

DYNAMICS OF PROTEIN INTERACTION SUB-NETWORKS

Barbara Bravi

Department of Mathematics, King's College London

Supervisor: Prof. Peter Sollich

UNIVERSITY OF BOLOGNA, ITALY

Faculty of Mathematical, Physical and Natural Sciences

- 1 September 2007 - October 2010:
Bachelor of science in Physics
- 2 October 2010 - July 2012:
Master of Science in Physics
Theoretical Physics curriculum
- 3 October 2010 - July 2012:
**Collegio Superiore of University
of Bologna**



Scientific experience

① July-September 2011:

CERN Summer Student Programme

- Project content: Monte Carlo event generators
- Application of Markov chains in hadronization events
- Supervisor: Dr. Peter Zeiler Skands

② February-June 2012:

Exchange programme ENS Paris

- Master thesis: "STATISTICAL MODELS FOR COEVOLUTION IN PROTEIN FAMILY PROFILES"
- Hidden Markov Models (HMMs) for protein profiles + Theoretical framework to incorporate interactions between residues + Mean-field approximation
- Supervisor: Prof. Remi Monasson

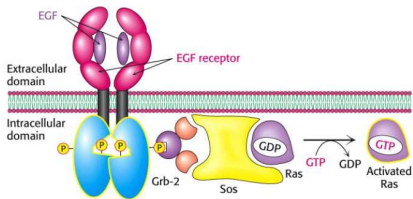
Biochemical reaction networks in systems biology

Intracellular Signalling pathways



Epidermal Growth Factor Receptor (EGFR)

- Phosphorylation
≡ Biochemical message
- Protein-protein interactions
≡ Signal transduction
- Protein dynamics
vs Gene regulation



Mean-field approximation: Langevin equation

$$\begin{aligned} \frac{\partial x_i(t)}{\partial t} = & \sum_{j,l,j \neq l} \left(k_{l,ij}^- x_l(t) - k_{ij,l}^+ x_i(t) x_j(t) \right) + \frac{1}{2} \sum_{j,l,j \neq l} \left(k_{jl,i}^+ x_j(t) x_l(t) - k_{i,jl}^- x_i(t) \right) \\ & + \sum_j \left(\lambda_{ji} x_j(t) - \lambda_{ij} x_i(t) \right) + \sum_l \left(2k_{l,ii}^- x_l(t) - k_{ii,l}^+ x_i(t) x_l(t) \right) + \\ & + \sum_j \left(\frac{1}{2} k_{jj,i}^+ x_j(t) x_j(t) - k_{i,jj}^- x_i(t) \right) + \xi_i(t) = \Phi_i(\mathbf{x}(t)) + \xi_i(t) \end{aligned}$$

Mass-action kinetics for elementary proteomic processes:

- Complex formation $i + j \xrightarrow{k_{ij,l}^+} l$
- Complex dissociation $l \xrightarrow{k_{l,ij}^-} i + j$
- Conformational change $i \xrightarrow{\lambda_{ij}} j$

STOCHASTICITY: **intrinsic noise** $\xi_i(t)$

→ Copy number fluctuations

Fundamental and practical limitations:

- Non-linear equations
 - Vast amount of information ($\sim 10^4$ equations)
- ↓
- **Uncertainty** → No complete qualitative understanding

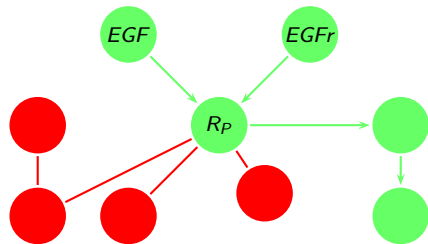
⇒ STATISTICAL APPROACH
& MODEL REDUCTION

Small subset of variables:

Sub-network

Embedded in a larger network:

Bulk



- 1 **DESCRIPTION**: dynamics of sub-network
- 2 **INFERENCE**: sub-network → bulk

Projection Methods

Projection operator $\mathcal{P} \Rightarrow$ Restriction to sub-network

$x_i(t) \rightarrow$ protein concentration from the sub-network

$$\frac{\partial x_i(t)}{\partial t} = x_j(t) \underbrace{\Omega_{ji}(t)}_{\text{Rate matrix}} \quad [\text{SUB-NETWORK REACTIONS}]$$

$$+ \int_0^t dt' x_j(t') \underbrace{M_{ji}(t-t')}_{\text{Memory function}} + \underbrace{r_i(t)}_{\text{Random force}} \quad [\text{BULK}]$$



ALTERNATIVE APPROXIMATION: VARIATIONAL
exhaustive and intuitive sub-network description?

Variational Gaussian approximation

Generating functional- Discretization time Δ :

$$\begin{aligned} Z_{\xi}[\psi] &= \int \prod_{it} dx_i(t) e^{i\Delta\psi_i(t)x_i(t)} \delta(x_i(t+\Delta) - x_i(t) - \Delta[\Phi_i(\mathbf{x}(t)) + \xi_i(t)]) = \\ &= \int \prod_{it} \frac{dx_i(t)d\hat{x}_i(t)}{2\pi} e^{i\Delta\psi_i(t)x_i(t)} \cdot e^{i\hat{x}_i(t)[x_i(t+\Delta) - x_i(t) - \Delta[\Phi_i(\mathbf{x}(t)) + \xi_i(t)]]} \end{aligned}$$

Average over noise:

$$Z[\psi] = \langle Z_{\xi}[\psi] \rangle_{\xi}$$

Gaussian Multiplicative White noise

$$\langle \xi_i(t) \rangle = 0 \quad \langle \xi_i(t) \xi_j(t') \rangle = \frac{1}{\Delta} \Sigma_{ij}(\mathbf{x}, t, t') = \frac{1}{\Delta} \Sigma_{ij}(\mathbf{x}) \delta_{tt'}$$

$$Z[\psi]|_{\psi=0} = \int d\mathbf{x}d\hat{\mathbf{x}} \cdot e^{iS}$$

⇓

“Distribution” (complex measure) over concentration paths

$$P(\mathbf{x}, \hat{\mathbf{x}}) = \frac{e^{iS(\mathbf{x}, \hat{\mathbf{x}})}}{Z} P_0(\mathbf{x}, \hat{\mathbf{x}}) \approx \mathcal{N}((\mathbf{x}, \hat{\mathbf{x}})|\boldsymbol{\mu}, \mathbf{C}) = Q(\mathbf{x}, \hat{\mathbf{x}})$$

Kullback-Leibler (KL) Divergence:

$$\text{KL}(Q||P) = \int d\mathbf{x}d\hat{\mathbf{x}} \cdot Q(\mathbf{x}, \hat{\mathbf{x}}) \ln \frac{Q(\mathbf{x}, \hat{\mathbf{x}})}{P(\mathbf{x}, \hat{\mathbf{x}})}$$

Measure of “distance” → Minimization

$$\frac{\partial \text{KL}}{\partial \boldsymbol{\mu}} = 0 \quad \frac{\partial \text{KL}}{\partial \mathbf{C}} = 0$$

Derivatives and $\Delta \rightarrow 0$

- $\hat{\mu}(t) \equiv 0$, $C_{ij}(t', t) = \langle \hat{x}_i(t') \hat{x}_j(t) \rangle \equiv 0$
- $\mu(t)$: Eqs. motion + Equal time correlations
- Correlation:

$$\frac{\partial \mathbf{C}(t', t)}{\partial t} = i\mathbf{C}(t', \hat{t})\Sigma(t) - \mathbf{C}(t', t)\mathbf{K}^T(t)$$

Propagation term $\approx \mathbf{K} \leftarrow$ Rate constants and $\mu(t)$

- Response function:

$$\mathbf{C}(t', \hat{t}) = -i \frac{\partial \langle \mathbf{x}(t') \rangle}{\partial \mathbf{I}(t)} = -i \theta(t' - t) \cdot \mathbf{T} \left[e^{-\int_t^{t'} \mathbf{K}(t'') dt''} \right]$$



External field



Implementation of causality

SUB-NETWORK/BULK Simplifications

- Block structure:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}^{SS} & \mathbf{C}^{SB} \\ \mathbf{C}^{BS} & \mathbf{C}^{BB} \end{pmatrix}$$

$$\left\{ \begin{array}{l} \text{Decoupling } \mathbf{C}^{BS} = \mathbf{C}^{SB} \equiv 0 \\ \quad \quad \quad + \quad \quad \quad \rightarrow \text{Closed equations} \\ \text{Bulk: Steady-state} \end{array} \right.$$

- Time evolution of \mathbf{C} has **source term** $\approx \Sigma \approx \epsilon = \frac{1}{V}$

$$\Sigma = \frac{\epsilon}{2} \mathbf{S} \text{diag}(\mathbf{f}) \mathbf{S}^T \quad \text{Flux Fluctuations}$$



Large $V \rightarrow$ Deterministic description: **Memory/Feedback?**

VARIATIONAL FACTORIZED APPROXIMATION

$$P(\mathbf{x}_S, \mathbf{x}_B) \approx Q_S(\mathbf{x}_S)Q_B(\mathbf{x}_B)$$

$$\ln Q_S^*(\mathbf{x}_S) = \langle \ln P(\mathbf{x}_S, \mathbf{x}_B) \rangle_{Q_B(\mathbf{x}_B)} + \text{const.}$$

$$\ln Q_B^*(\mathbf{x}_B) = \langle \ln P(\mathbf{x}_S, \mathbf{x}_B) \rangle_{Q_S(\mathbf{x}_S)} + \text{const.}$$

Free-form decoupling $\rightarrow \mu$: original dynamics of \mathbf{x}
+ coupling to bulk



VARIATIONAL APPROACH:

No intermediate approximation $\left\{ \begin{array}{l} \text{COMPLEX} \rightarrow \text{Fully correlated} \\ \text{SIMPLIFIED} \rightarrow \text{No feedback} \end{array} \right.$

Model reduction vs Relevant information

Future Developments

Approximation strategies

- 1 Marginalization via Diagrammatic expansion
- 2 Reverse form $KL(P||Q)$: **Expectation-Propagation** Network \rightarrow Graph: temporal trajectory at each node
- 3 Coloured Intrinsic Noise

Long-term objectives

- 1 Effective noise from bulk \equiv Extrinsic noise
- 2 Coupling to gene regulation dynamics
- 3 **Inference**: e.g. identification of “boundary species”

↓
MEMORY TERMS

SECONDMENTS

Complementarities: Biochemical reaction networks



Inference in signalling and regulatory networks

① **Berlin:** METHODS

- Dynamical functional methods
- Approximate inference methods (VARIATIONAL TECHNIQUES + EP)

② **Rome/Torino:** APPLICATIONS

- Rome → Analysis of real cellular biochemical networks
- Torino → Inference of signal-transduction networks in cancer cell lines

Thank you for your attention!