramètres sont changés de manière indépendante avec une log-variance individuelle croissante ; ii) un sous-ensemble de cardinal r (le nombre de cibles) des paramètres est changé à log-variance fixée, mais pour des valeurs croissantes de r . Les deux graphes résultants, la log-variance de la propriété quantitative en fonction de la log-variance des paramètres et en fonction du nombre de cibles sont discriminants par rapport aux trois types de robustesse.

Nous souhaitons appliquer cette procédure de manière extensive aux modèles de la base de données BIOMODELS www.ebi.ac.uk/biomodels/, à une collection de circuits bistables disponible via une collaboration avec U.Bhalla de NCBS Bangalore, ainsi qu'à des modèles de régulation transcriptionnelle de gènes du développement chez la *Drosophila* en collaboration avec J.Reinitz de U.of Chicago [14].

Des efforts similaires en France et à l'international proposent des études de la robustesse en testant la sensibilité ou la stabilité structurelle des modèles, mais sans une idée théorique unificatrice, telle que la nôtre. Ainsi, ces études sont spécifiques à des modèles individuelles et mettent rarement en évidence des principes généraux [15-18]. L'ambition de notre projet est de proposer un nouveau paradigme dont l'impact en biologie théorique et biologie des systèmes pourrait être considérable.

Comprendre les mécanismes générales de la robustesse permet de guider des études expérimentales ou des applications dans le domaine de la santé. Ainsi, il est important de savoir quels modèles et quelles

propriétés sont susceptibles de résister à des attaques localisés (c'est le cas de la robustesse du type cube ou simplex) et quels modèles ont des endroits critiques. Certains des modèles que nous allons tester concernent la relation entre signalisation et prolifération, importante pour la compréhension des mécanismes moléculaires du cancer.

Méthodes, Algorithme, Langages

Nous utilisons un algorithme de Monte-Carlo pour estimer la log-variance qui consiste en plusieurs exécutions indépendantes d'une procédure principale déterminant une propriété quantitative du modèle.

Les modèles, sous format SBML sont importés de la base BIOMODELS et traités dans un environnement MATLAB. Nous avons développé des fonctions MATLAB permettant la lecture, la conversion en systèmes d'équations différentielles des modèles SBML et la génération automatique du code nécessaire à l'application de nos tests. Le code est ensuite compilé par un compilateur MATLAB et peut être utilisé sur une machine ne possédant pas de licence MATLAB.

La procédure principale du programme consiste à générer les solutions numériques des équations différentielles et les utiliser pour calculer des caractéristiques quantitatives des modèles dynamiques. Par exemple, pour des oscillateurs nous utilisons trois méthodes pour tester l'existence des oscillations, et pour déterminer leur période et leur degré d'amortissement. La première méthode utilise la transformée de Fourier rapide, la deuxième détermine les positions des extrema locaux successifs des trajectoires et la troisième utilise le spectre de la matrice Jacobienne aux états stationnaires du système. Dans le cas des modèles multistationnaires nous utilisons les positions des états stationnaires (obtenues par des méthodes d'homotopie) et le spectre de la matrice Jacobienne dans les états stationnaires pour caractériser la robustesse de la multistationnarité. Dans le cas des modèles de transcription, nous utilisons un modèle du type champs moyen proposé par l'équipe de J.Reinitz. Nous souhaitons aussi prendre en compte les fluctuations stochastiques de ce modèle, ce qui peut également se faire à l'aide de simulations Monte-Carlo.

Nos algorithmes sont trivialement parallélisables, il suffit de disposer d'un système de fichiers communs et lancer les processus plusieurs fois avec des paramètres différents. Le nombre d'exécutions peut être reparté de manière quelconque sur les processeurs qui seront mis à notre disposition. Les valeurs résultantes seront enregistrées dans un ou plusieurs fichiers communs qu'on traitera ensuite par nos propres moyens. On peut estimer que pour un modèle, nous avons besoin de 10.000 à 1.000.000 d'exécutions scalaires de la procédure principale pour une bonne statistique et que nous souhaitons tester moins d'une centaine de modèles. Les modèles d'oscillateurs sont les plus gourmands en temps de calcul car il faut générer des longues trajectoires. Le temps de calcul par processeur scalaire varie entre quelques secondes et une heure, avec une distribution qui peut être considérée comme uniforme en échelle logarithmique. Globalement nous avons besoin d'environ 150000 h de calcul scalaire pour un seul processeur, ainsi que d'un espace de stockage de 30-50Go.

Au préalable, nous comptons tester nos codes (avant la fin de cette année) sur les machines du centre HPC@LR de l'Université de Montpellier 2.