

## Fehérjék feltekeredése Hálózatok

Hegedűs Tamás

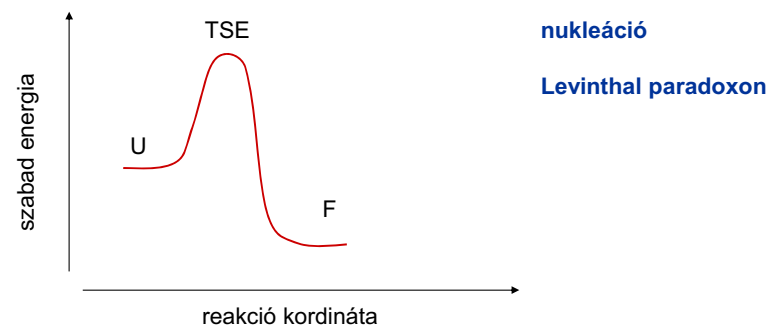
tamas@hegelab.org



MTA-SE Molekuláris Biofizikai Kutatócsoport  
SE Biofizikai és Sugárbiológiai Intézet



## Kétállapotú fehérje feltekeredés Two-state model of protein folding



## Kétállapotú fehérje feltekeredés Two-state model of protein folding

$$G_{\text{protein}} = H_{\text{protein}} - TS_{\text{protein}}$$

$$-RT \ln K = \Delta G_{\text{protein}} =$$

$$= \Delta H_{\text{protein}} + \Delta H_{\text{solvent}} + \Delta H_{\text{protein-solvent}} - T\Delta S_{\text{protein}} - T\Delta S_{\text{solvent}}$$

## Fehérje stabilitás I.

Konformációs stabilitást elősegítik:

- Hidrofób kölcsönhatások
- Intramolekuláris H-híd kötések
- Intramolekuláris ionos kölcsönhatások
- Intramolekuláris van der Waals kölcsönhatások
- Intramolekuláris diszulfid hidak

Destabilizáló tényezők:

- H-híd az oldószerrel
- Van der Waals kölcsönhatás az oldószerrel
- Az ionos csoportok solvatációja
- entrópia

## Fehérje stabilitás II.

A fehérje stabilitás nem ér el maximális értéket.

Erre utalnak:

termofil baktériumok fehérjéi;  
igen stabil, tervezett fehérjék.

Ennek okai lehetnek:

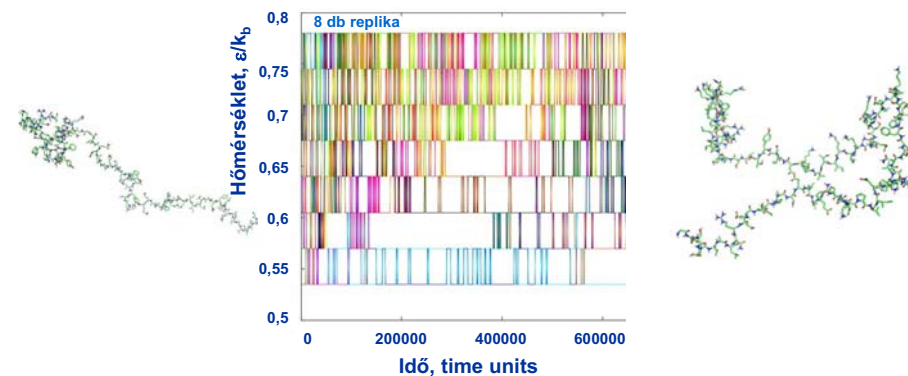
az evolúció nem igényel stabilabb fehérjét mint a funkció önmaga;  
a fehérjéknek le is kell bomlaniuk;  
a funkcióhoz flexibilitás szükséges.

Folding szimulációk során mi az abszolút energiaminimumot  
(maximum stabilitást) keressük.

## Fehérje feltekeredés szimulációja

All atom force-field:

Potenciál függvény számolása erőforrásigényes  
Reprezentatív konformációs mintavételezés problémás  
Umbrella sampling, [replica exchange](#).



## Összefoglalás

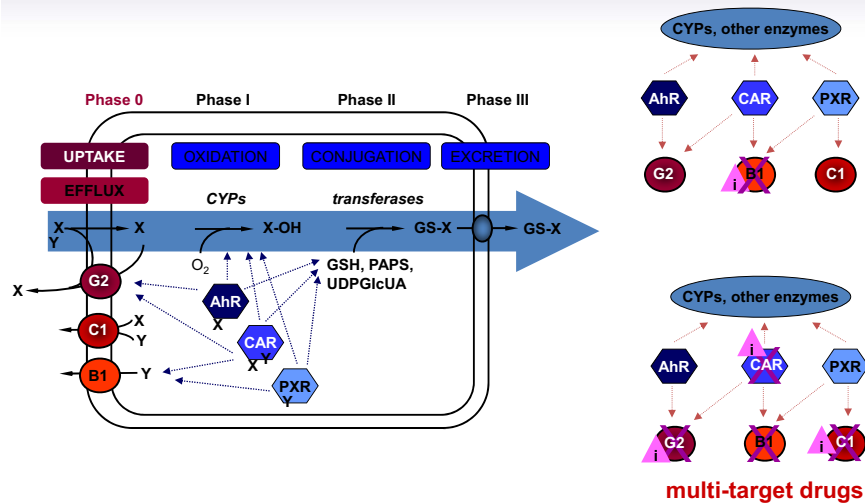
- Bevezetés – a fehérje dinamika és a szimulációk jelentősége
- Fehérjék jellemzése bioinformatikai eszközökkel
- Fehérjék dinamikájának modellezése
- Fehérjék feltekeredésének szimulációja

## Hálózatok – fehérje hálózatok

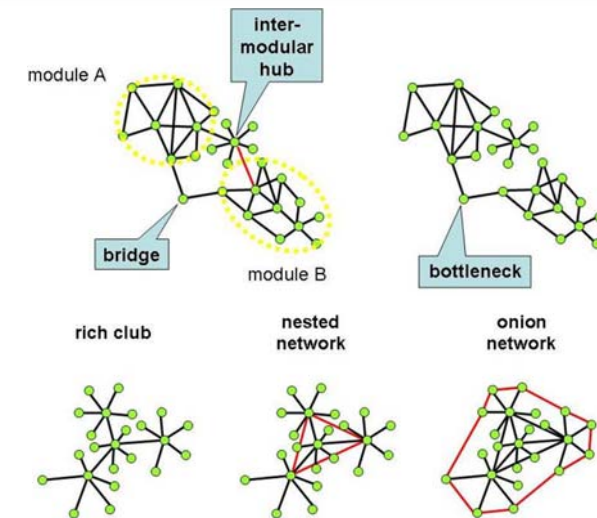
- Rendszerbiológia
- Fehérjék-fehérjék kölcsönhatási hálózata
- Gének-fehérjék-drogok kölcsönhatása
- Kapcsolatok aminosavak mozgásaiban

Csermely P. et al. 2012, <http://arxiv.org/abs/1210.0330>

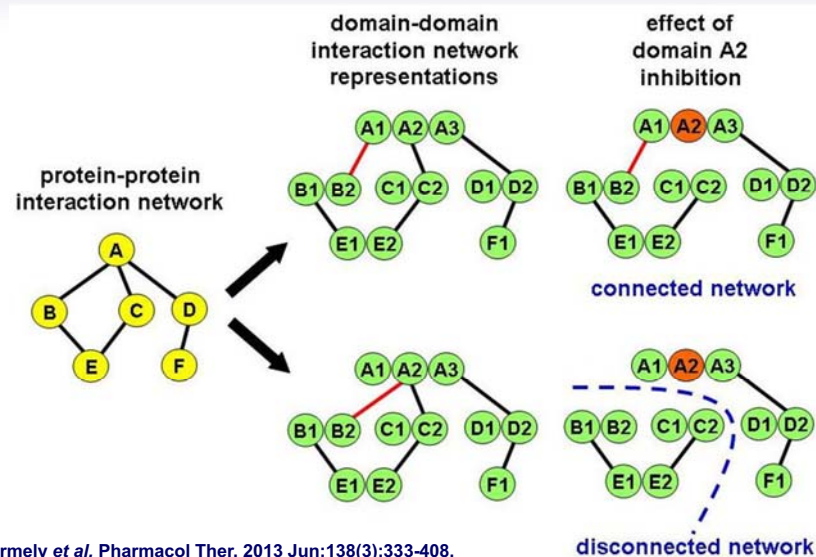
## A kemoimmun (ChI) rendszer (sejtszintű immunitás)



## Hálózatok szerkezete

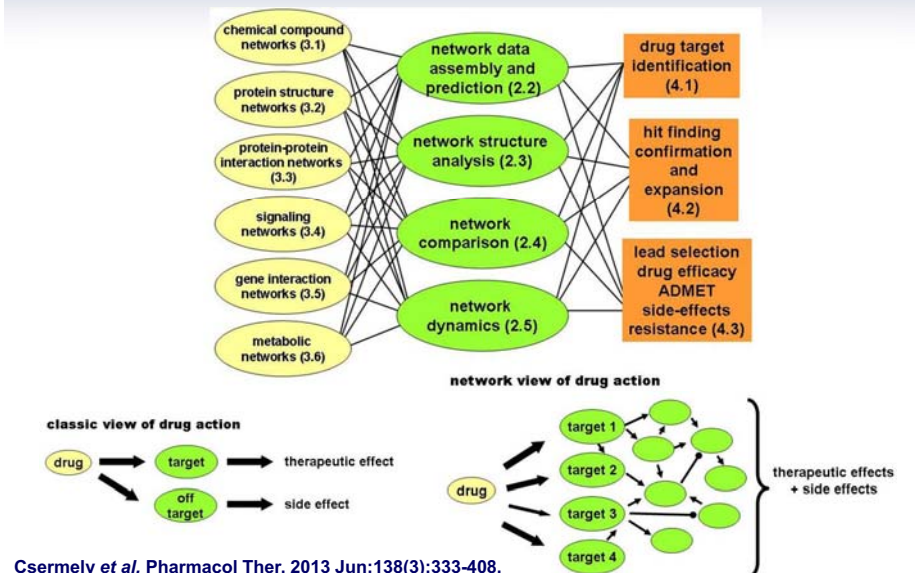


## Fehérje-fehérje kölcsönhatási hálózatok



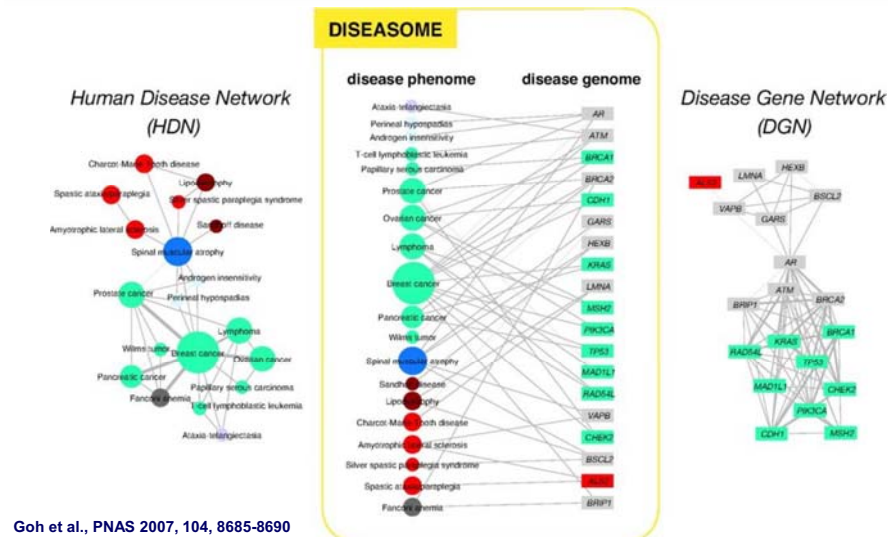
Csermely et al. Pharmacol Ther. 2013 Jun;138(3):333-408.

## Hálózatok és gyógyszertervezés



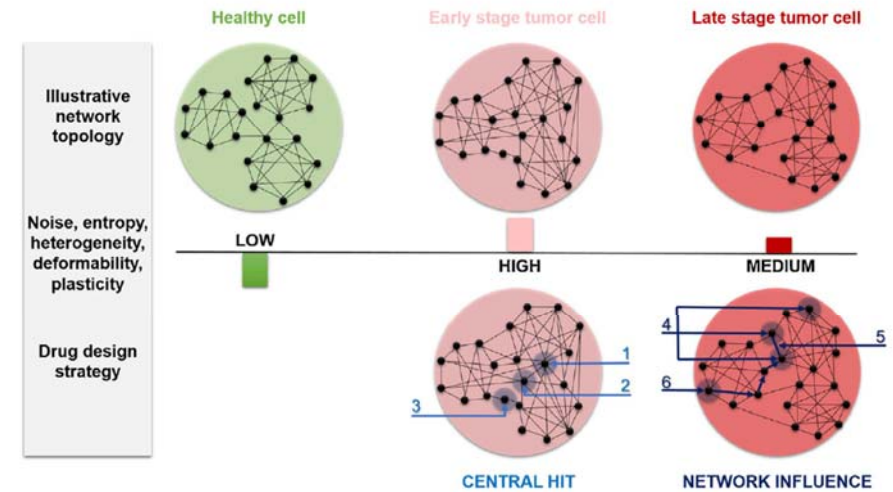
Csermely et al. Pharmacol Ther. 2013 Jun;138(3):333-408.

## Network assembly

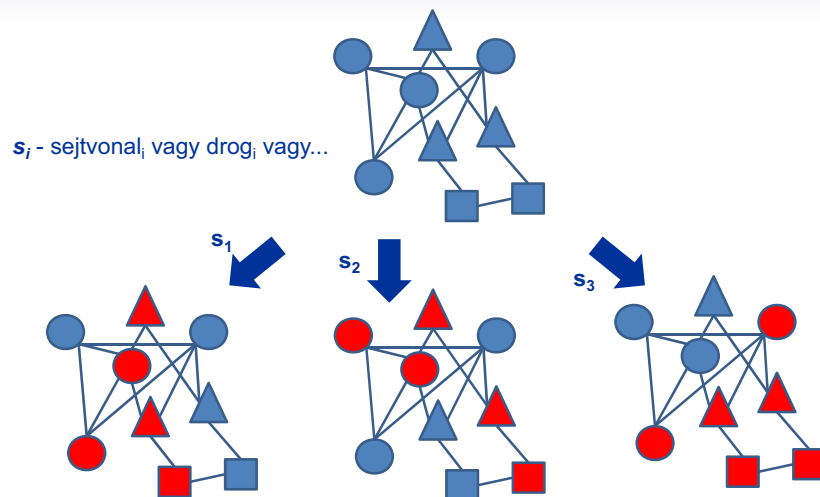


## Rákos sejtek hálózatai

D.M. Gyurkó et al. / Seminars in Cancer Biology 23 (2013) 262–269



## A kemoimmunitási hálózat működésének feltérképezése



Data

Expression-changes

Expression-patterns

Biochemical pathways

## Pipeline of analysis

- Human samples treated with drugs
- NCBI Gene Expression Omnibus (GEO) database
- Preprocessed (by GEO) data & quality check

	our interest	example(s)
DataSet (GDS) <sup>1</sup>	180 (1 335 human)	Anti-cancer agent saphyryn PCI-2050 effect on lung cancer cell line: dose response (GDS2499)
Experiment <sup>2</sup>	883 (2786 cont.+treat.)	treatment: 1) Actinomycin-D 5 ug/ml 2-3) Saphyryn PCI-2050 1.25 ug/ml, 2.5 ug/ml
Tissue/cell	132	lung cancer cell line, MCF-7, HUVEC, primary fetal astrocytes, tumor biopsies ...
Drug or xenobiotic	222	actinomycin D, saphyryn PCI-2050, thapsigargin, tunicamycin, doxorubicin ...
Microarray platform (GPL)	26	Affymetrix - Human Genome U133 Plus 2.0 Array (GPL570)

<sup>1</sup>Collection of coherent experiments (by GEO)

<sup>2</sup>One celltype, one agent, one timecourse, one dose



## Whole dataset

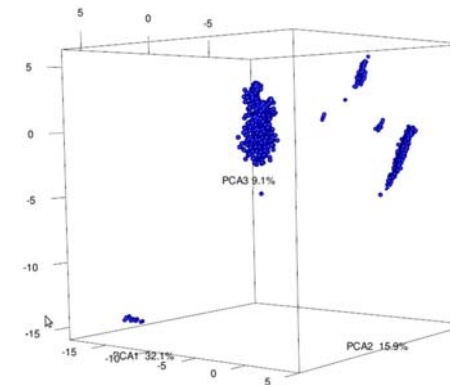
- Calculate the expression changes
  - Discretization

	Expression change (fold)	Discret value
upregulated	>2x	1
downregulated	<0,5x	-1
no change	-	0
no probe on chip	-	2

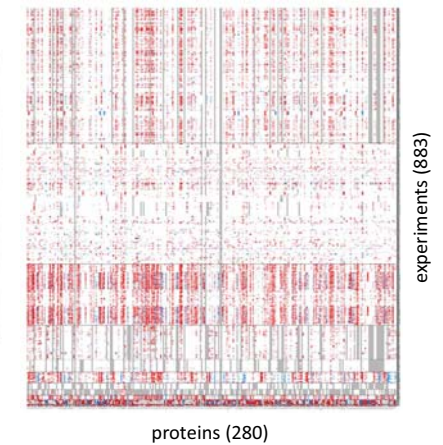
- Vectors

[illegible]

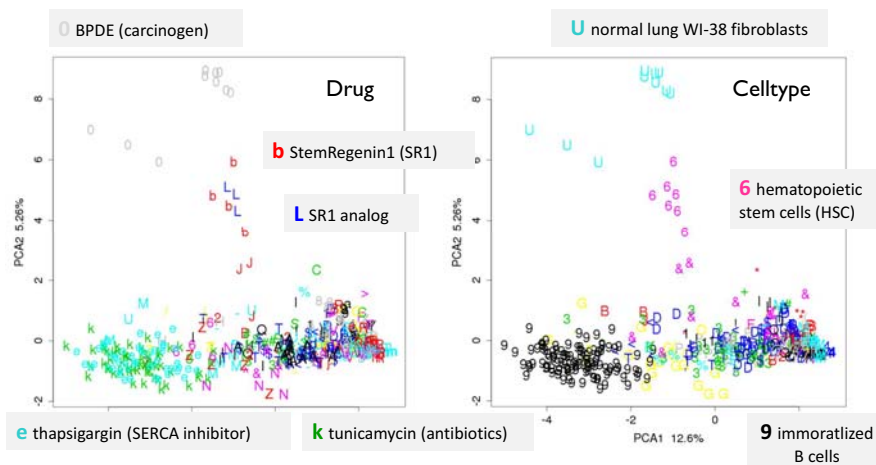
PCA analysis (the first 3 component)



Heatmap (result of clustering)



## PCA analysis



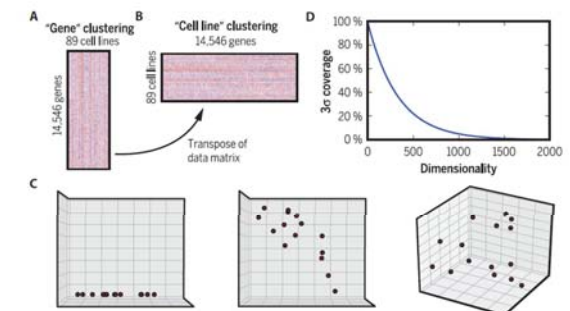
## REVIEW

## COMPUTATIONAL BIOLOGY

## Avoiding common pitfalls when clustering biological data

Tom Ronan, Zhijie Qi, Kristen M. Naegle\*

www.SCIENCESIGNALING.org 14 June 2016 Vol 9 Issue 432 re6



**Fig. 1. Determining the dimensionality of a clustering problem.** (A and B) Representation of the mRNA clustering problem consisting of >14,000 mRNAs measured across 89 cell lines. Data are from Lu et al. (6). When the mRNAs are clustered, the mRNAs are the objects and each cell line represents a feature, resulting in an 89-dimensional problem (A). When attempting to classify normal and tumor cell lines using gene expression, the cell lines are the objects and each mRNA is a feature, resulting in a clustering problem with thousands of dimensions (B). (C) Effect of dimensionality on sparsity. (D) Effect of dimensionality on coverage of the data based on SD from the mean.

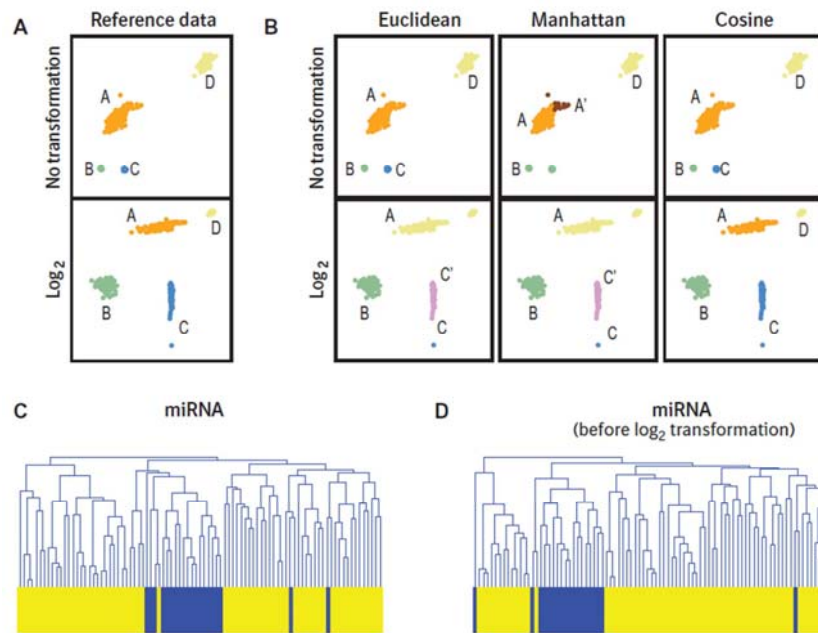


Fig. 3. Effect of transformations and distance metric on clustering results. (A) Demonstration of how trans-

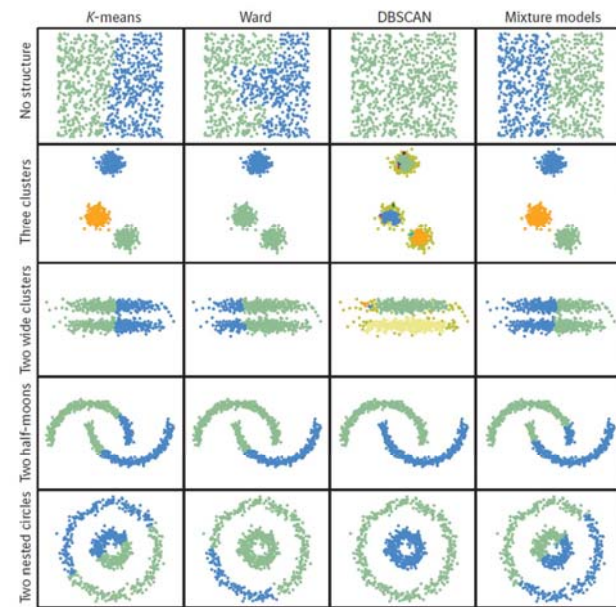


Fig. 4. The effect of algorithm on clustering results. Four toy data sets (<https://github.com/knaeigle/clusteringReview>) demonstrate the effects of different types of clustering algorithms on various known structures in two-dimensional data.

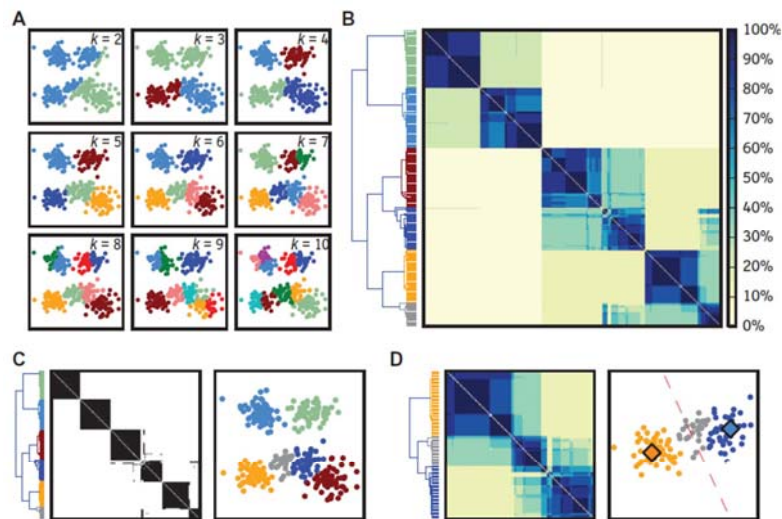
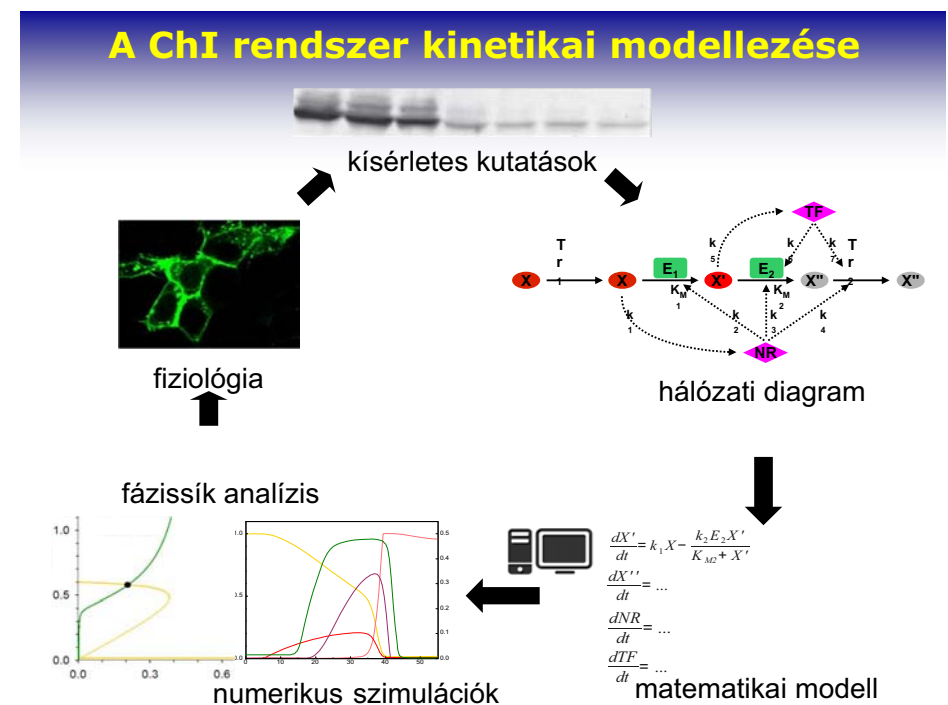
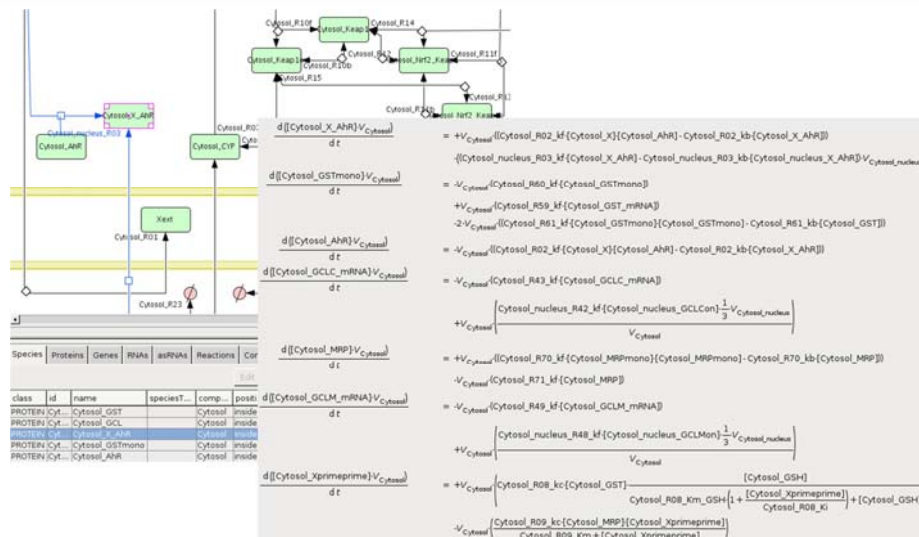


Fig. 5. Ensemble clustering overview. Finishing techniques were applied to random toy data (see file S1 for analysis details). (A) Set of clustering results obtained using the  $k$ -means algorithm with various values of  $k$  (a  $k$ -sweep). (B) Hierarchically clustered (Ward linkage) co-occurrence matrix for the ensemble of results in (A). The heatmap represents the percentage of times any pair of data points coclusters across the ensemble. (C) A majority vote analysis was applied (left panel) using a threshold of 50% on the co-occurrence matrix in (B). Six clusters (see dendrogram color groupings) result from the majority vote (right panel). (D) Application of fuzzy clustering to the ensemble. The left panel shows the details of the

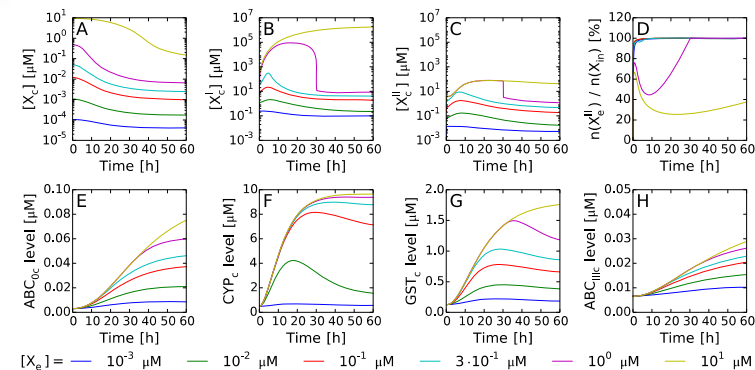


## Mathematical modeling

to understand relationships in the chemodefense network



## A rendszer kinetikai modellezése



Tóth et al. 2015 PLOS ONE

## Informatikai "szemlélet"

- Adat - RDBMS, ORM
- Logika - Python
- Megjelenítés - browser, templating

## Saját programok – programozási nyelvek

- C/C++:** lassú fejlesztés  
ha sebesség kell; mégis ritkán tanácsolt
- Java:** lassú fejlesztés; általában a szükségelt csomag beta ☹
- Script nyelv:** igen gyors fejlesztés  
bizonyos feladatokhoz igen lassú

- Könyvtárak
- Olvashatóság, dokumentálhatóság
- Objektum orientáltság
- Több fejlesztő: subversion vagy hasonló megoldások
- Adatbázis kezelés (Big Data); ORM!
- Egyéni száj-íz

# A Kiválasztott szkriptnyelv

## Java

```
public class HelloWorld
{
    public static void main (String[] args)
    {
        System.out.println("Hello!");
    }
}
```

```
int myCounter = 0;
String myString = String.valueOf(myCounter);
if (myString.equals("0")) ...
```

## Perl

```
sub foo {
    my $arg1 = shift;
    my @rest_of_args = @_;
    ...do stuff...
}
```

## Python

```
print "Hello, world!"
```

```
myCounter = 0
myString = str(myCounter)
if myString == "0": ...
```

```
def foo (arg1, rest_of_args):
    ...do stuff...
```

# KÖSZÖNÖM A FIGYELMET!

➤ [tamas@hegelab.org](mailto:tamas@hegelab.org)

➤ Álláslehetőség: [www.hegelab.org](http://www.hegelab.org)