2.9.16

Project description: Master Thesis

Considering biological constraints via L1-penalization

Background

Understanding of complex biochemical networks as they occur in living cells requires the combination of experimental work with mathematical modelling. Ordinary differential equation models (ODEs) can be used as mathematical representation for analyzing known biochemical interaction networks. A major goal is the calibration of such models, i.e. to estimate the parameters like initial concentrations or rate constants based on experimental data.

A limitation of the standard approach is that biological constraints are usually provided for the dynamic compounds (concentrations) and cannot be directly translated to parameters. Moreover, biological knowledge is often difficult to specify precisely, i.e. as hard constraints like x>y. Therefore, constraints may be slightly violated. This requirement is related to another challenge: sometimes it is impossible to find a model satisfying all constraints and then a solution with a least total violation should be found.

State of the art

- Experimental data is often limited and researcher therefore use literature knowledge for establishing models with biologically valid parameters [1].
- In our group, a comprehensive implementation of the parameter estimation methodology is available [2,3].
- A clear strategy/algorithm for accounting for biological constraints is not available in the literature.
- L1-penalization, i.e. changing the least-squares objective function like $\chi^2_{pen} = \chi^2 + \Sigma_i \lambda \, |\theta_i - \theta_{i,\,target}|$ is a current state-of-the-art approach in other statistical and bioinformatics fields for finding "sparse" solutions for high-dimensional problems.
- Recently the approach has been transferred to the systems biology setting for identification of cell-type-specific parameters [4].

Master thesis

In this master thesis, L1-penalization will be used to account for biological constraints. An existing implementation of parameter estimation and L1 penalization [2-4] can be used.

First, an illustrative example is established. Then, the methodology is applied to a realistic application model. The goal is to publish L1-penalization as a new method for considering biological knowledge and constraints. The project timeline is planned to have time for writing a paper.

Literature

[1] J. Rausenberger et al. Photoconversion and nuclear trafficking cycles determine phytochrome A's response profile to far-red light. *Cell* 146, 2011, 813-825.

[2] A. Raue, et al. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics* **31**, 2015, 3558-3560

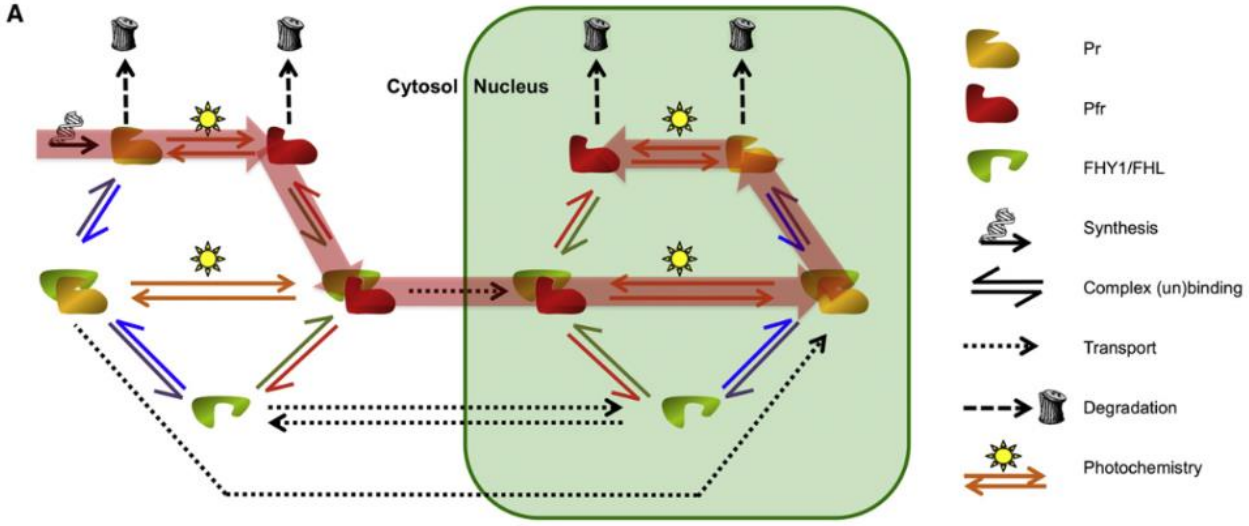
[3] Data2Dynamics Software Website
https://bitbucket.org/d2d-development/d2d-software/wiki/Home

[4] B. Steiert et al. L1 regularization facilitates detection of cell type-specific parameters in dynamical systems, *Bioinformatics*, 2016.

[5] D. Vidaurre et al. A Survey of L1 Regression, *International Statistical Review*, 2013, 81, 3, 361–387.

Figure(s)

A



B

Input conditions

Condition	Experimental phentoype	Transfer to model	Reference
phyA-YFP in D	phyA-YFP localizes to the cytosol	$phy_{nuc} \leq 5\%$ in D	Hiltbrunner et al., 2005; Kim et al., 2000
phyA-YFP in cFR	phyA-YFP accumulates in the nucleus	$phy_{nuc} \geq 10\%$ in cFR	Hiltbrunner et al., 2005; Kim et al., 2000
phyA Y242H-YFP in D	Hypocotyls are shorter in phyA Y242H-YFP plants than in wt	Pfrn(wt) < Pfrn(phyA Y242H-YFP) in D	Fig. S3
phyA Y242H-YFP in cFR	Hypocotyls are longer in phyA Y242H-YFP plants than in wt	Pfrn(wt) > Pfrn(phyA Y242H-YFP) in cFR	Fig. S3
fhy1 in cFR	Hypocotyls are longer in *fhy1* than in wt	Pfrn(wt) > Pfrn(*fhy1*) in cFR	Desnos et al., 2001; Whitelam et al., 1993; Zeidler et al., 2001
FHY1 OX in cFR	Hypocotyls are shorter in FHY1 OX lines than in wt	Pfrn(wt) < Pfrn(FHY1 OX) in cFR	Desnos et al., 2001
Varying wavelength λ	wt exhibits peak in action spectrum (any position)	Calculate Pfrn(wt) for different λ; find peak: $Pfrn(wt(\lambda_{max} \pm 20\ nm)) \leq 0.9 \times Pfrn(wt(\lambda_{max}))$	

Fig. 1: Panel A shows a model of light perception in plants. Panal B summarizes constraints used in [1] to establish the model. phyA denotes phytochrom A which is a photoreceptor, FHL and FHY1 are proteins (signal transducers), “nuc” is the abbreviation for nucleus, wt=wild type, “D”=darkness, “cFR” = continuous far red light, “Pr” denotes PhyA in the red light absorbing state, “Pfr” = PhyA in far-red absorbing state, and “Pfrn” = nuclear-localized Pfr. Y242H-YFP is an experimentally used fluorescent label where “YFP” is the abbreviation for yellow fluorescent protein.