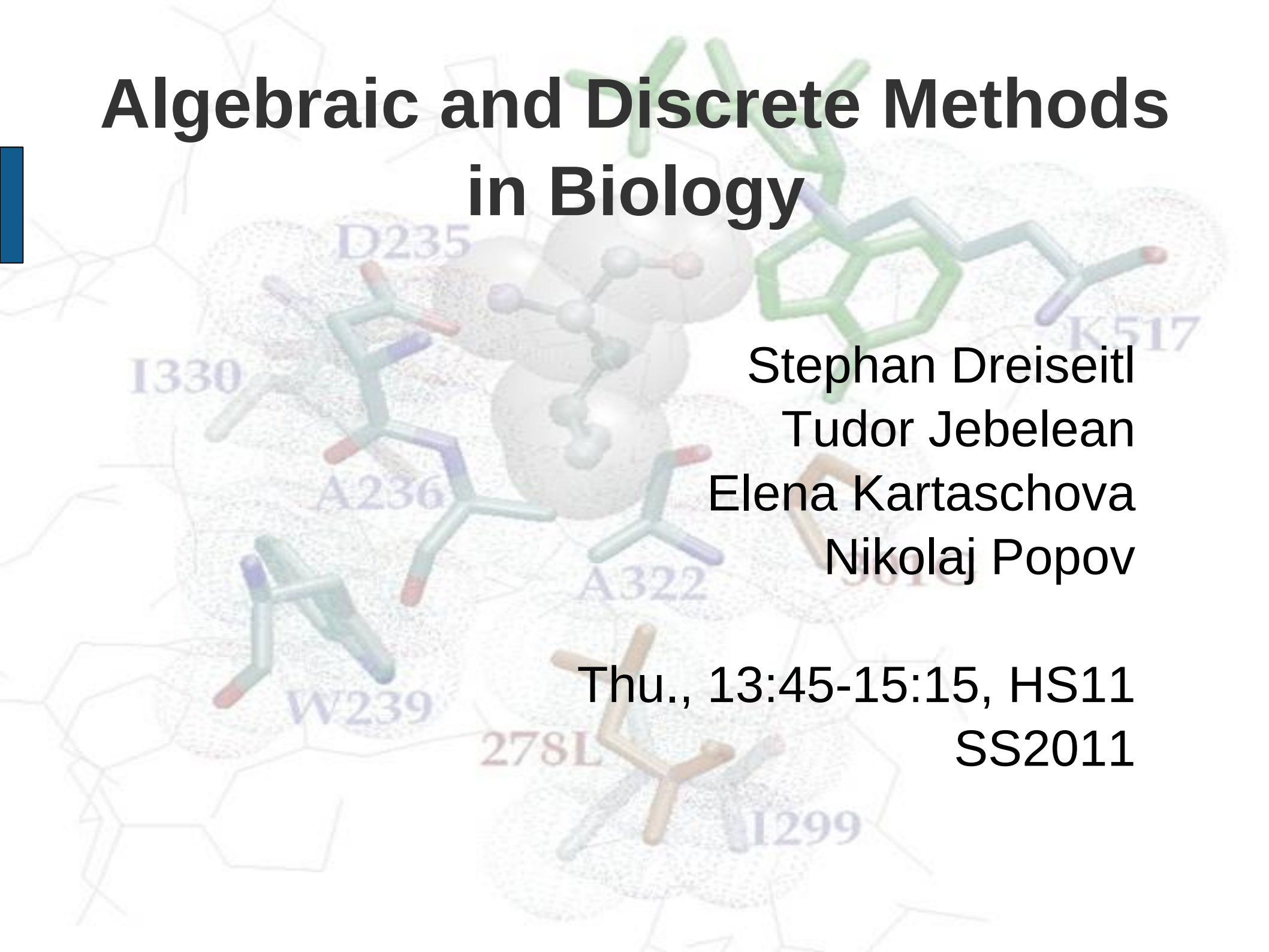
Algebraic and Discrete Methods in Biology



Stephan Dreiseitl
Tudor Jebelean
Elena Kartaschova
Nikolaj Popov

Thu., 13:45-15:15, HS11
SS2011

Meeting dates and lecturers

- Stephan Dreiseitl: March 17th, 24th
- Tudor Jebelean: March 31st, April 7th, 14th
- Nikolaj Popov: March 10th, May 5th, 12th, 19th
- Elena Kartaschova: May 26th, June 9th, 16th
- Exam: June 30th

Text book

- Leah Edelstein-Keshet. *Mathematical Models in Biology*. SIAM, 2005.

ROC Analysis

- In machine learning, want to be able to compare performance of one algorithm with that of another
- Bioinformatics/medical informatics mostly use Receiver Operating Characteristics analysis – known as ROC analysis

ROC Analysis

- ROC analysis provides graphical representation and one-number summary of discriminatory power
- We will talk about this measure, derive its properties, and show how it can be applied

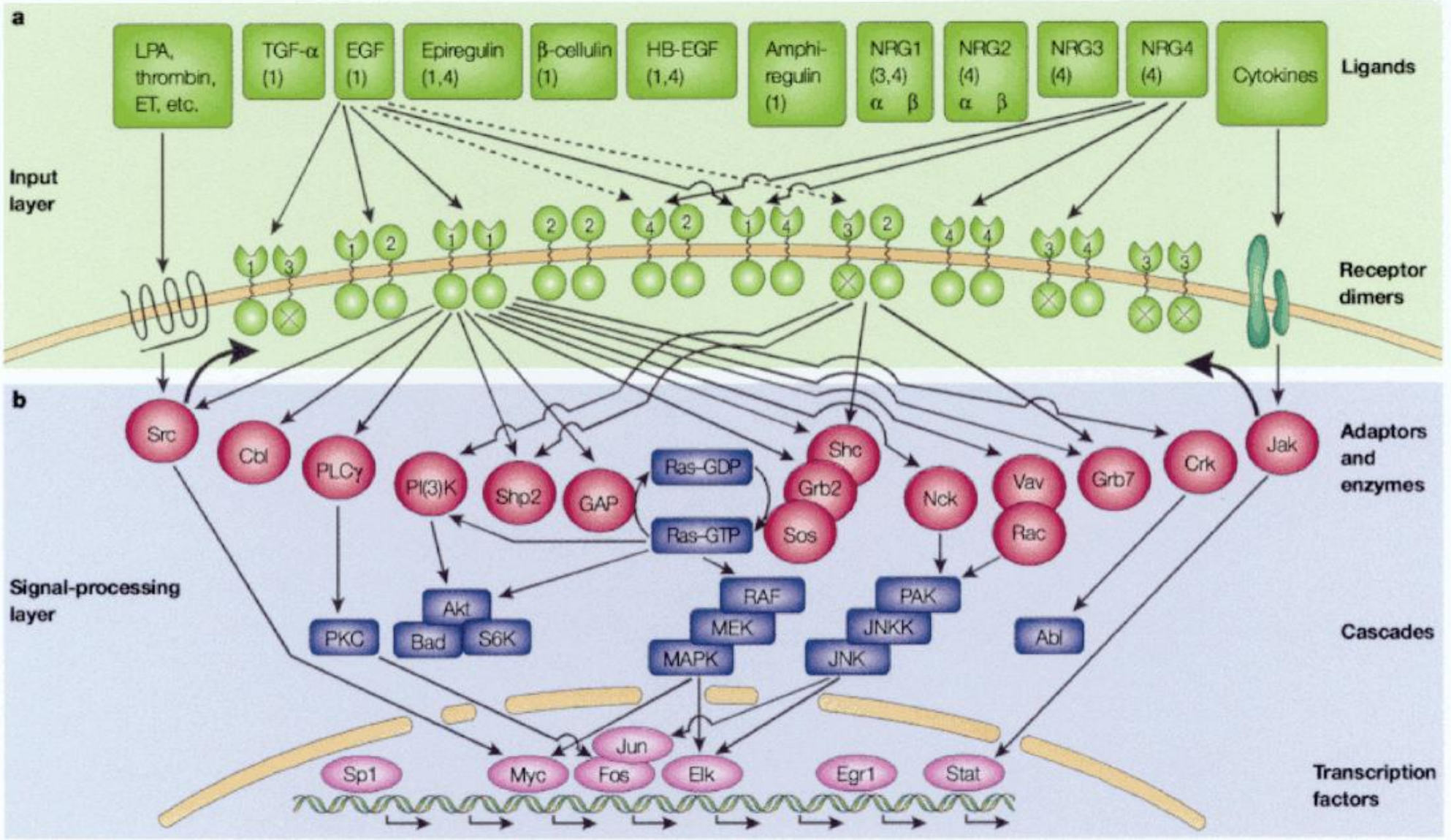
Haplotype tagging SNPs

- Single-Nucleotide Polymorphisms (SNPs) are locations where genomic information differs between individuals
- Problem is to find the smallest number of SNPs that uniquely determine an individual's haplotype

Haplotype tagging SNPs

- Identification of ht-SNPs computationally complex (known to be NP-hard)
- We will approach the problem by relating it to the minimum set cover problem, and present a greedy algorithm for approximating the optimal solution

Logic: Proteins in the cell



Logic: Proofs

Wolfram Mathematica 6.0 - [Activation-by-Presence.nb *]

File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

Activation-by-Presence.nb *

Prove:

(Lemma (Activation by Presence))

$$\forall_{C,P,Q} \left(\text{IsPresent}[P, C] \wedge \text{IsPresent}[Q, C] \wedge \text{Activates}[P, Q] \Rightarrow \text{IsPresent}[\text{active}[Q], C]\right),$$

under the assumptions:

(Definition (Activation))

$$\forall_{P,Q} \left(\text{Activates}[P, Q] \Leftrightarrow \forall_{C} \left(\text{IsPresent}[P, C] \Rightarrow \text{BecomesActive}[Q]\right)\right),$$

(Axiom (Active Presence))

$$\forall_{C,P} \left(\text{IsPresent}[P, C] \wedge \text{BecomesActive}[P] \Rightarrow \text{IsPresent}[\text{active}[P], C]\right).$$

For proving (Lemma (Activation by Presence)) we take all variables arbitrary but fixed and prove:

(1)
$$\text{IsPresent}[P_0, C_0] \wedge \text{IsPresent}[Q_0, C_0] \wedge \text{Activates}[P_0, Q_0] \Rightarrow \text{IsPresent}[\text{active}[Q_0], C_0].$$

We prove (1) by the deduction rule.

We assume

(2) $\text{IsPresent}[P_0, C_0] \wedge \text{IsPresent}[Q_0, C_0] \wedge \text{Activates}[P_0, Q_0]$

and show

200%

Logic: Rewriting

Rewrite rules are used to express biochemical events.

Example

- A biochemical event of binding EGF to the EGFR, the first step in the activation of the EGFR signaling pathway:

  Activated Erk1 is translocated to the nucleus where it is functionally sequestered and can regulate the activity of nuclear proteins including transcription factors.

- The same event expressed in rewrite rules:

```
rl[410.Erk1/2.to.nuc]:
   {CM | cm:Soup
            {cyto:Soup [Erk1 - act] {NM | nm:Soup {nuc:Soup}}}}
   =>
   {CM | cm:Soup
            {cyto:Soup {NM | nm:Soup {nuc:Soup [Erk1 - act]}}}} .
rl[438.Erk.act.Elk]:
   [?Erk1/2 - act] Elk1 => [?Erk1/2 - act] [Elk1 - act] .
```

Difference equations

- A difference equation is any equation of the form $x_{n+1} = f(x_n , x_{n-1} , \ldots , x_1)$, where x_k is the $k-$th element of a sequence $\{x_k\}$ and f is a function of n arguments.

- Depending on the functional form of f, we distinguish linear and nonlinear difference equations

Application of linear difference equations

- Cell division
- Insect population
- Propagation of annual plants
- Growth of segmental organisms
- Schematic model of red blood cell production

Application of nonlinear difference equations

- Density dependence in single-species populations
- Two-species interactions: host-parasitoid systems

Pattern formation in biological systems

Diffusion equations

A simplest classical model for pattern formation in an organism (morphogenesis) is based on ***at least*** two diffusion equations:

$$\frac{\partial C_1}{\partial t} = R_1(C_1, C_2) + D_1 \frac{\partial^2 C_1}{\partial x^2} \tag{1}$$

$$\frac{\partial C_2}{\partial t} = R_2(C_1, C_2) + D_2 \frac{\partial^2 C_2}{\partial x^2} \tag{2}$$

with C_1, C_2, D_1, D_2 being concentrations and diffusion rates of two chemicals.

PDEs generating patterns

Regard (1) as equation of one variable C_1

$$\frac{\partial C_1}{\partial t} = R_1(C_1) + D_1 \frac{\partial^2 C_1}{\partial x^2} \tag{3}$$

with $R_1(C_1)$ being nonlinear function of C_1. Eq.(3) belongs to the class of nonlinear PDEs which can generate pattern formation.
Example of a nonlinear PDE from this class is Swift-Hohenberg equation (SH), it has been used to model patterns in a variety of biological materials, such as neural tissues. On-line *numerical simulations* with SH showing formation of various patterns (depending on the initial conditions and values of a parameter ε) can be found here:

- http://www.cmp.caltech.edu/ mcc/Patterns/SwiftHohenberg.html