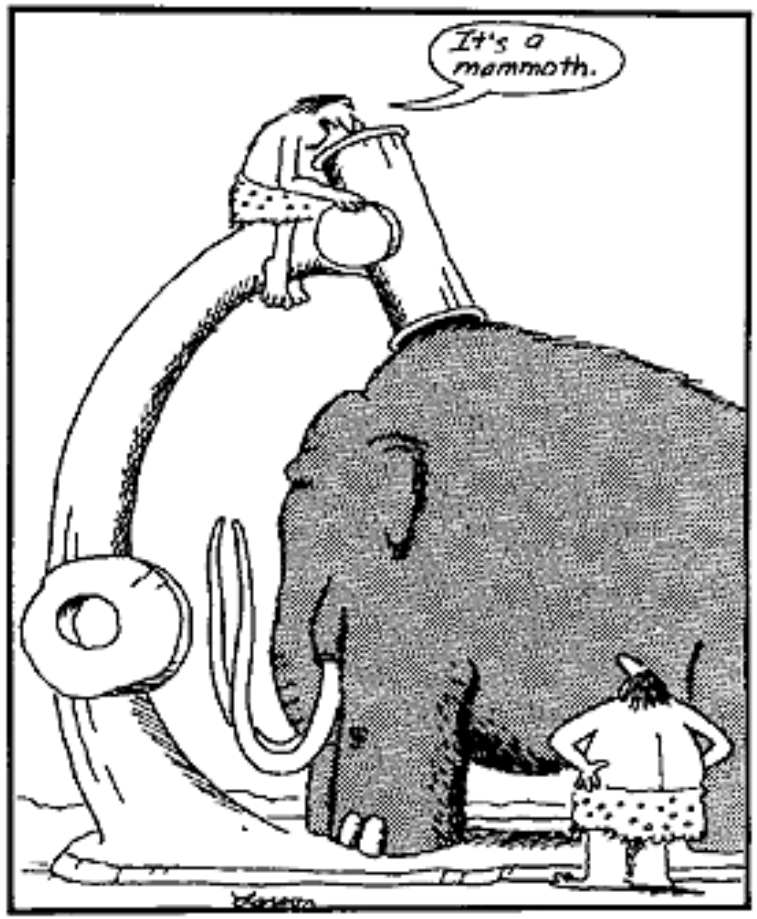
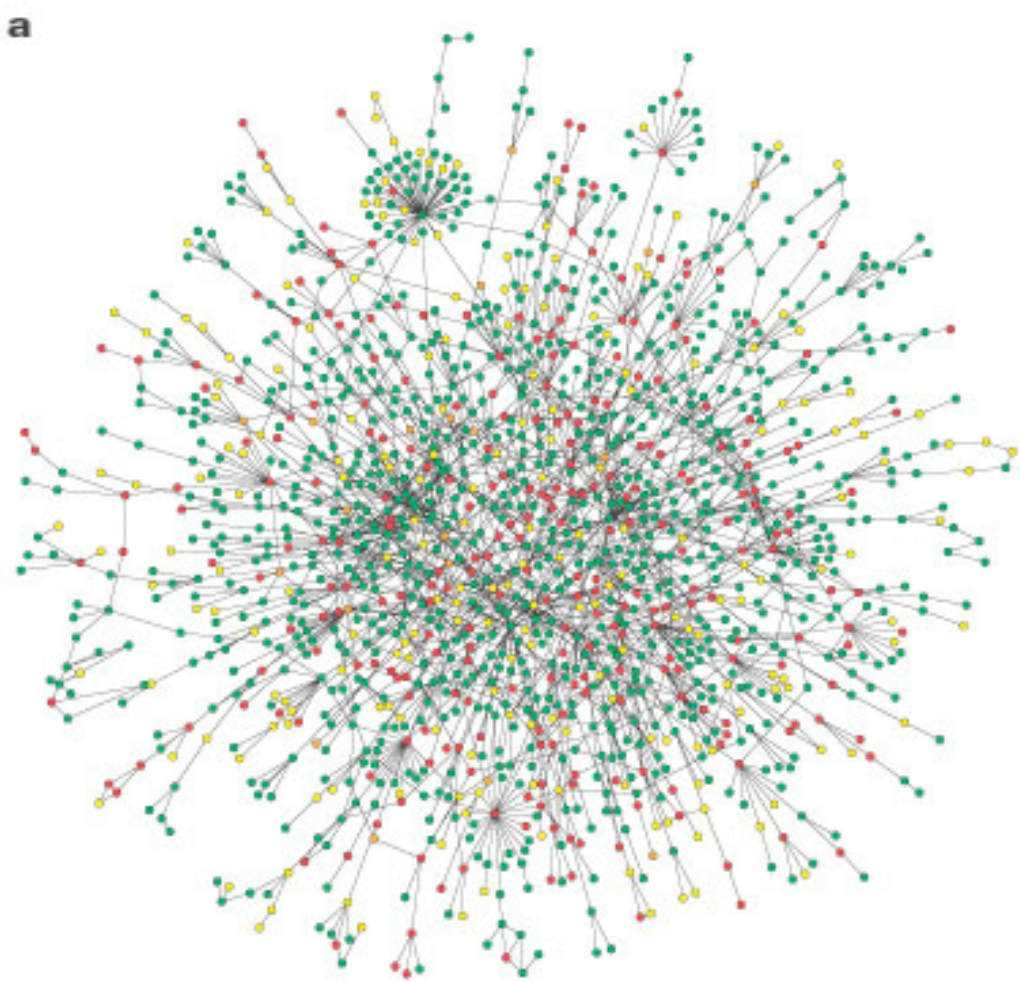
# Using Evolutionary Approaches To Study Biological Pathways

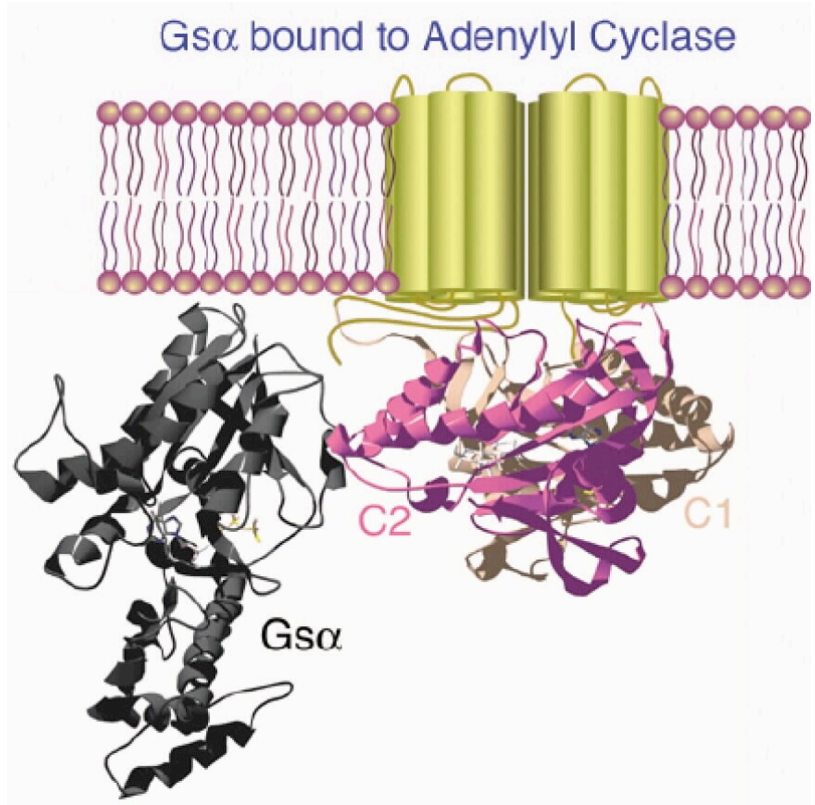



**Orkun S. Soyer**
The Microsoft Research - University of Trento Centre for Computational and Systems Biology



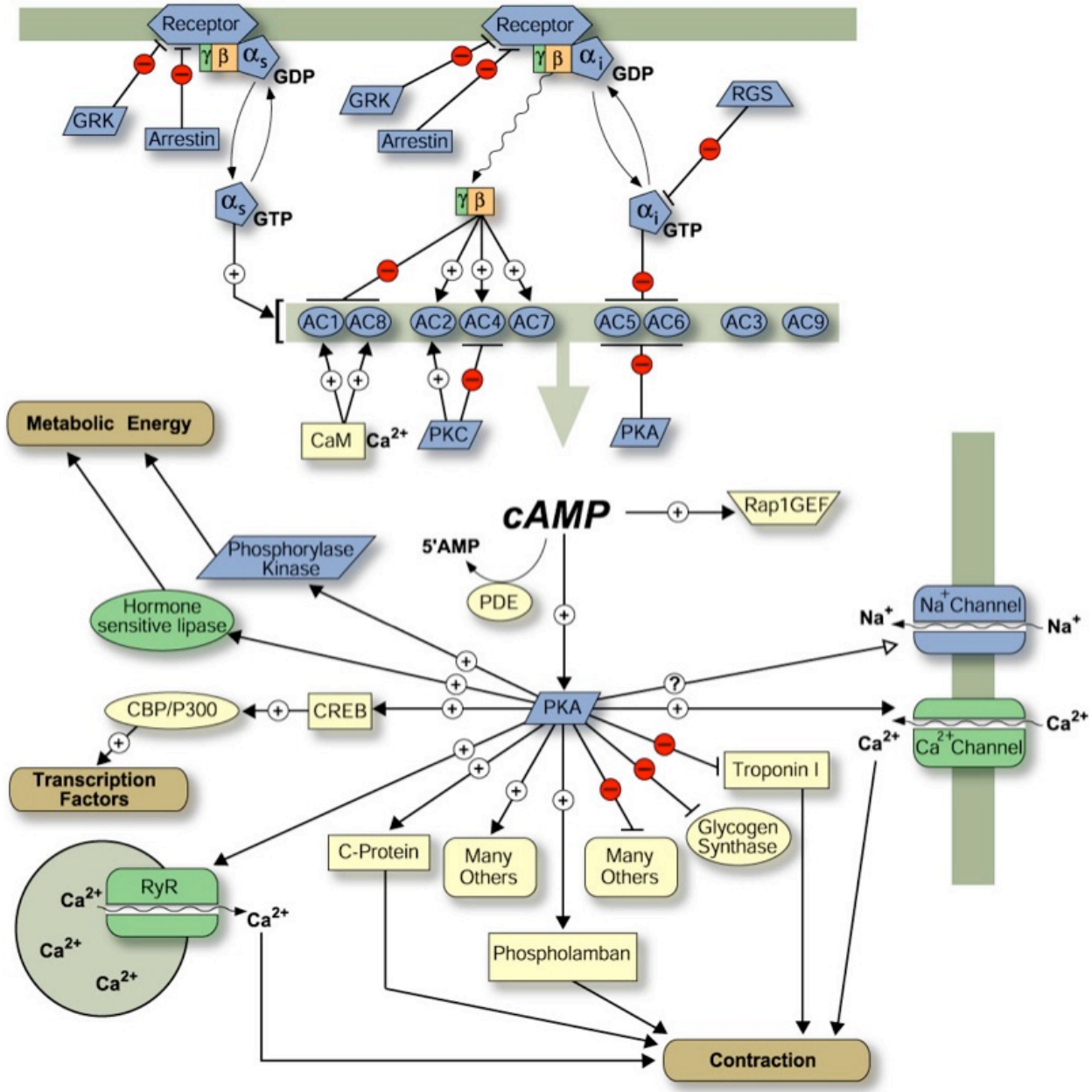

Protein-protein interactions map from yeast. Jeong H. et. al., Nature 411, 2001

# Proteins do interact...

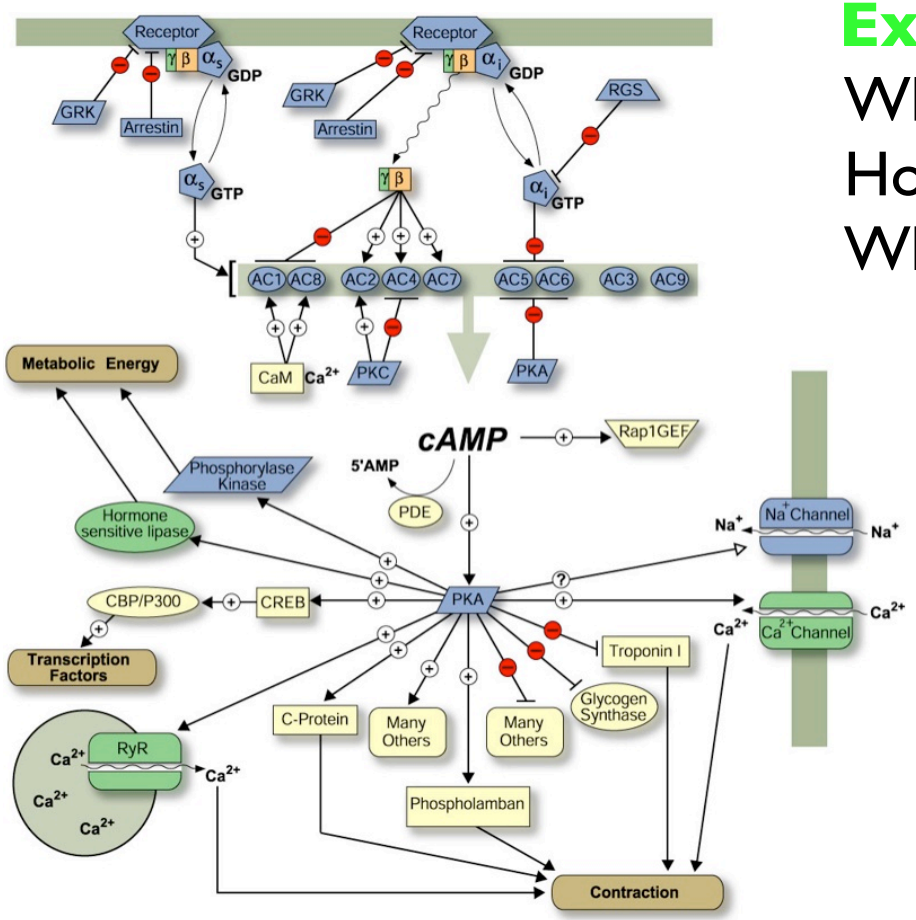


... constituting biological pathways.

Generic representation of pathways involving cAMP



# Studying Biological Pathways: Common Approaches




## Experimental techniques

Which proteins are involved?
How do specific proteins interact?
What are the interaction kinetics?

## Conventional Modeling

What are the response dynamics? How do they change with different conditions (e.g. under drug treatment)?

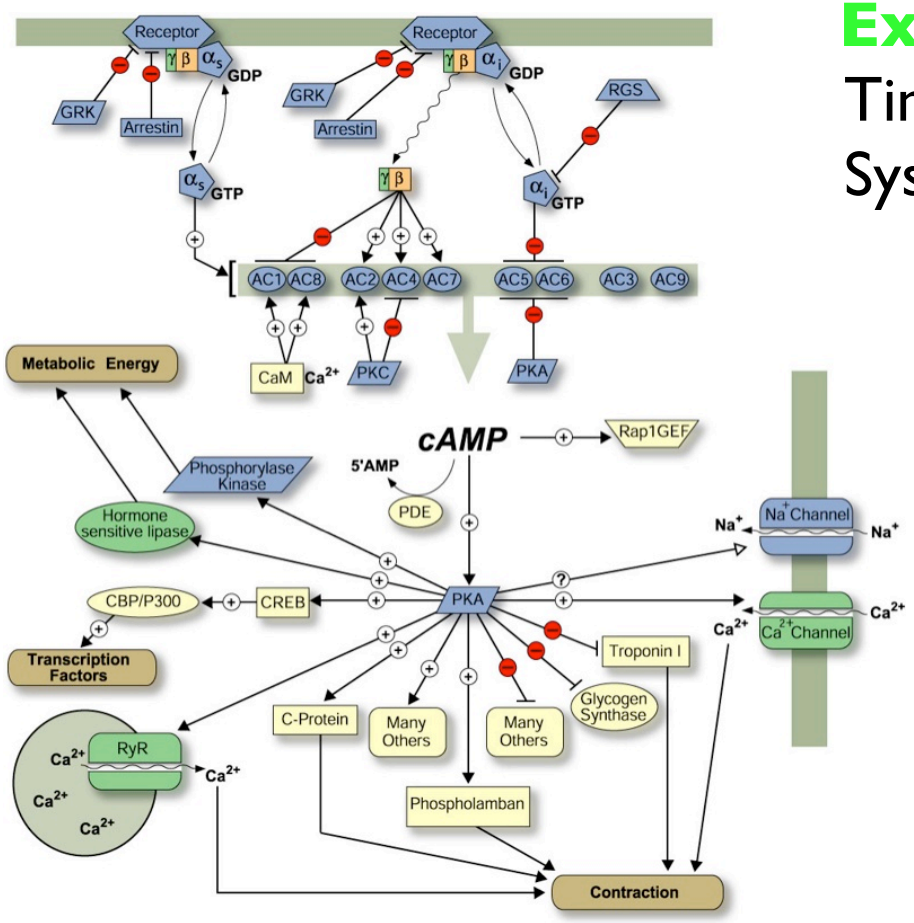
## Pathway Discovery

Can we detect new pathways from low level data like protein sequences?

## Large Scale Data Analysis

Are there any general network properties?
Connectivity distribution, re-occuring interaction motifs, etc. "Design Principles"???

# Studying Biological Pathways: "Limitations"




## Experimental techniques

Time, techniques, and money

System specific results

## Conventional Modeling

Sparse data

System specific results

## Pathway Discovery

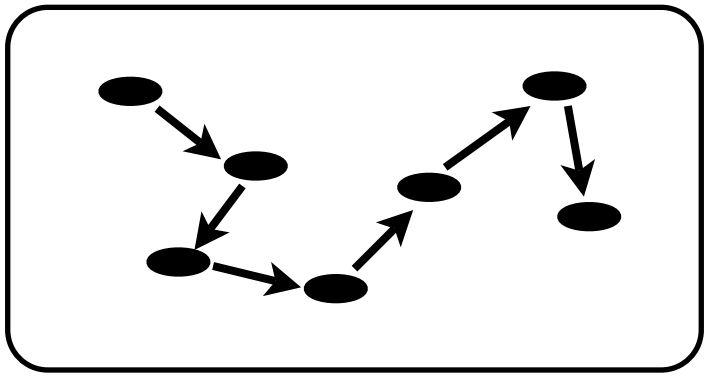
Usual suspects; reliability of data, algorithms, etc.

## Large Scale Data Analysis

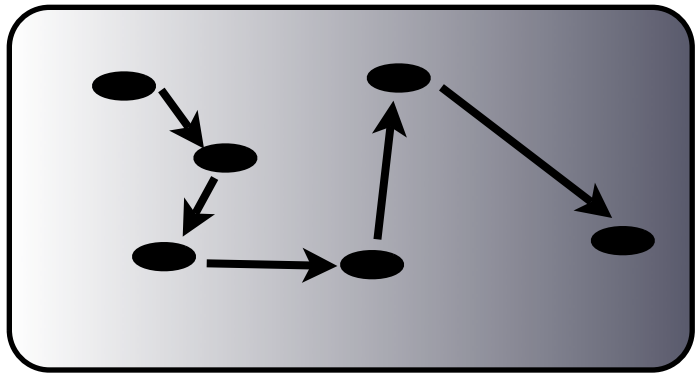
Broad conclusions (too broad ?)

# Bacterial Chemotaxis: The Behavior

Chemotaxis is a biased random walk:

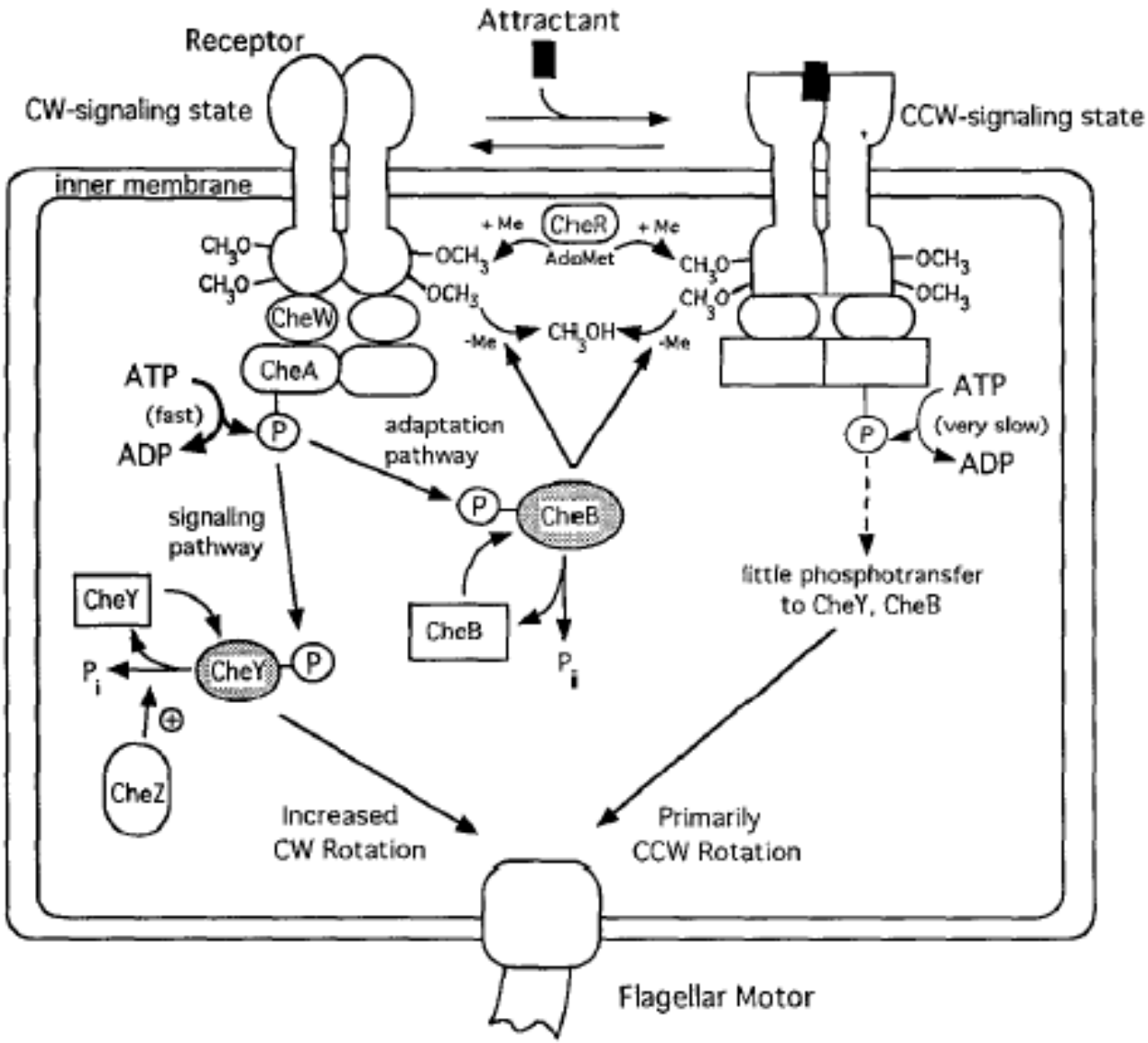


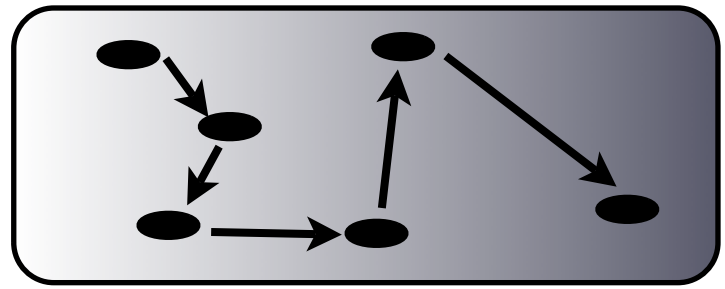
Constant Tumbling Frequency



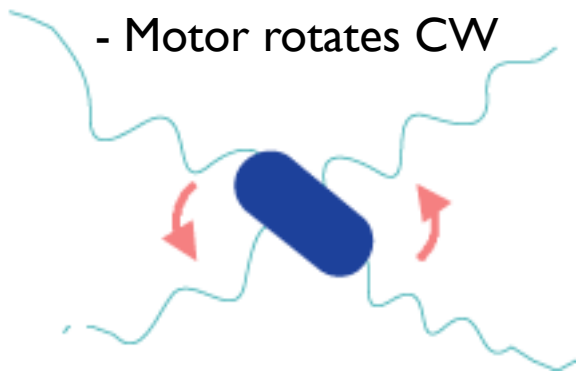
Decreasing Tumbling Frequency

# Bacterial Chemotaxis: The Pathway



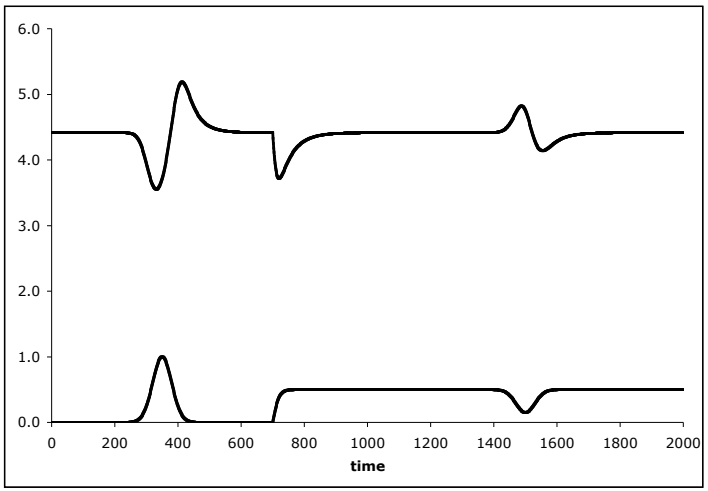
Tumbling:
- CheY-P bound
- Motor rotates CW



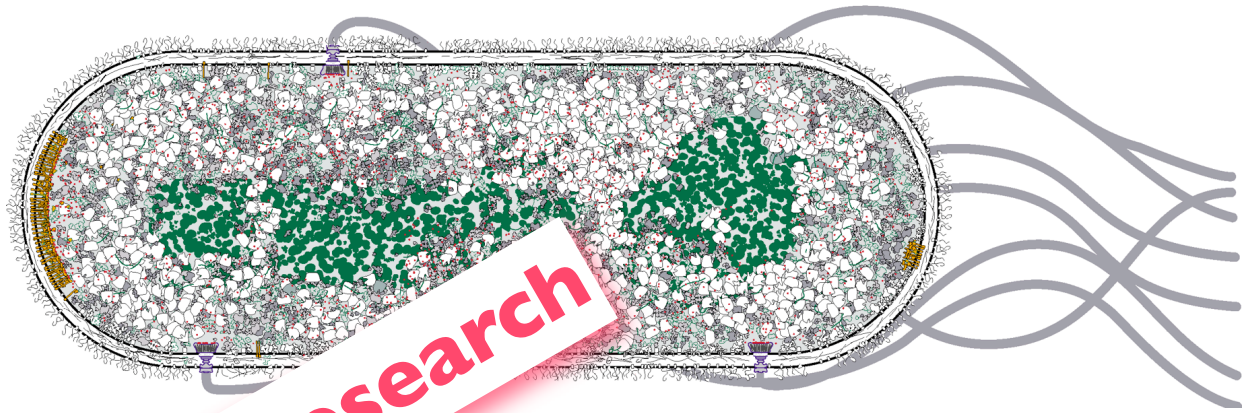
Smooth Swimming:
- CheY-P unbound
- Motor rotates CCW

# Bacterial Chemotaxis Solved

## Adaptive response



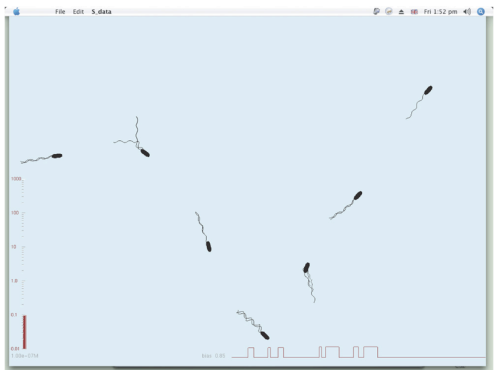

## Receptor and protein localization



30 years of research

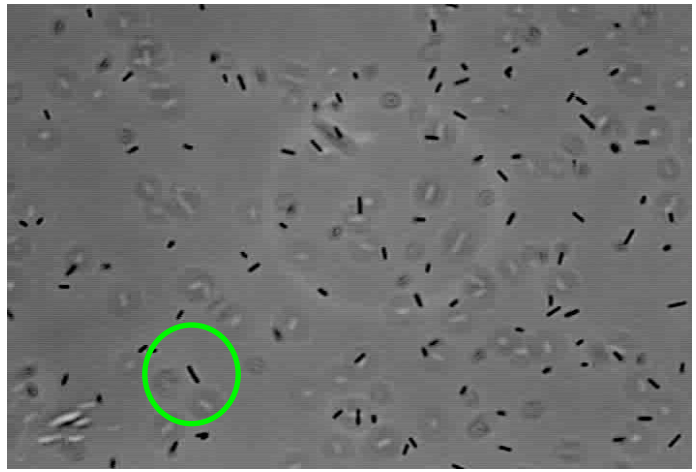
## ased random walk

# Bacterial Chemotaxis Solved Or Not

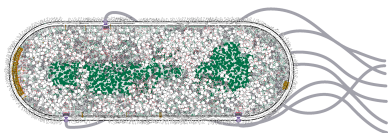


?

?



*Halobacterium salinarum*

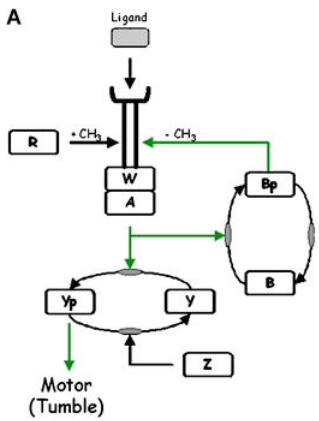


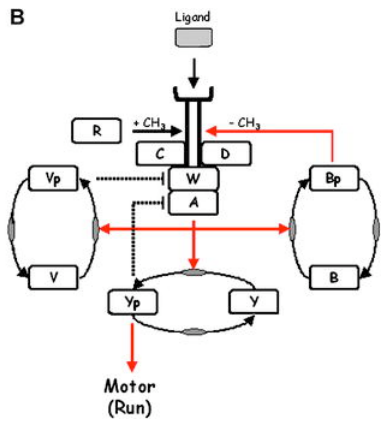
*Escherichia Coli*

*Bacillus subtilis*

*Rhodobacter sphaeroides*

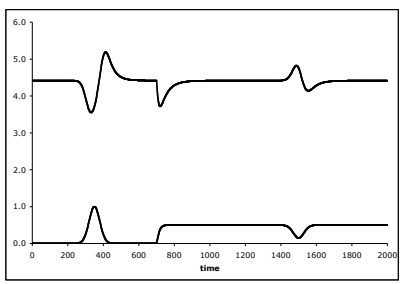
*Helicobacter pylori*




?

?

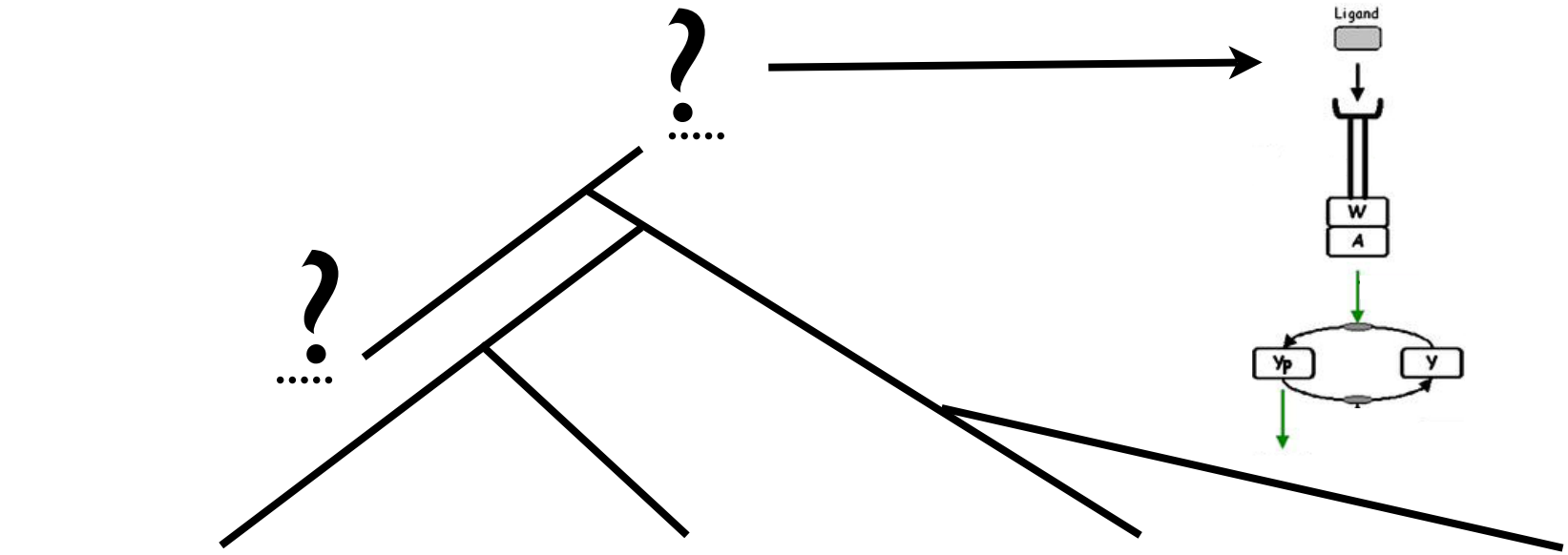



?

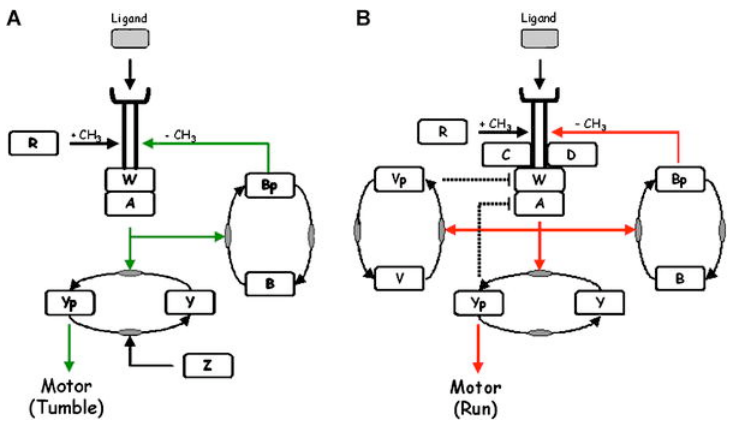
?

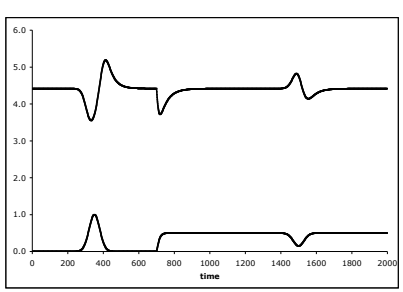
?

# Evolutionary Approaches for Studying Biological Pathways




*Escherichia Coli* *Bacillus subtilis* *Rhodobacter sphaeroides* *Helicobacter pylori*

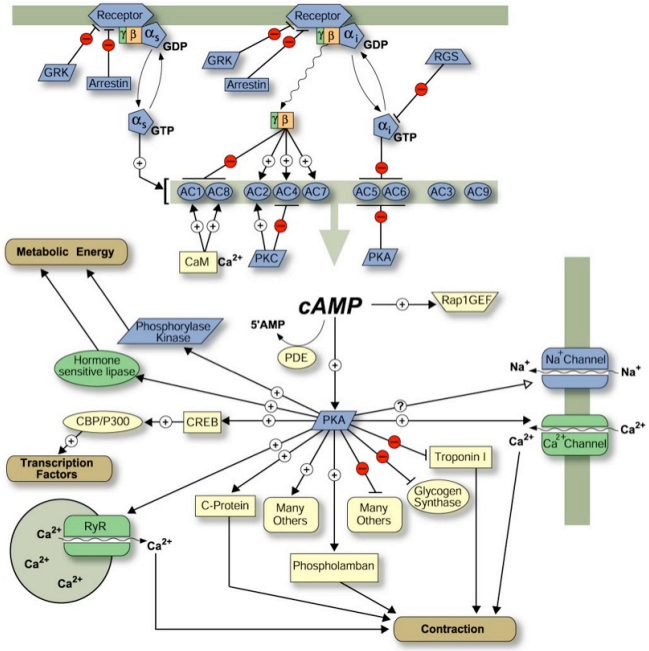


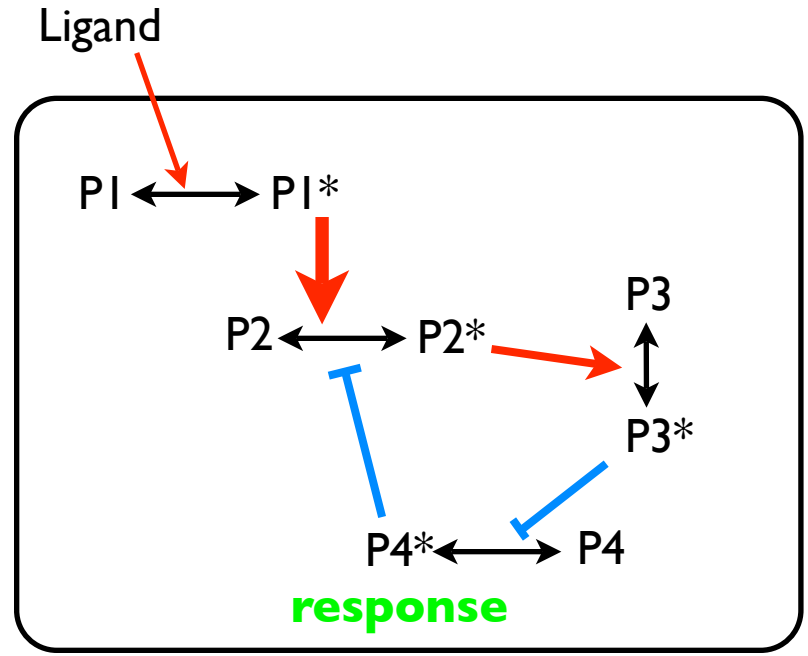




? ? ?

? ? ?

# Abstract Models for Studying Biological Pathways



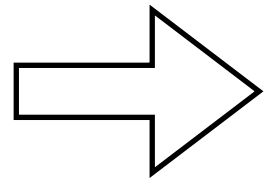

$$\frac{\mathrm{d}[P_i]}{\mathrm{d}t} = \left[[P_i^*] \cdot \sum_j l_{ij} \cdot [P_j^*]\right] - \left[[P_i] \cdot \left(\delta_{i1} \cdot [L] + \sum_j k_{ij} \cdot [P_j^*]\right)\right]$$

| | Ligand | P1* | P2* | P3* | P4* |
|---|---|---|---|---|---|
| P1↔P1* | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| P2↔P2* | 0.0 | 0.9 | 0.0 | 0.0 | -0.5 |
| P3↔P3* | 0.0 | 0.0 | 0.7 | 0.0 | 0.0 |
| P4↔P4* | 0.0 | 0.0 | 0.0 | -0.7 | 0.0 |

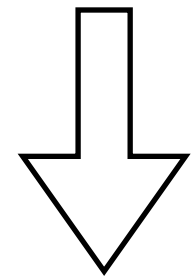
**A Generic Pathway Model**

# Putting The Two Together: *In Silico* Evolution

| | Ligand | P1* | P2* | P3* | P4* |
|---|---|---|---|---|---|
| P1↔P1* | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| P2↔P2* | 0.0 | 0.9 | 0.0 | 0.0 | -0.5 |
| P3↔P3* | 0.0 | 0.0 | 0.7 | 0.0 | 0.0 |
| P4↔P4* | 0.0 | 0.0 | 0.0 | -0.7 | 0.0 |

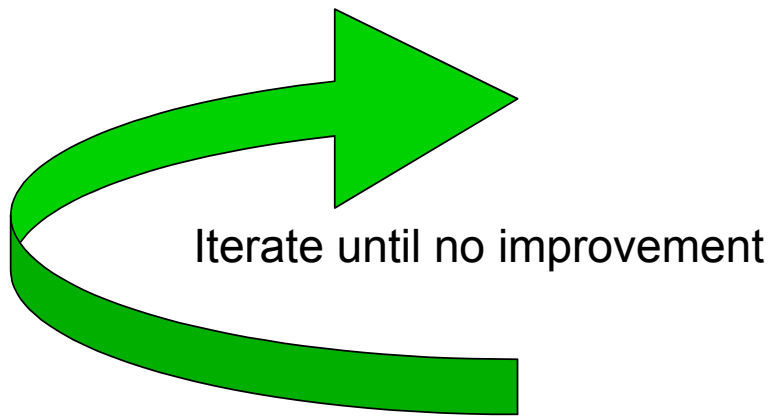


Evaluate Response



Introduce Mutations

| | Ligand | P1* | P2* | P3* | P4* |
|---|---|---|---|---|---|
| P1↔P1* | 0.5 | 0.0 | 0.0 | 0.3 | 0.0 |
| P2↔P2* | 0.0 | 0.9 | 0.0 | 0.0 | -0.5 |
| P3↔P3* | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 |
| P4↔P4* | 0.0 | 0.0 | 0.0 | -0.1 | 0.0 |

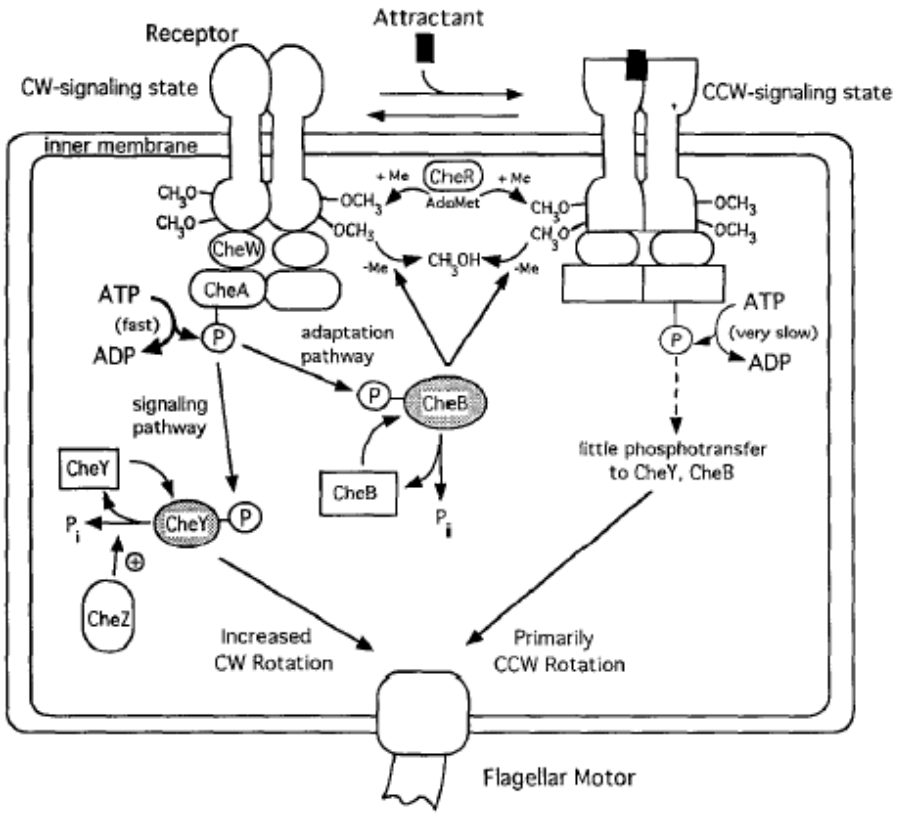


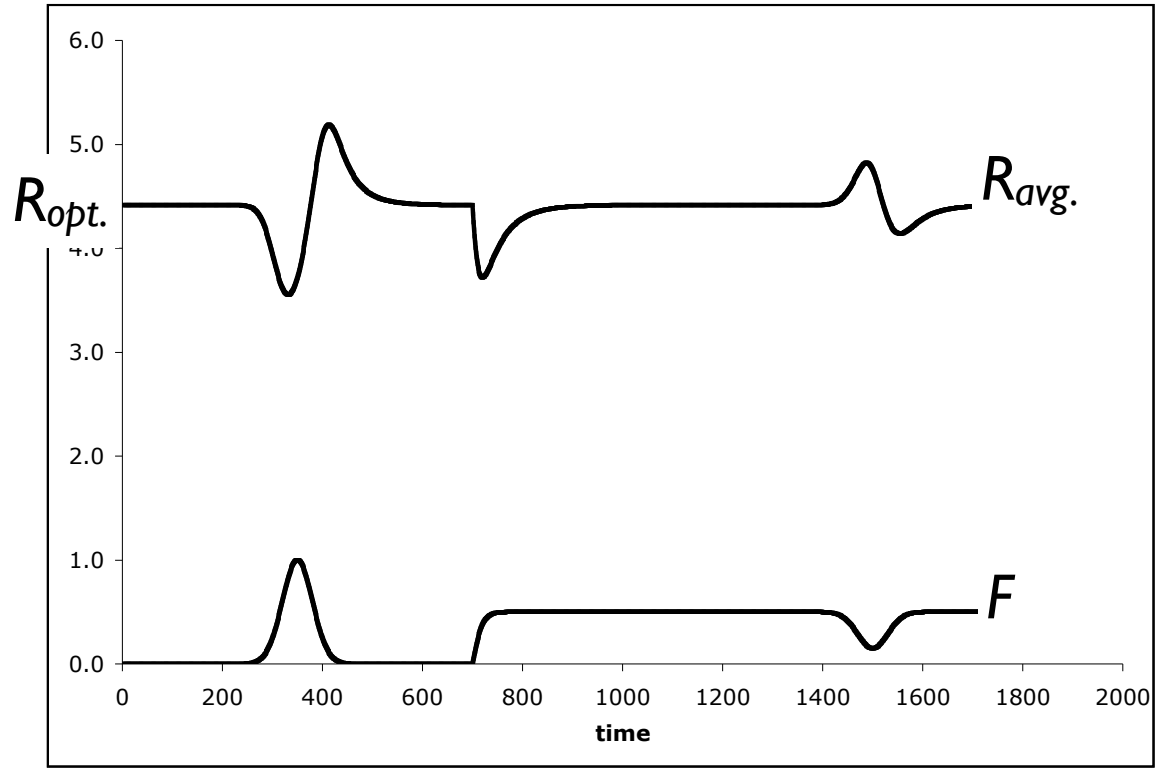
Iterate until no improvement

**Analysis of pathway evolution using generic, extendable pathway models**

# Fitness: Capturing Bacterial Chemotaxis

Chemotaxis is achieved via a **derivative-like** response:



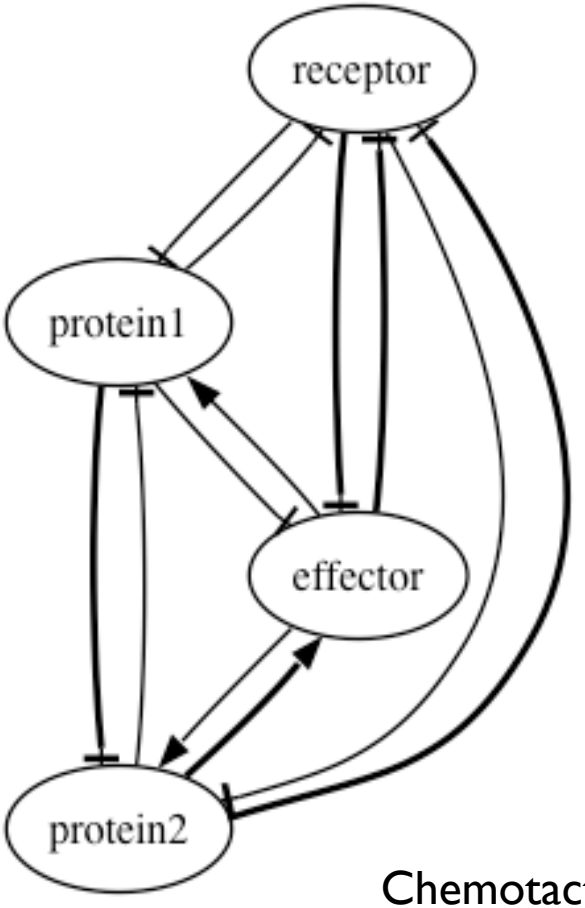

Rao CV et. al., PLOS Biology, 2(2), 2004

$$\text{Chemotactic Ability} = \sum_{t} \left[ C \cdot \left[ \Delta F \cdot (R_{opt.} - R_{avg.}) \right] - \left[ (R_{opt.} - R_{avg.})^2 \right] \right].$$

Reward food intake over a time interval

Penalize deviation from a base (optimum) response

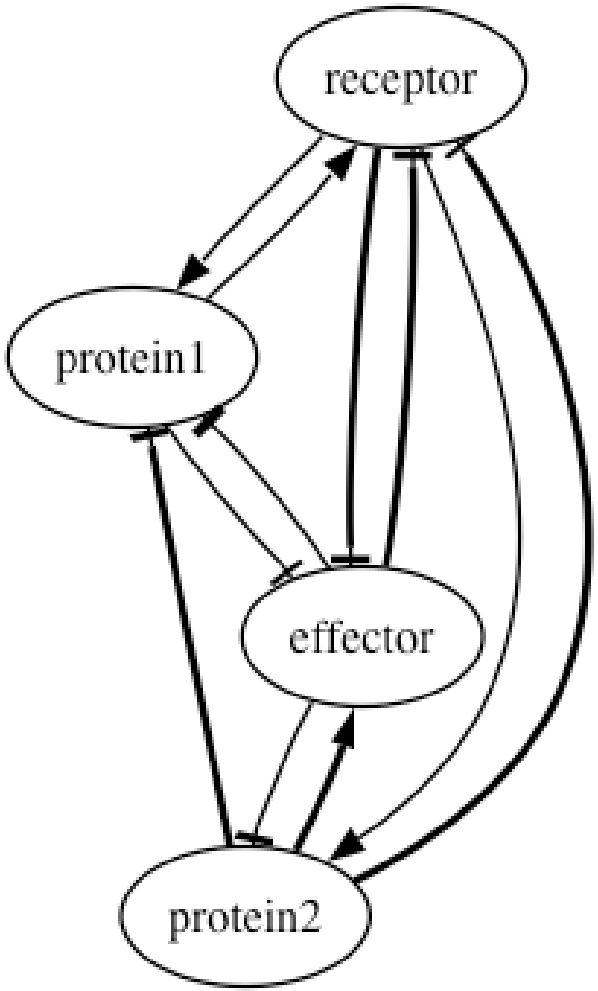
# Evolving Chemotaxis Pathways




Introduce Mutations

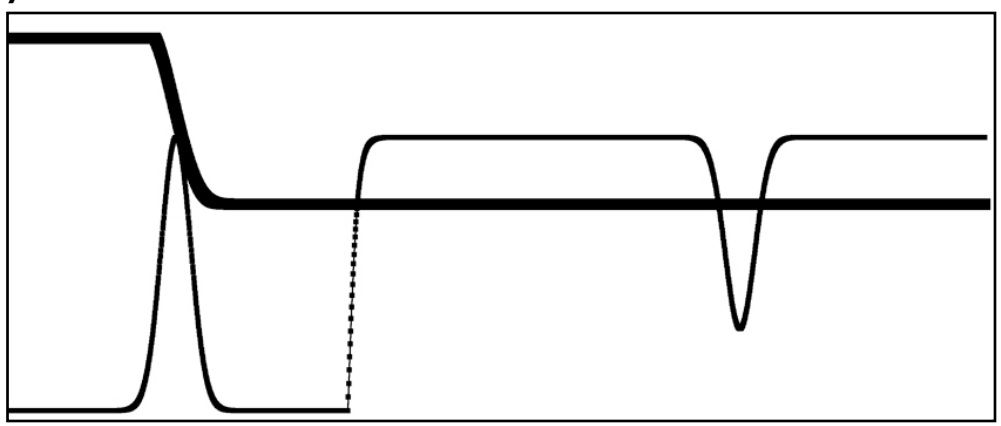
Iterate Until No Improvement

Evaluate Response




Chemotactic Ability $= \sum_t \left[ C \cdot \left[ \Delta F \cdot (R_{opt.} - R_{avg.}) \right] - \left[ (R_{opt.} - R_{avg.})^2 \right] \right].$

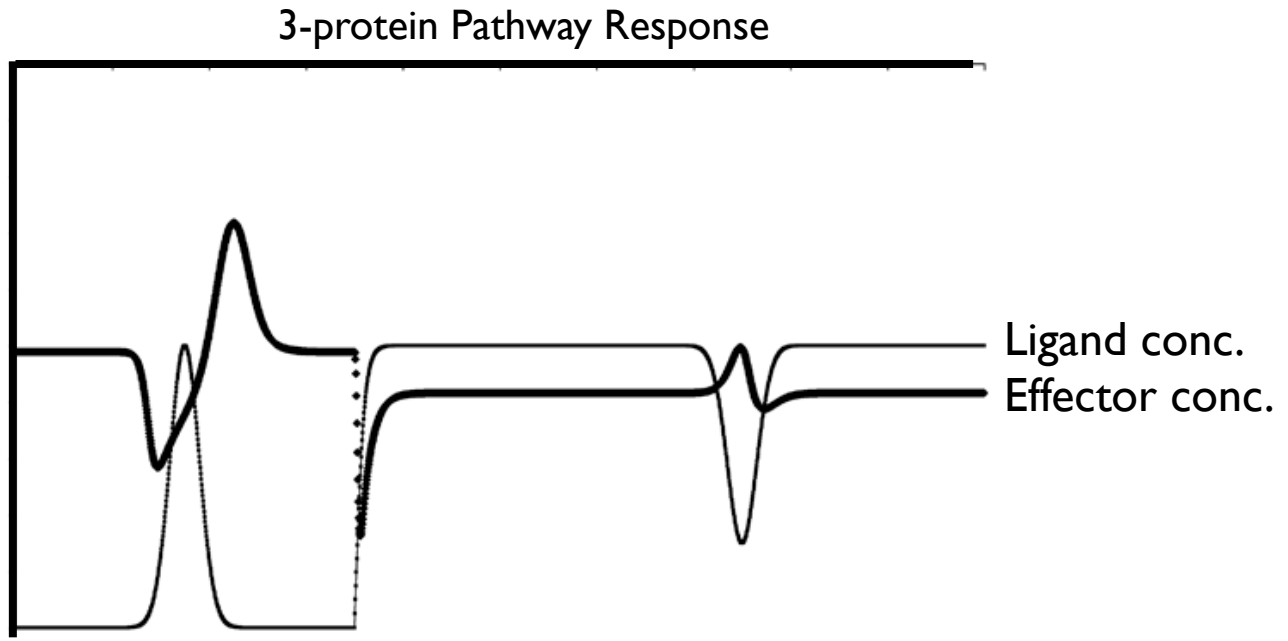
~-50

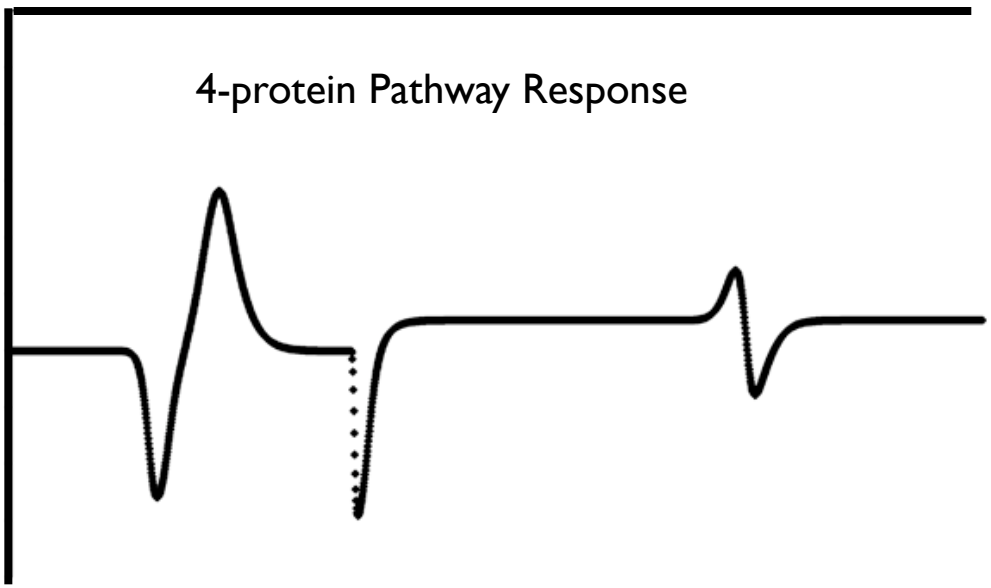


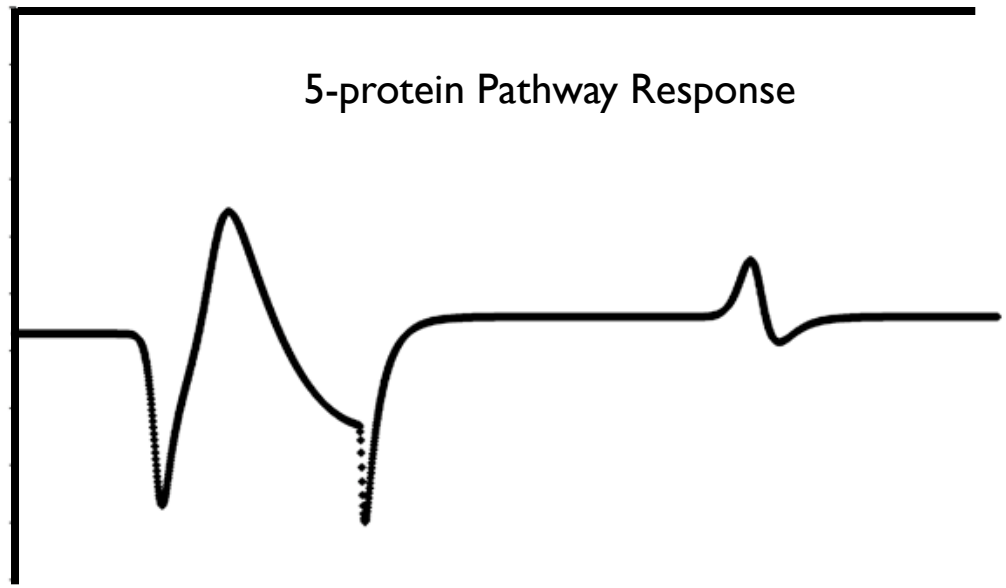
Ligand conc.
Effector conc.

# Evolved Pathways

Are able to give a derivative-like response:

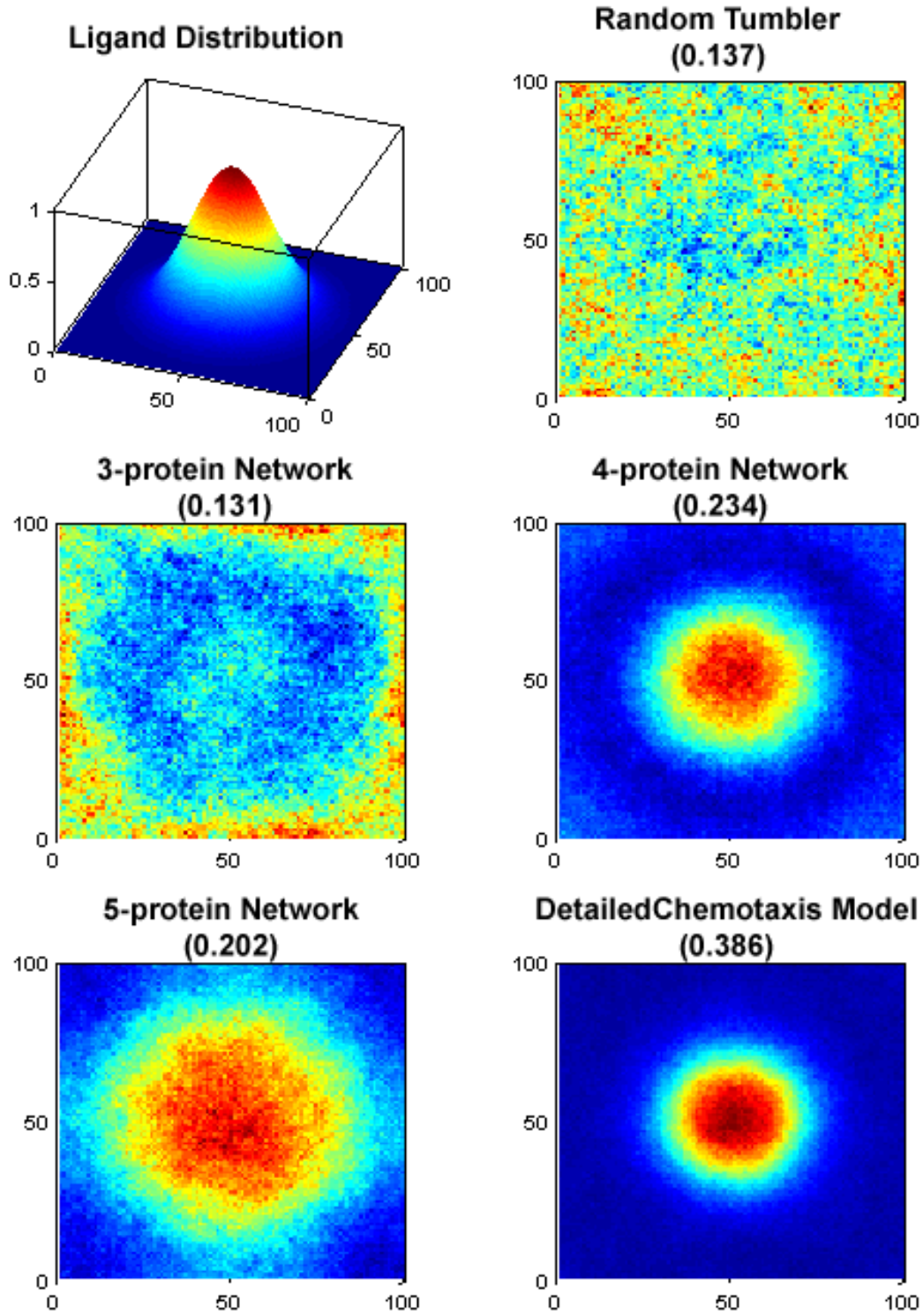


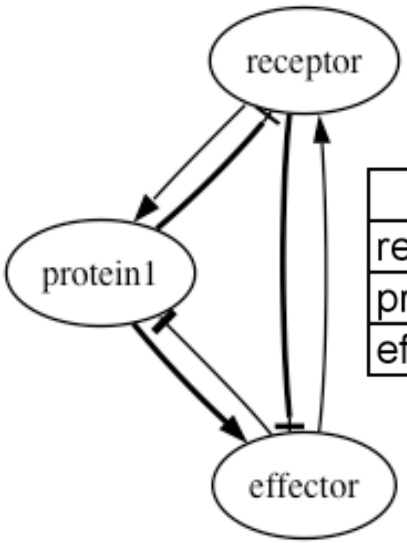



# Evolved Pathways

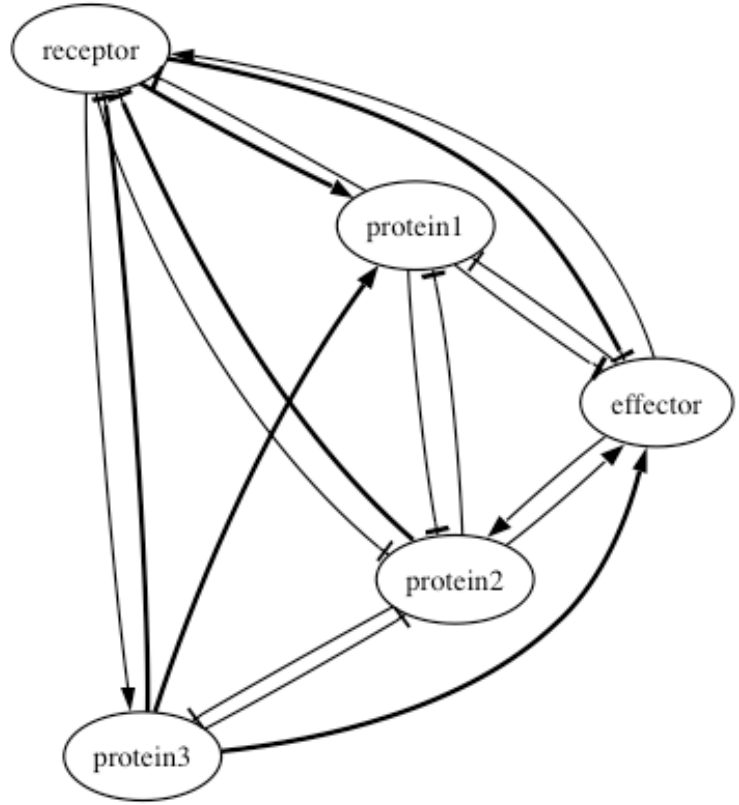
Are able to mediate chemotaxis-like behavior:



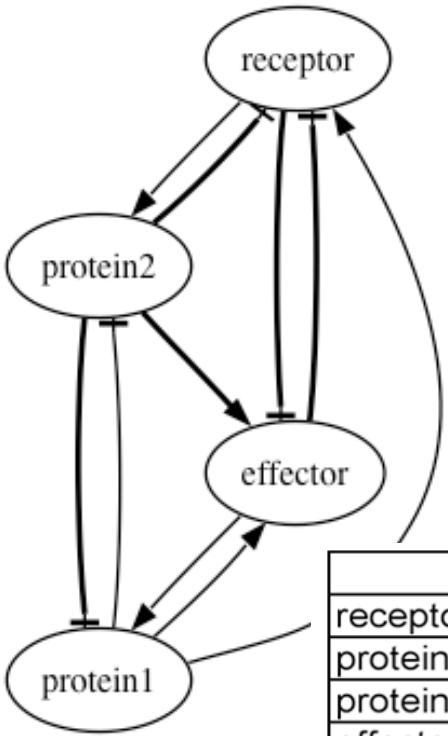
# Evolved Pathways




| | receptor | protein 1 | effector |
|---|---|---|---|
| receptor | 0.000 | -0.986 | 0.007 |
| protein 1 | 0.020 | 0.000 | -0.040 |
| effector | -0.733 | 0.726 | 0.000 |

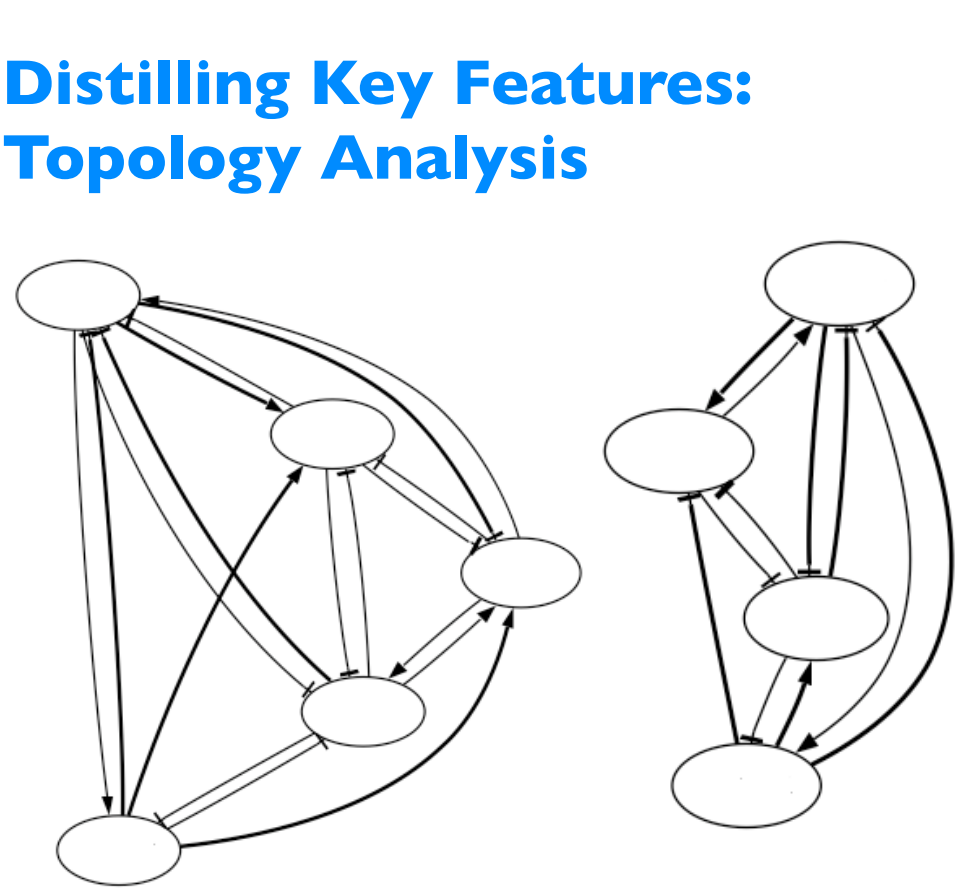


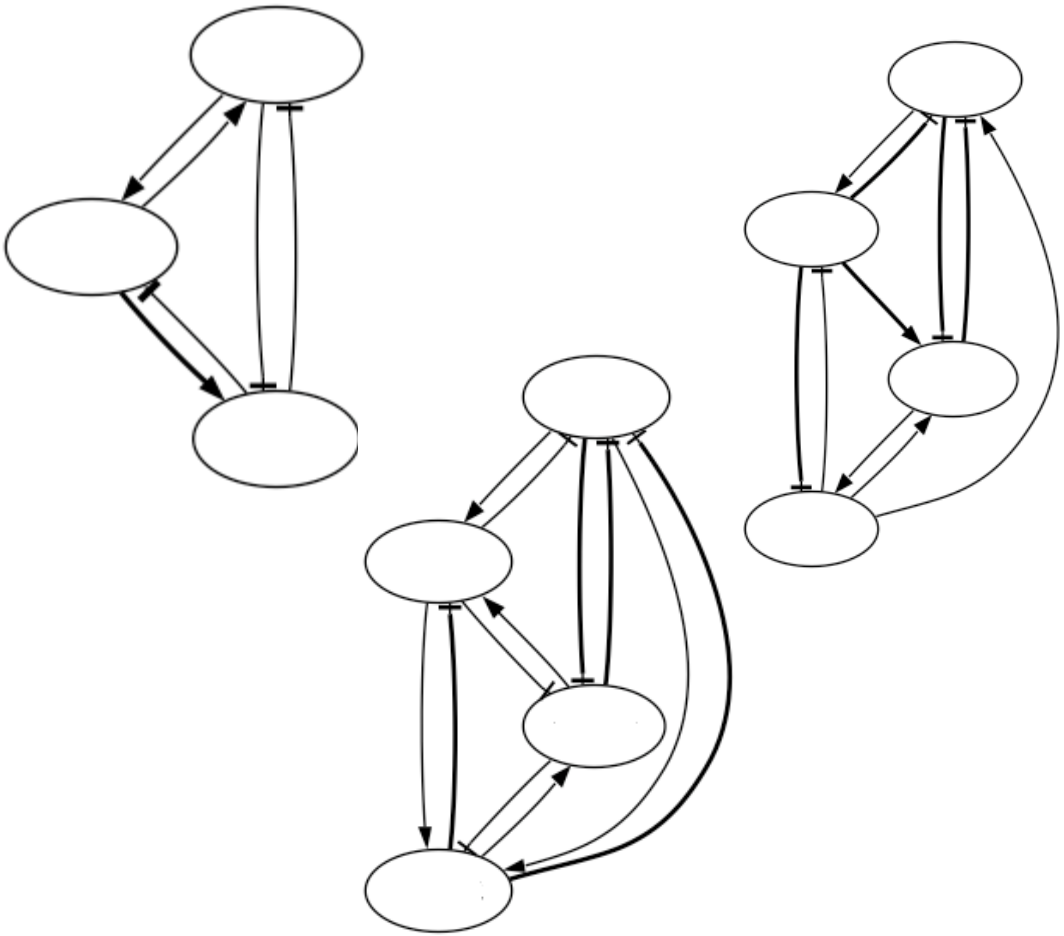

| | receptor | protein 1 | protein 2 | protein 3 | effector |
|---|---|---|---|---|---|
| receptor | 0.000 | -0.137 | -0.990 | -0.984 | 0.010 |
| protein 1 | 0.574 | 0.000 | -0.070 | 0.618 | -0.050 |
| protein 2 | -0.303 | -0.180 | 0.000 | -0.086 | 0.150 |
| protein 3 | 0.030 | 0.000 | -0.040 | 0.000 | 0.000 |
| effector | -0.995 | -0.143 | 0.010 | 0.679 | 0.000 |



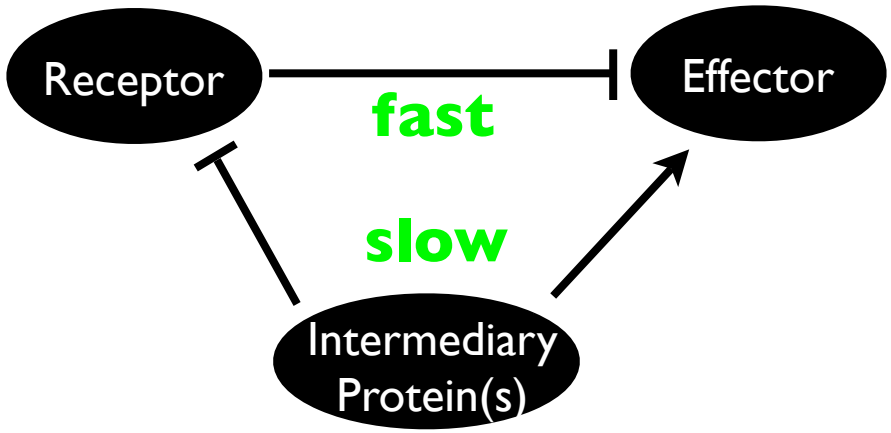

| | receptor | protein 1 | protein 2 | effector |
|---|---|---|---|---|
| receptor | 0.000 | 0.021 | -0.934 | -0.991 |
| protein 1 | 0.000 | 0.000 | -0.981 | 0.060 |
| protein 2 | 0.028 | -0.290 | 0.000 | 0.000 |
| effector | -0.981 | 0.027 | 0.979 | 0.000 |

# Distilling Key Features: Topology Analysis





Important Interactions

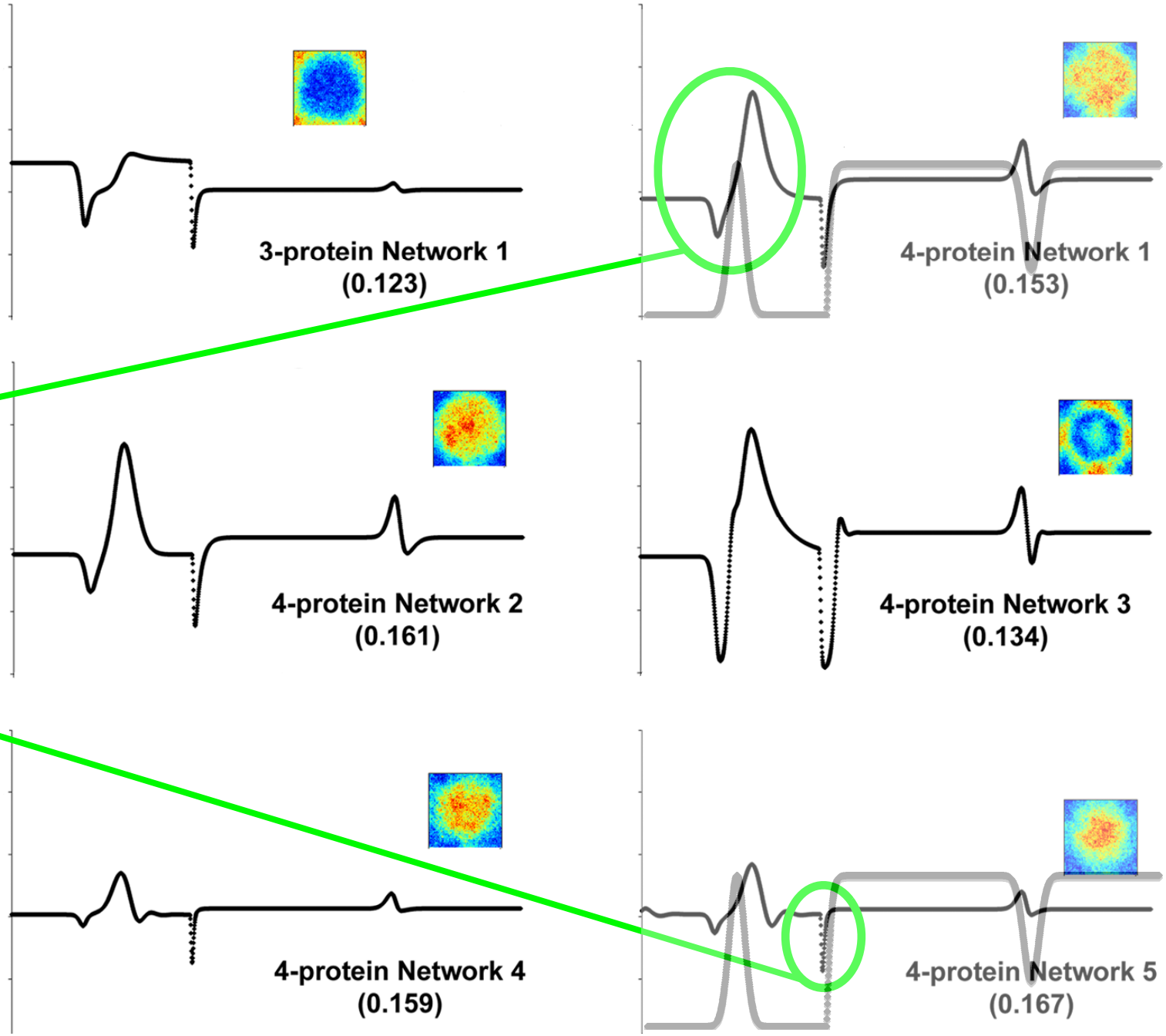



Important Dynamical Features

Parameters For 3-protein Network

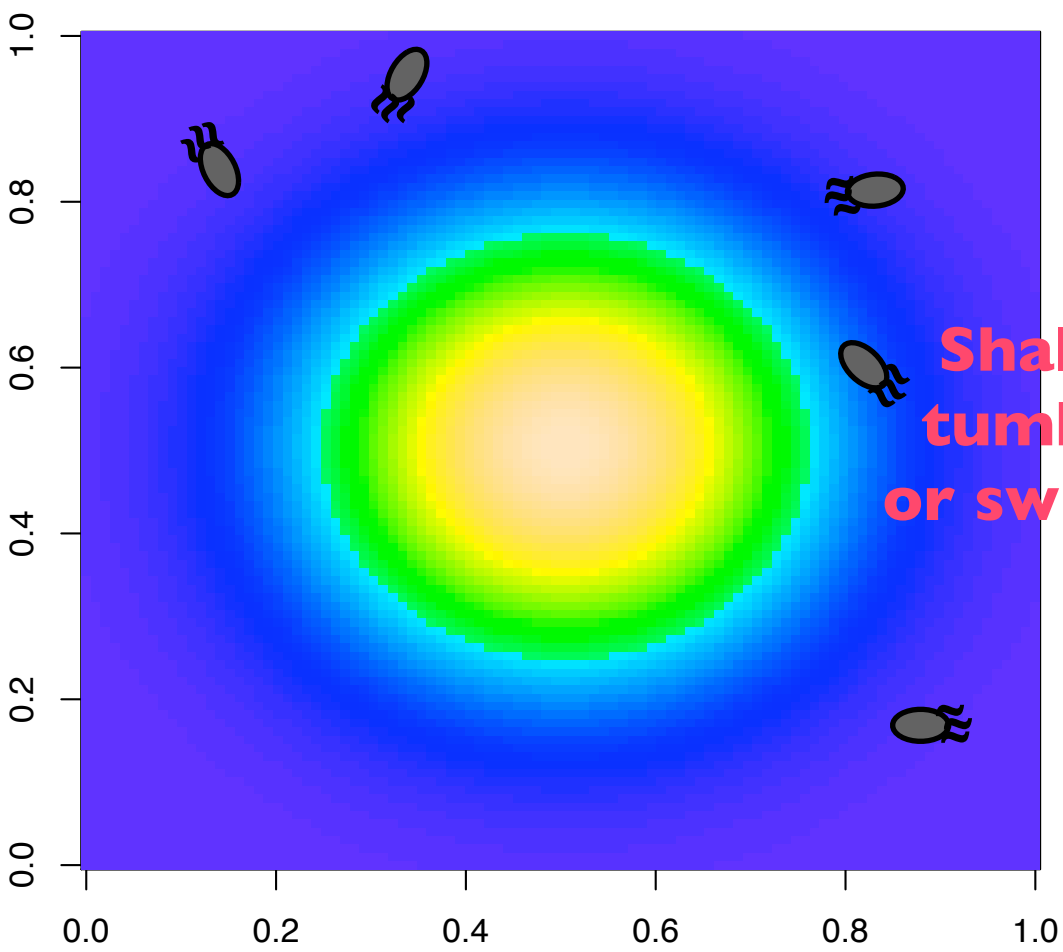
| | receptor | protein 1 | effector |
|---|---|---|---|
| receptor | 0.000 | -0.986 | 0.007 |
| protein 1 | 0.020 | 0.000 | -0.040 |
| effector | -0.733 | 0.726 | 0.000 |

# Distilling Key Features: Response Analysis

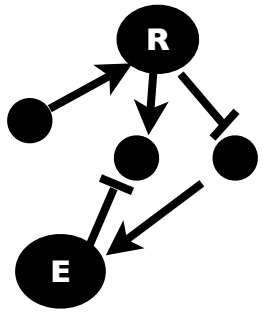


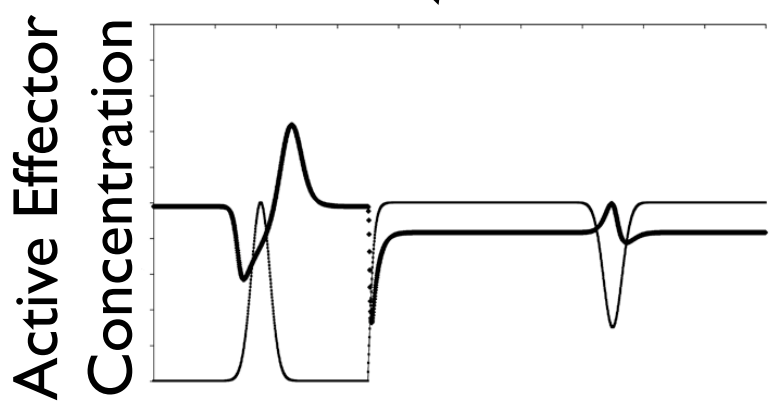
# Evolutionary Systems Biology: Chemotaxis

## Pathway => Dynamics => Behavior



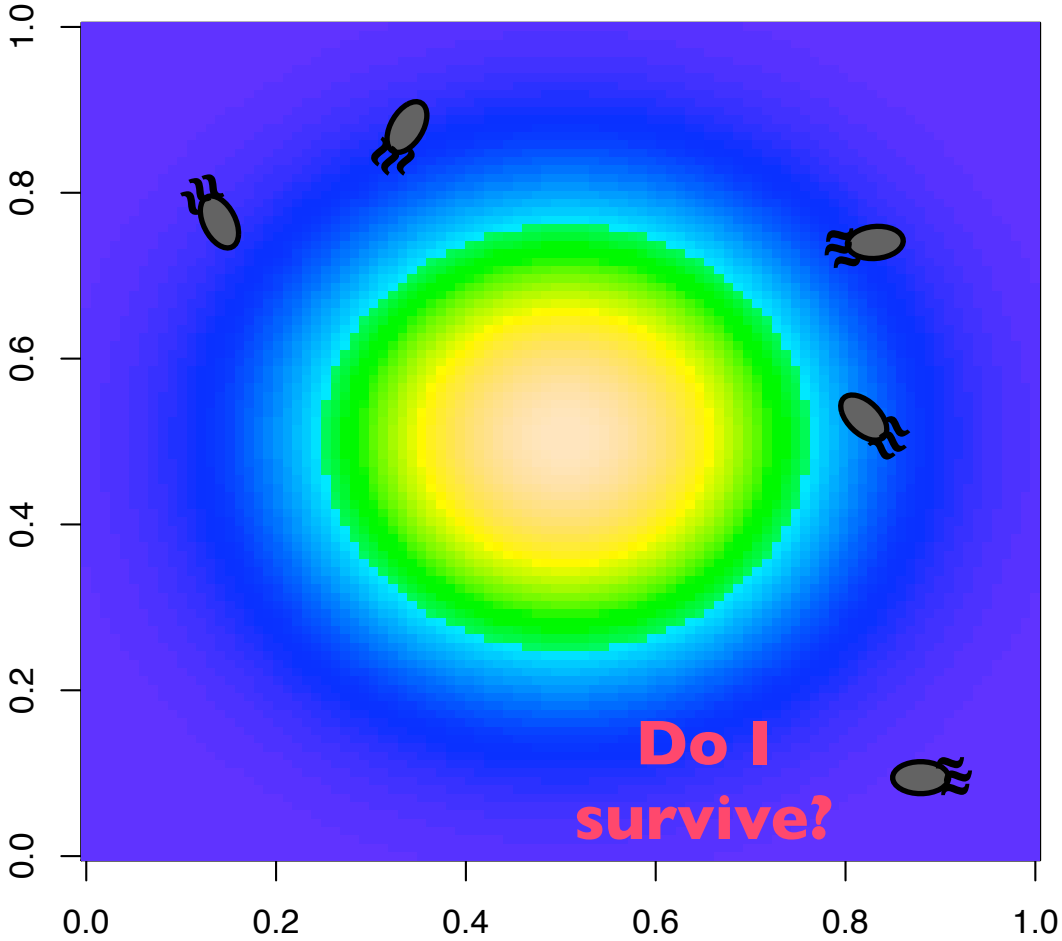

Shall I tumble or swim??



# Evolutionary Systems Biology: Chemotaxis

## Behavior => Fitness => Evolution



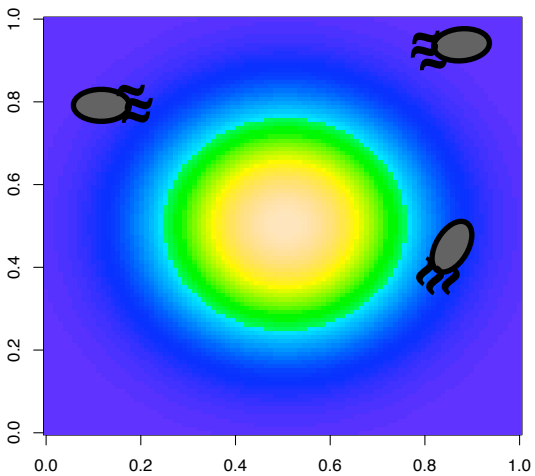

Can pathways evolve that allow bacteria to find the food?

How would the structure/dynamics of these pathways change with environment/pathway constraints?

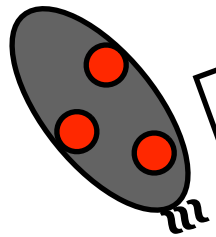
Do all bacteria evolve the same strategy (i.e. pathway structure)?

Do we find population wide variances?

# A more realistic evolutionary setup




Initial Population



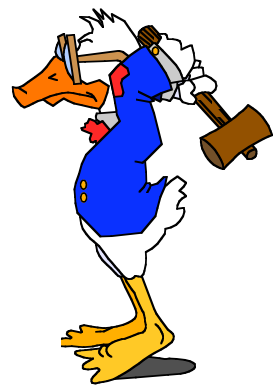
dynamics

$$\frac{d[P_i^*]}{dt} = \left(k_{ii} + \sum_{j \neq i} k_{ij}[P_j^*] + \delta_{i1} k_{1A}[A]\right)(1 - [P_i^*])$$

$$- \left(l_{ii} + \sum_{j \neq i} l_{ij}[P_j^*] + \delta_{i1} l_{1A}[A]\right)[P_i^*]$$

behavior

Time

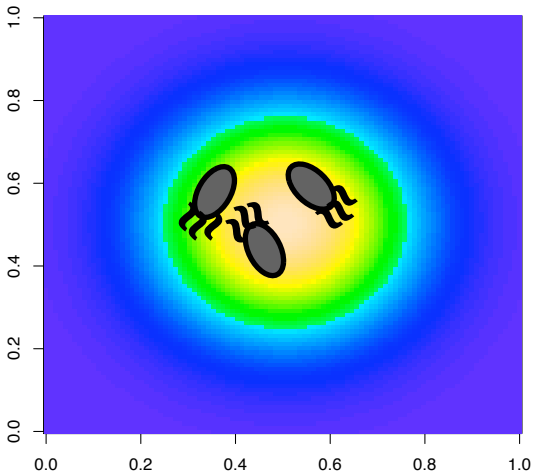


evolution

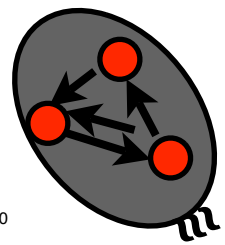
$$p_{\text{Tumble}} = \frac{\gamma_m [P_{N_p}^*]}{1 + \gamma_m [P_{N_p}^*]}$$

Selection

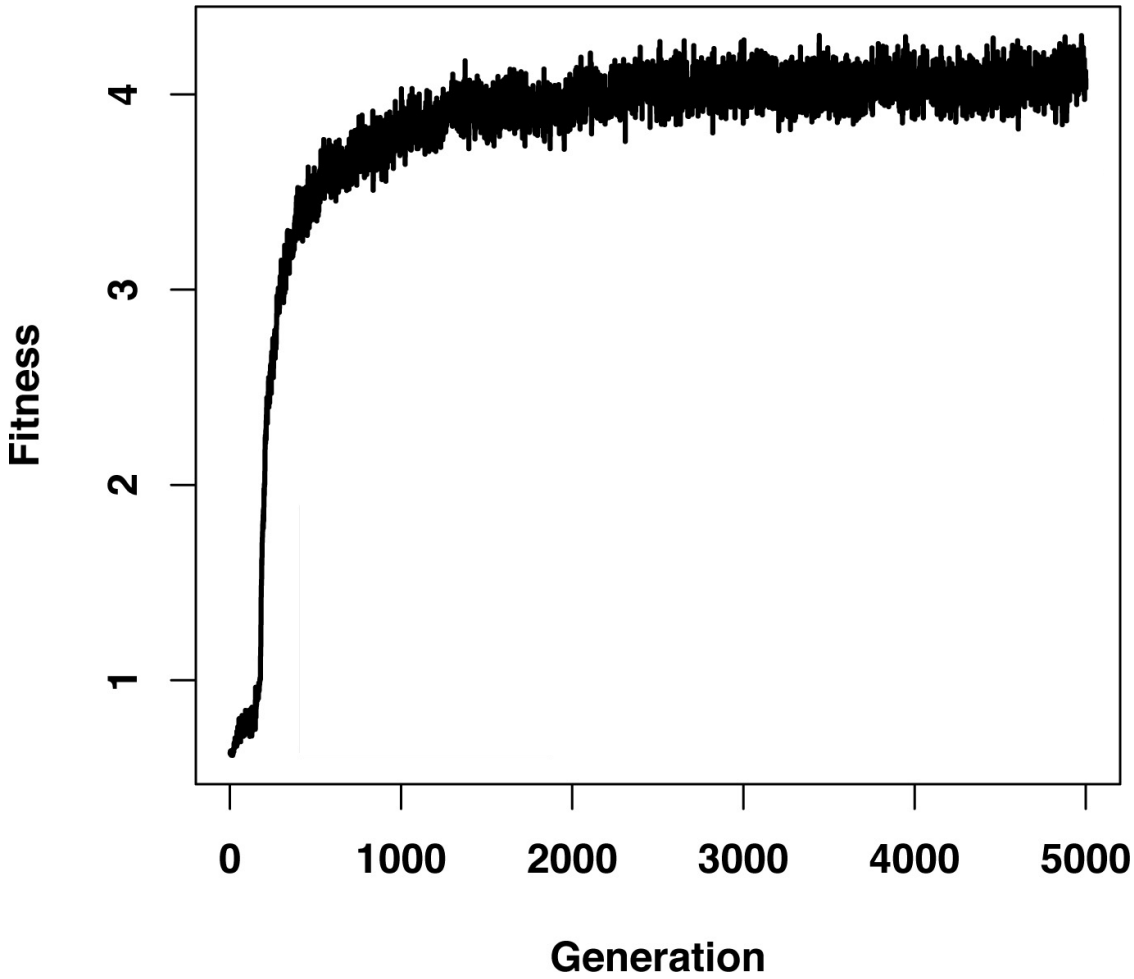
Fitness = Food Encountered

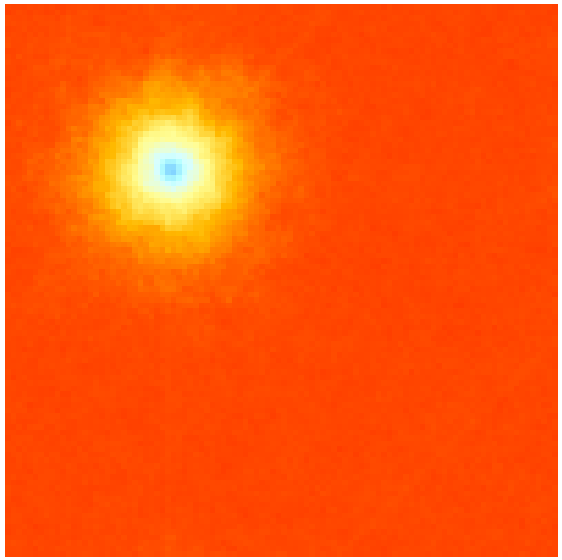



Final (evolved) Population

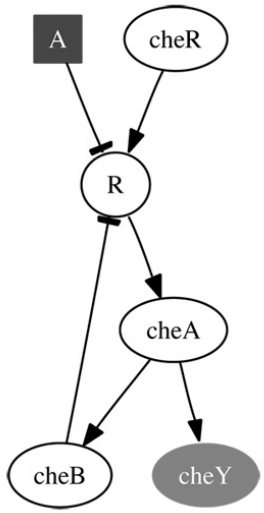


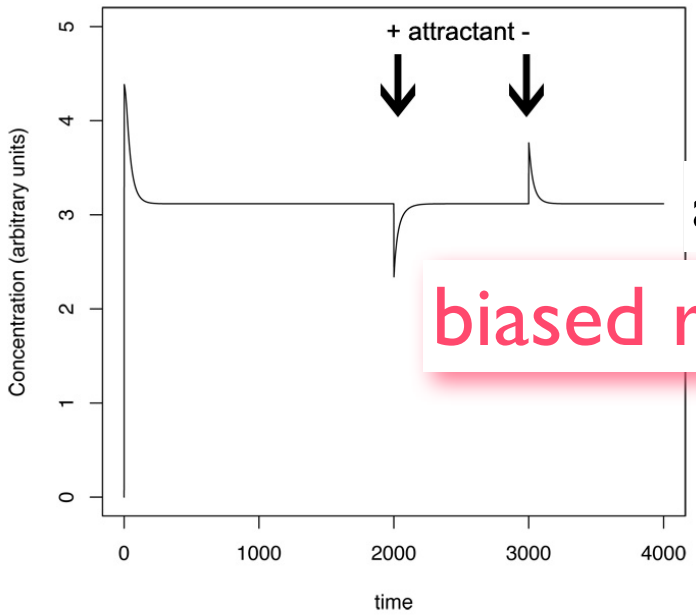
# Evolution Of Chemotaxis



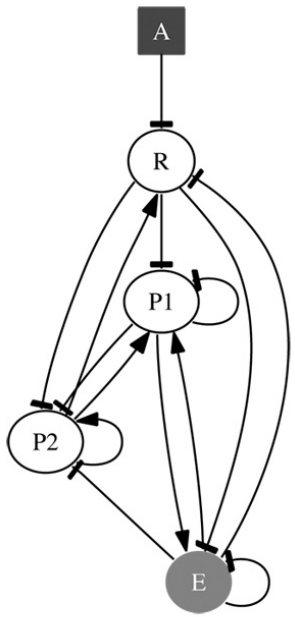
# Evolution Of Chemotaxis Pathways

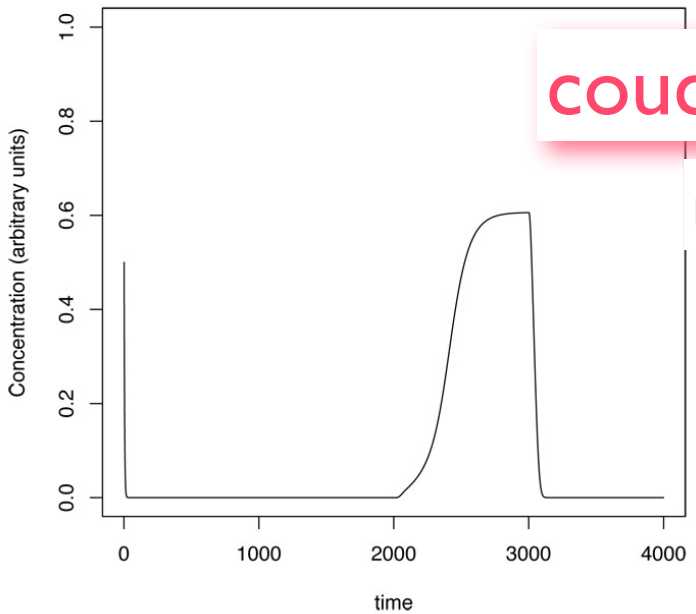



adaptive dynamics

biased random walk

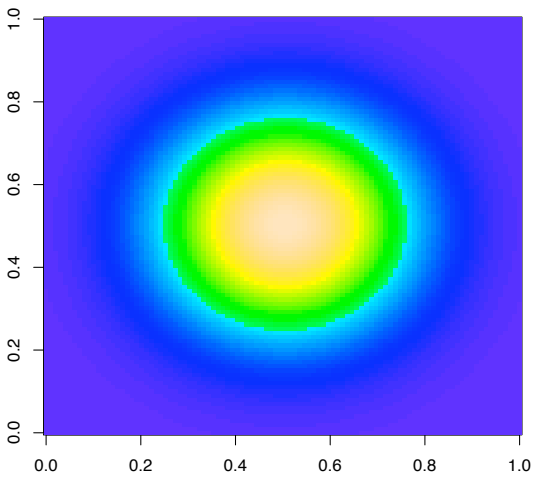


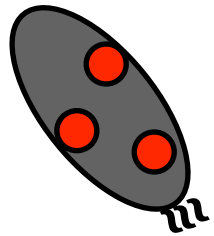

couch potato

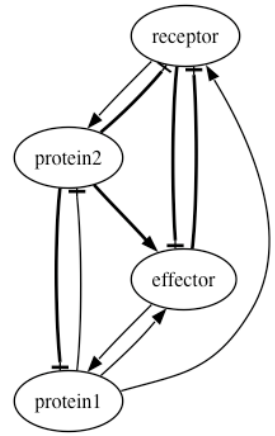
non - adaptive dynamics

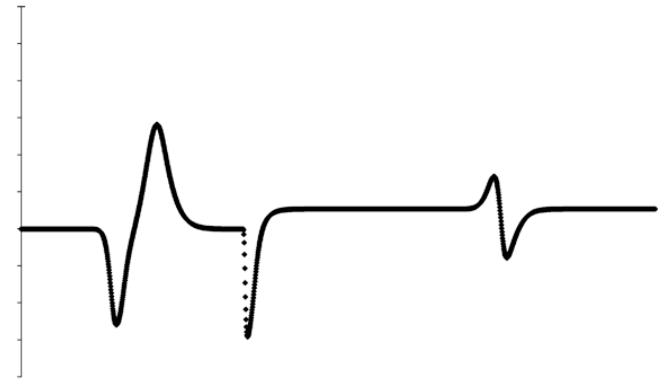
# Non - Adaptive Chemotaxis is not due to



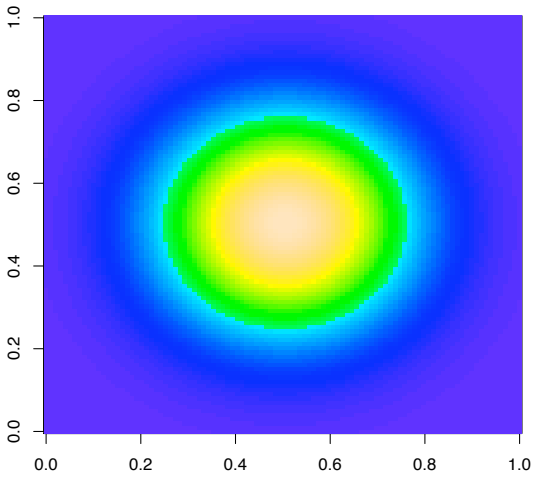

Initial Population



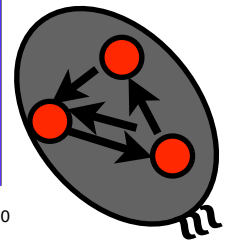


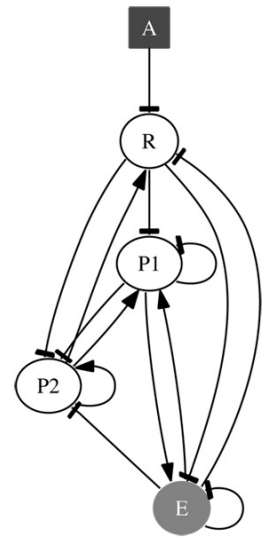



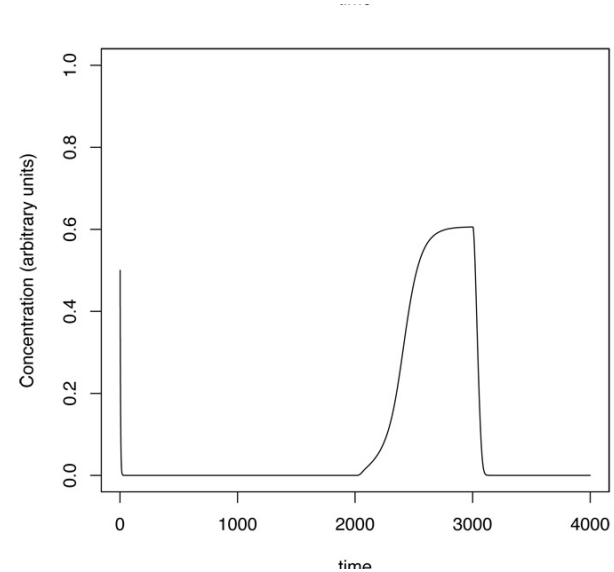
Initial Population with adaptive dynamics




Final (evolved) Population

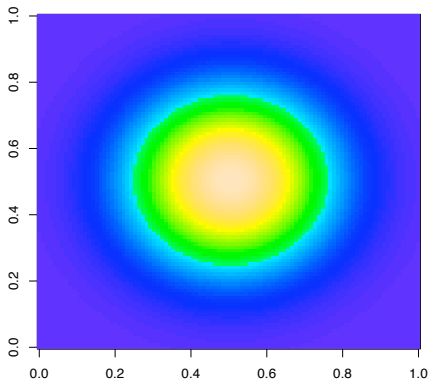


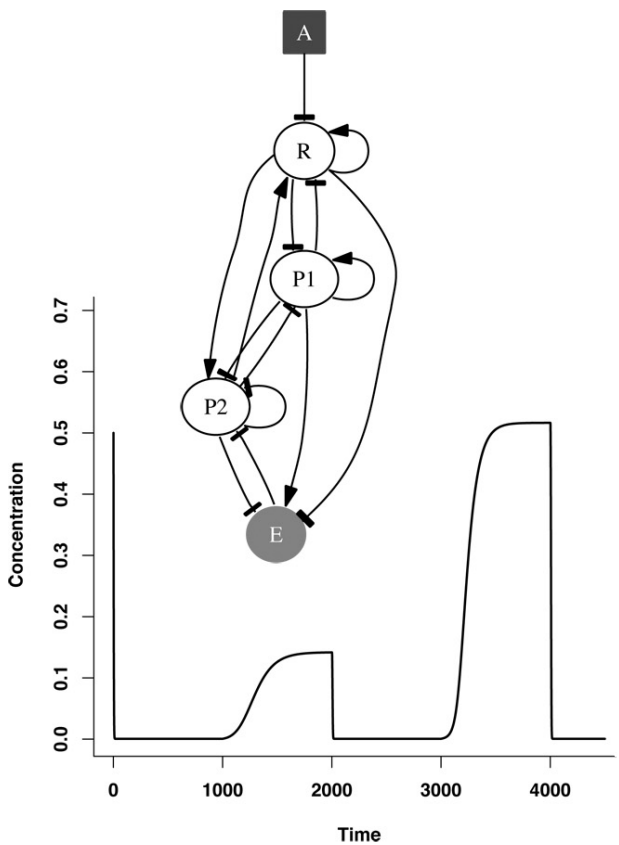


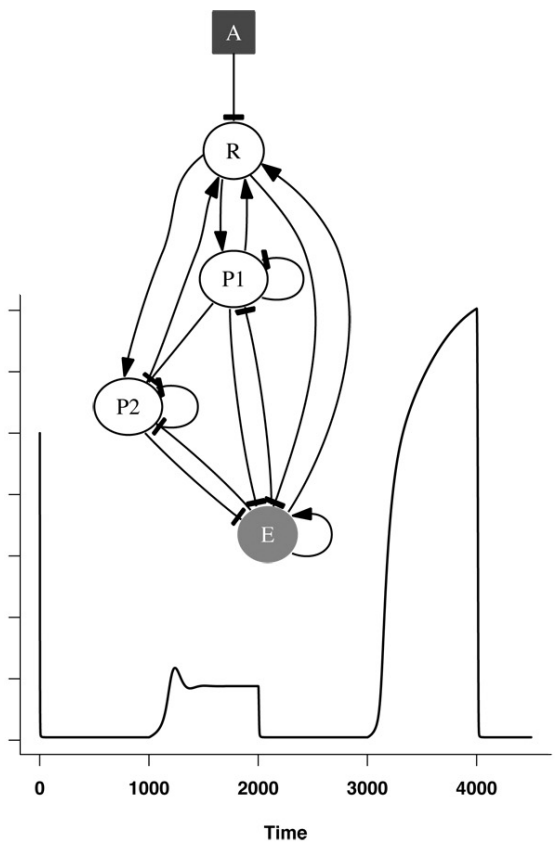



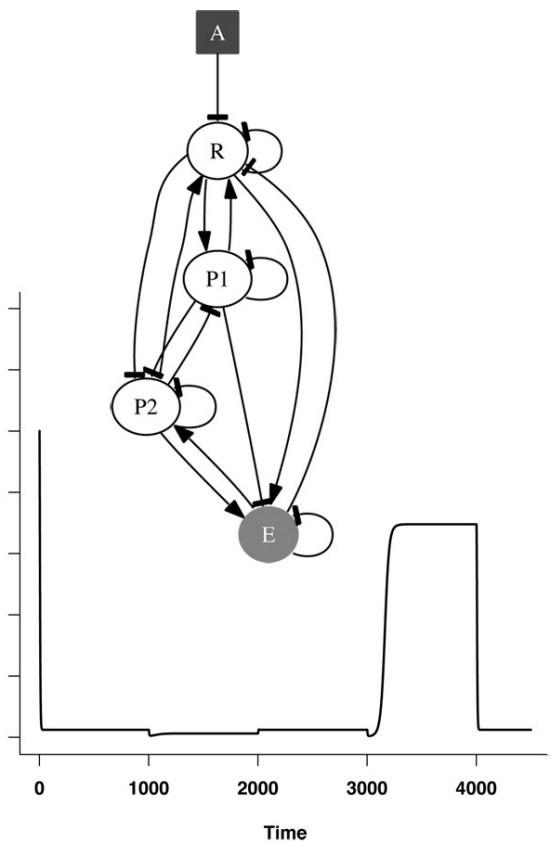

Final population with non-adaptive dynamics

# Non - Adaptive chemotaxis under fluctuating environments

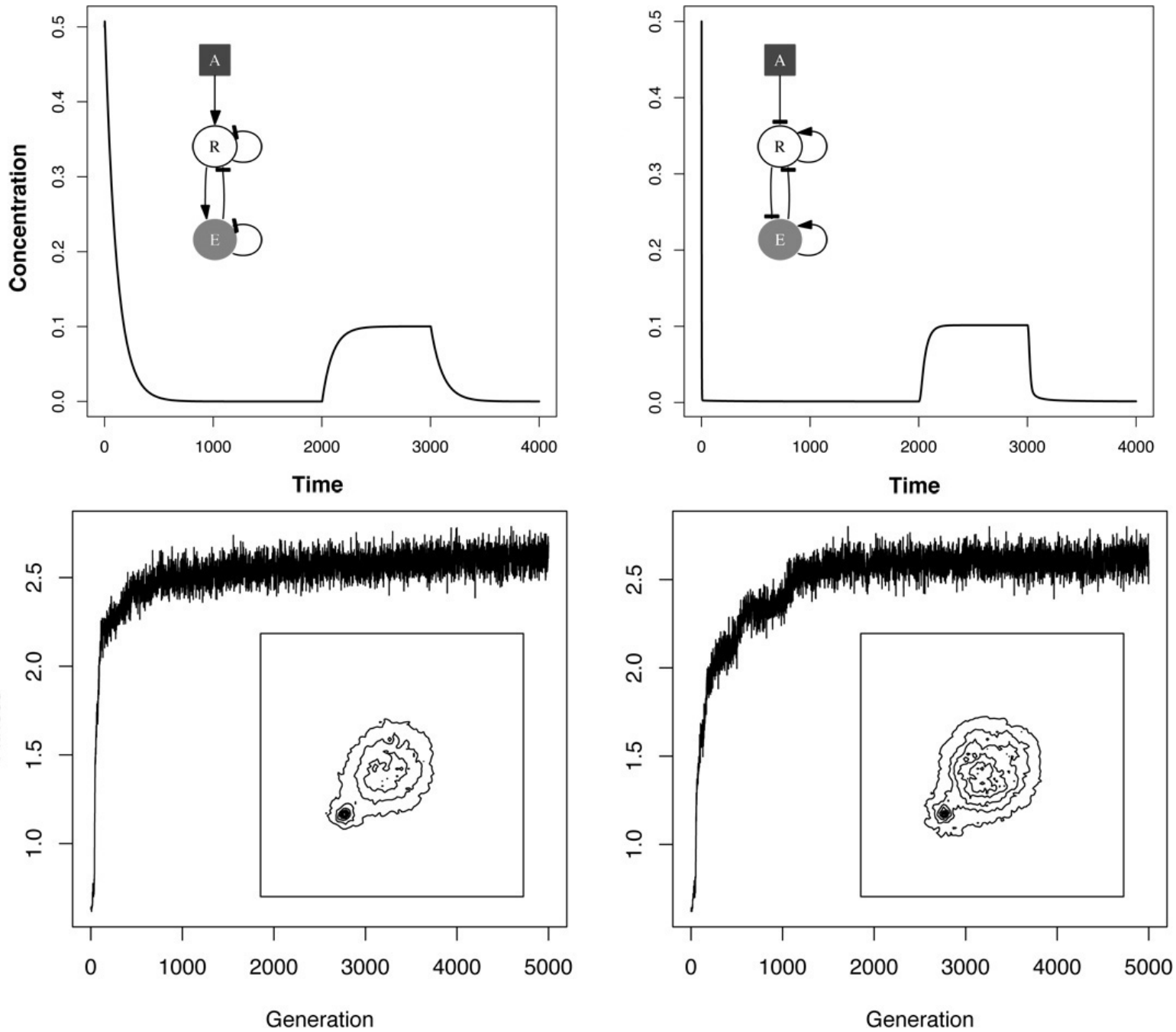


# Minimal Non-Adaptive Chemotaxis Mechanisms



# Reality or Modeling Curiosity ?

Chemotaxis in absence of adaptation has already been observed in nature in *Rhodobacter sphaeroides* and in mutant strains of *Escherichia coli*.

Poole PS and Armitage JP, J. Bacteriology, 170, 1988
Barak R and Eisenbach M, Mol. Microbiology, 31, 1999

**Pathways with non-adaptive dynamics**

- possible existence in many bacterial species

- a simple mechanism to couple metabolic and/or other signals to conventional chemotaxis

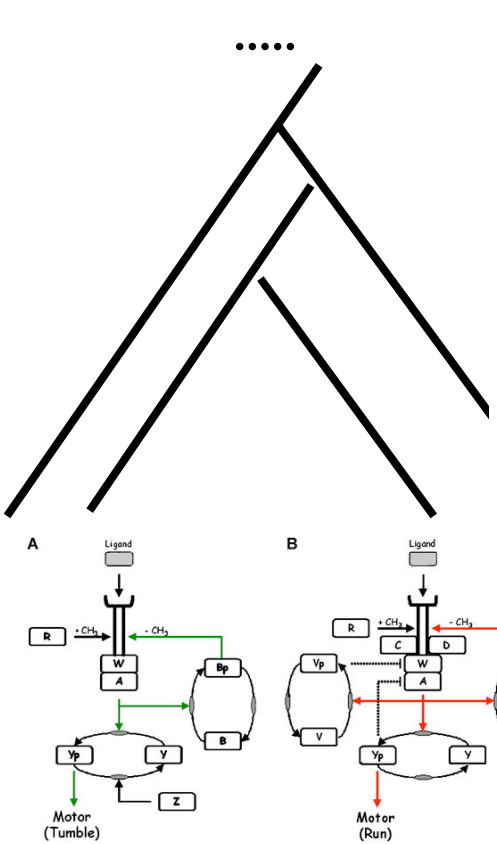
- evolutionary origins of chemotaxis

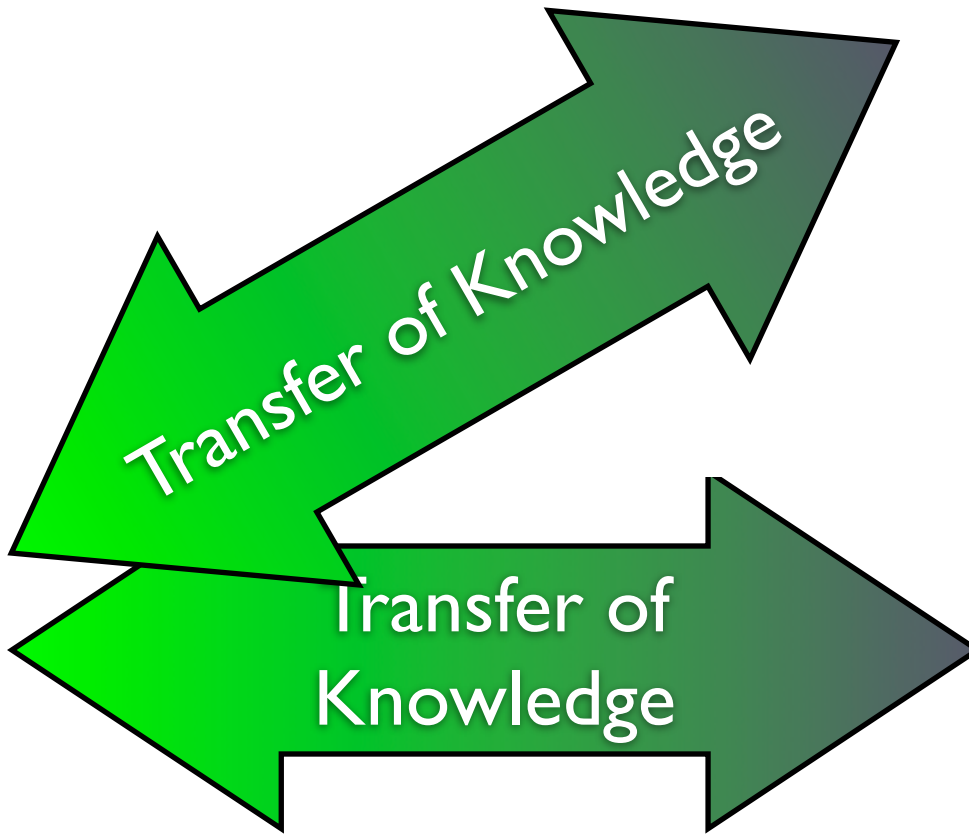
# Insights from Evolutionary Systems Biology

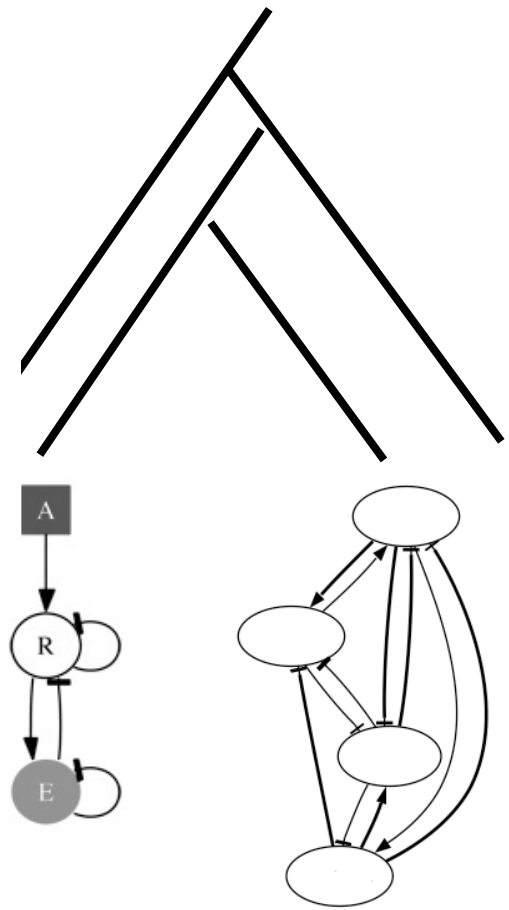
**Natural Evolution**

How do evolutionary processes affect pathway properties?

***In Silico* Evolution**



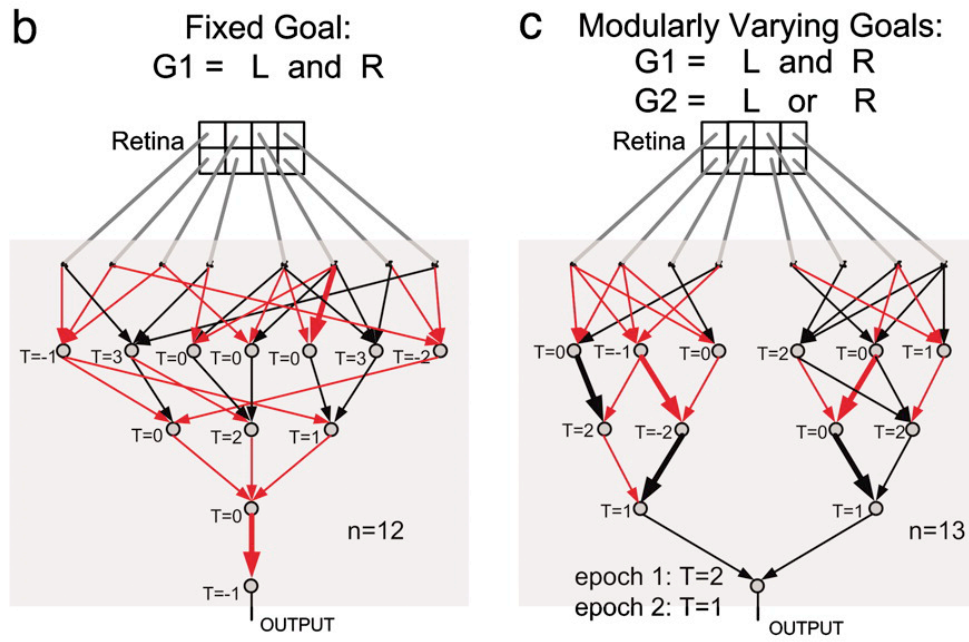

Soyer OS, Pfeiffer T, Bonhoeffer S, JTB, 2006, 241(2)
Goldstein RA, Soyer OS, *submitted*

# Pathway Modularity

## How does modularity arise in biological systems and how is it maintained?

Adaptation to alternating environments...

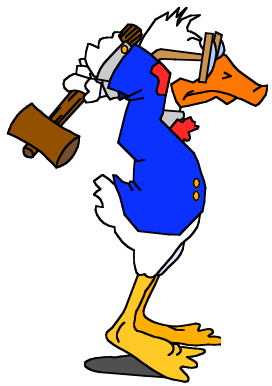
Kashtan, N. & Alon, U. (2005) *PNAS* 102, 13773-8.

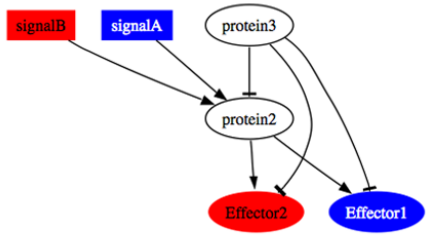



Selection for evolvability...

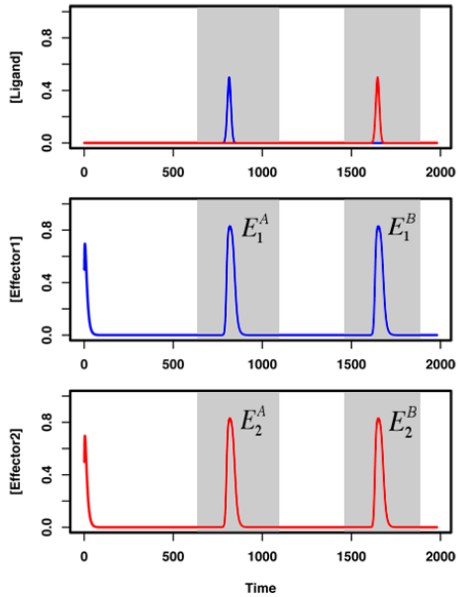
Kirschner, M. & Gerhart, J. (1998) *PNAS* 95, 8420-7.

# Evolution of Modularity



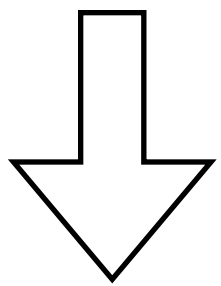



| Protein2 | Protein3 | Effector1 | Effector2 |
|---|---|---|---|
| 0.0 | -0.5 | 0.0 | 0.0 |
| 0.0 | 0.0 | 0.0 | 0.0 |
| 1.0 | -0.2 | 0.0 | 0.0 |
| 1.0 | -0.2 | 0.0 | 0.0 |

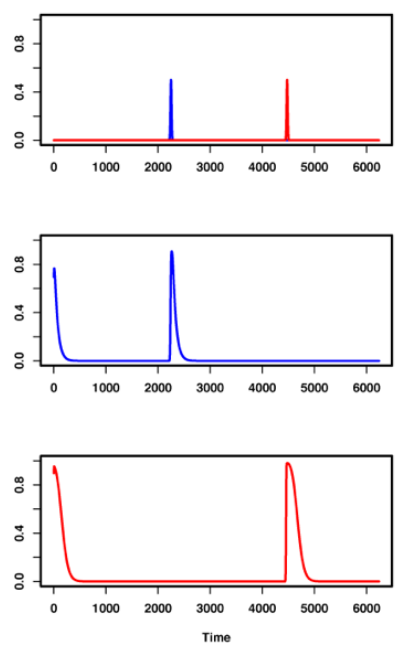



Selection

Fitness = "Ability to mediate separate responses"

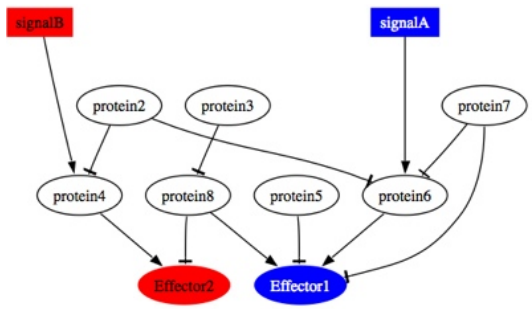


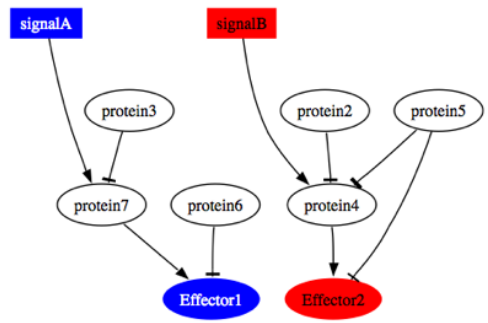
**Through simple evolutionary processes....**



# Two responses mediated by different structures

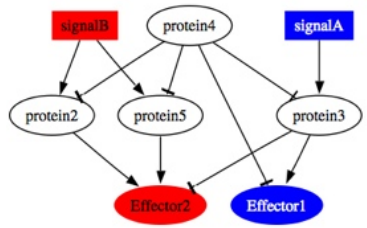
**modular**

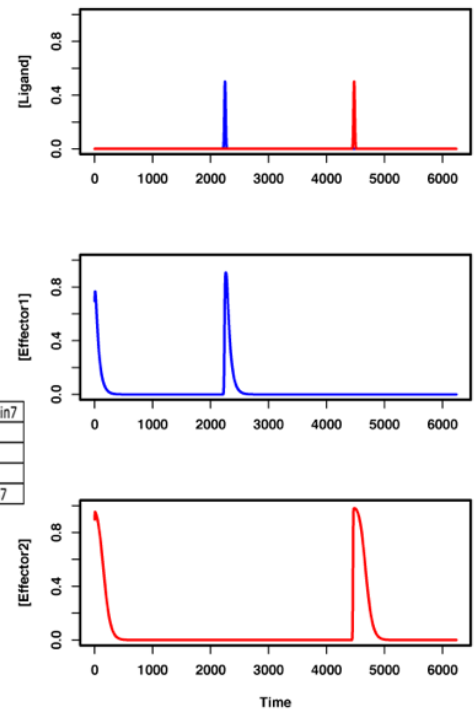


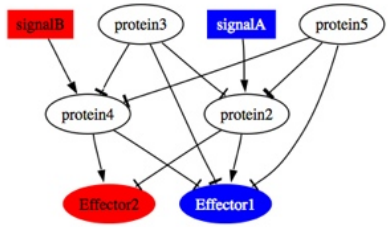

| | SignalA | SignalB | Protein2 | Protein3 | Protein4 | Protein5 | Protein6 | Protein7 |
|---|---|---|---|---|---|---|---|---|
| Protein4 | | 0.72 | -0.03 | | | -0.02 | | |
| Protein7 | 0.25 | | | -0.21 | | | | |
| Effector1 | | | | | 1.00 | -0.40 | | |
| Effector2 | | | | | | | -0.03 | 0.37 |

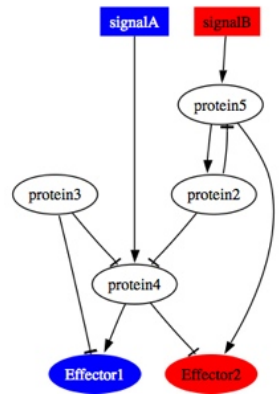
**cross-talk**



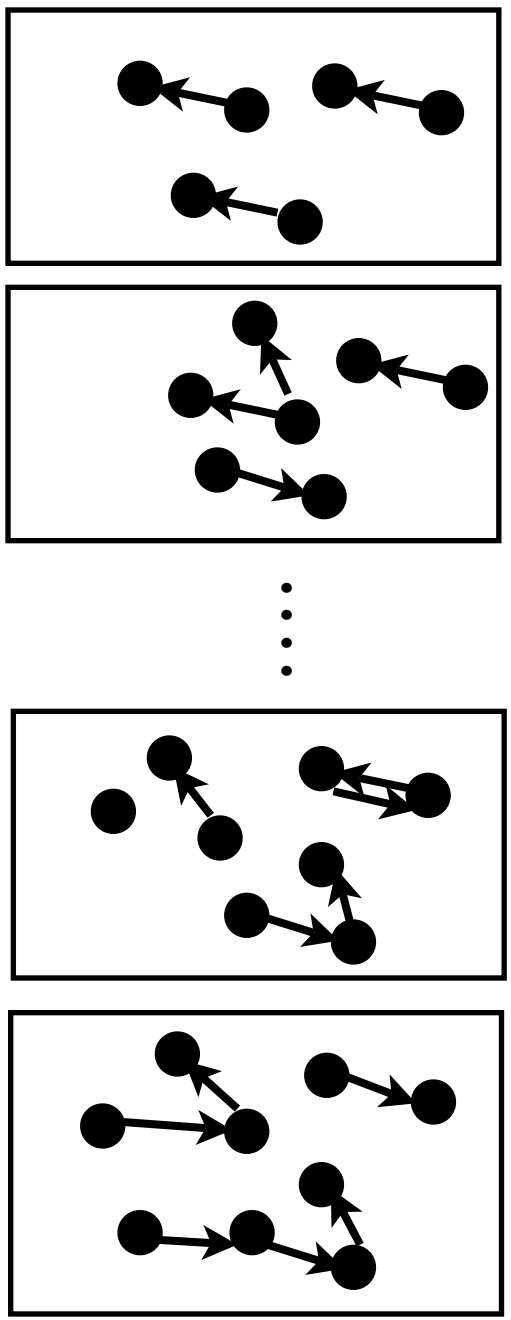




**complex**



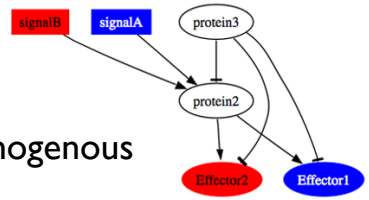
# The Model



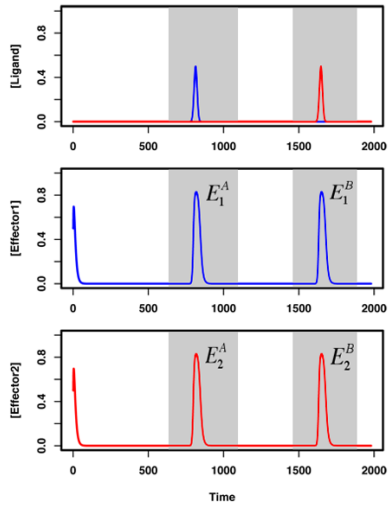
Initial Homogenous Population

Time

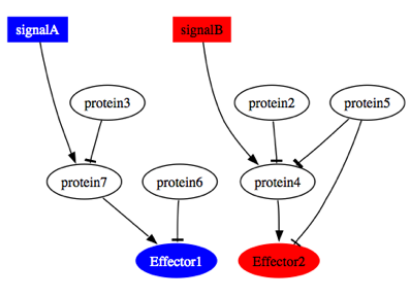
Final Population



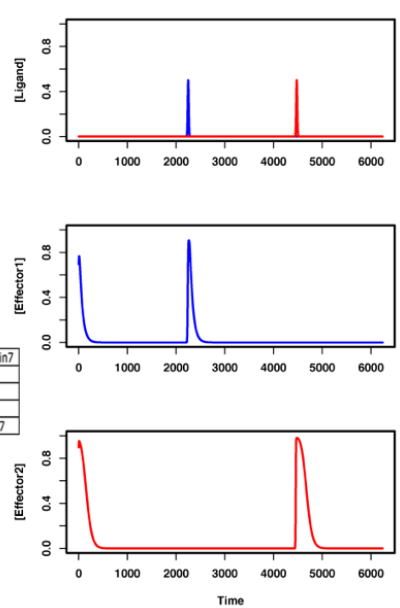

| | SignalA | SignalB | Protein2 | Protein3 | Effector1 | Effector2 |
|---|---|---|---|---|---|---|
| Protein2 | 0.5 | 0.5 | 0.0 | -0.5 | 0.0 | 0.0 |
| Protein3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Effector1 | 0.0 | 0.0 | 1.0 | -0.2 | 0.0 | 0.0 |
| Effector2 | 0.0 | 0.0 | 1.0 | -0.2 | 0.0 | 0.0 |

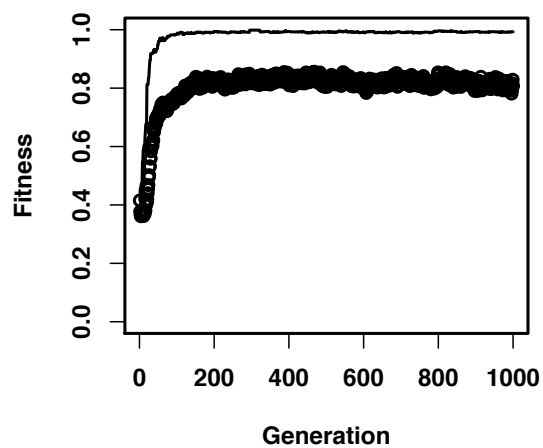



$$F = \frac{1}{4} \cdot \left[ (E_1^A + E_2^B) + (E_1^A + E_2^B - E_1^B - E_2^A) \right] - n \cdot c$$

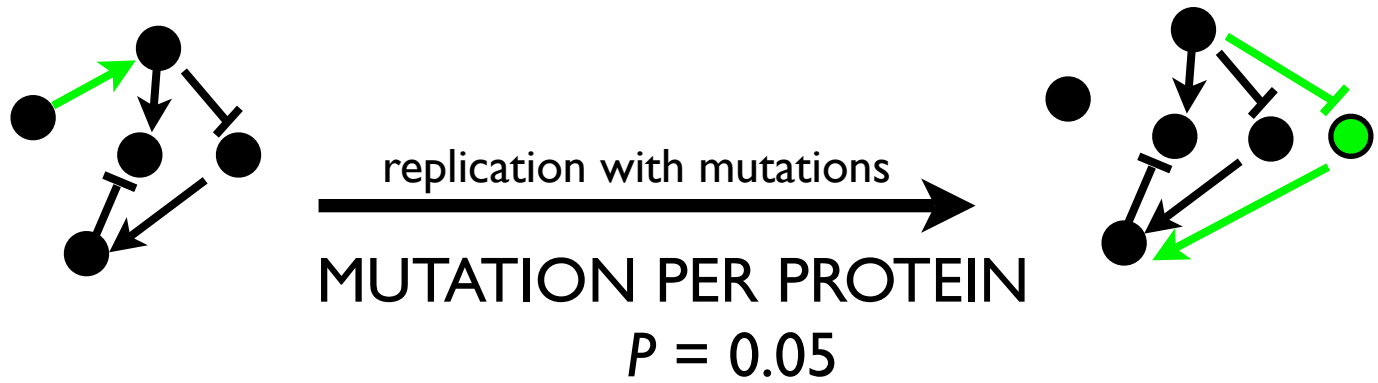



| | SignalA | SignalB | Protein2 | Protein3 | Protein4 | Protein5 | Protein6 | Protein7 |
|---|---|---|---|---|---|---|---|---|
| Protein4 | | 0.72 | -0.03 | | | -0.02 | | |
| Protein7 | 0.25 | | | -0.21 | | | | |
| Effector1 | | | | | 1.00 | -0.40 | | |
| Effector2 | | | | | | | -0.03 | 0.37 |

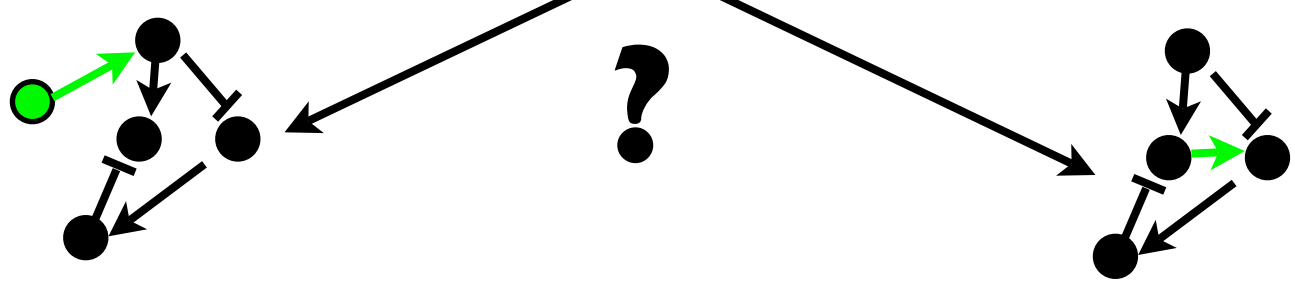


# The Model



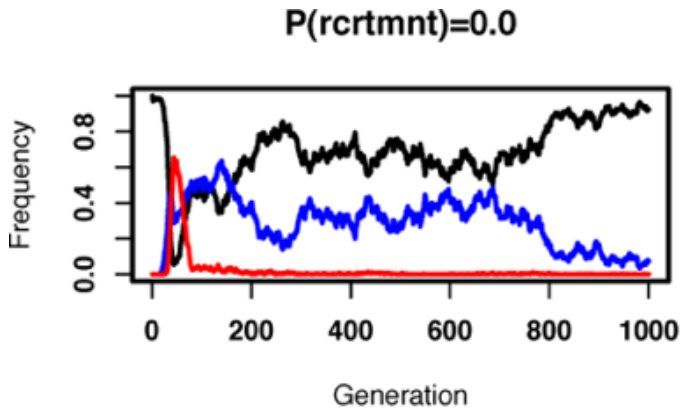

- Delete Interaction (P = 0.4)
- Delete Protein (P = 0.2)
- Change Interaction (P = 0.2)
- Duplicate Protein (P = 0.1)
- Add Interaction (P = 0.1)

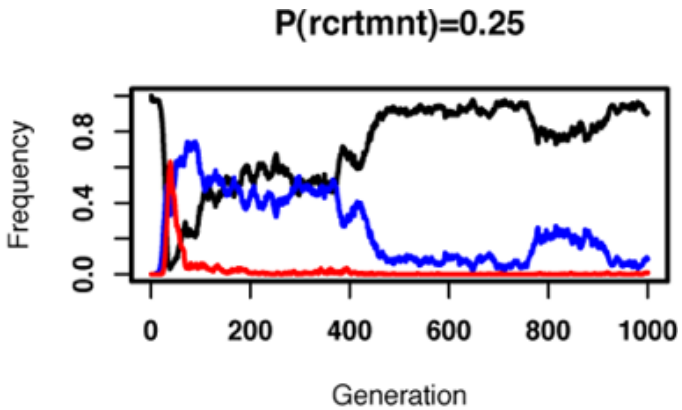



with new protein
(i.e. Protein Recruitment)

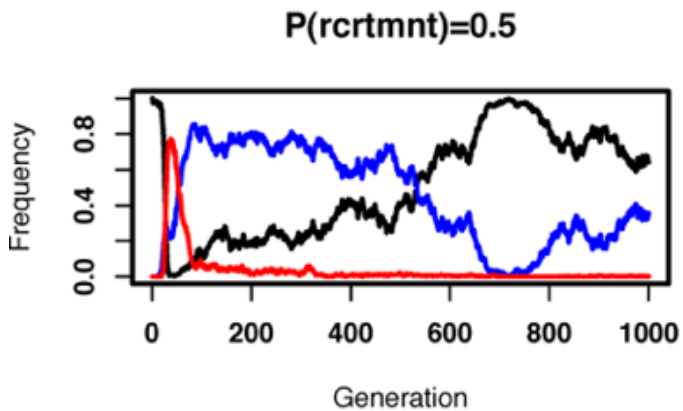
between existing proteins

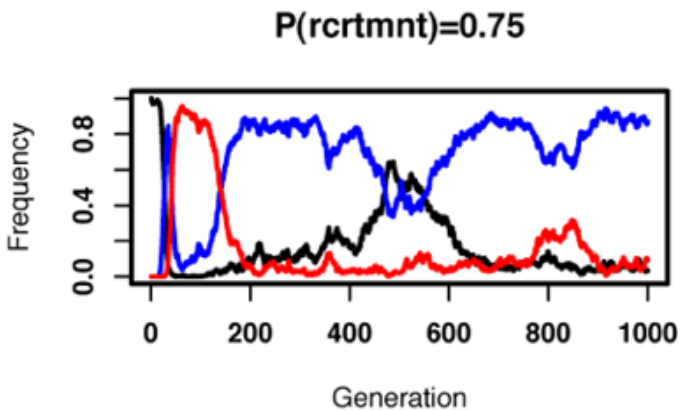
# Modularity maintenance depends on mutational mechanisms

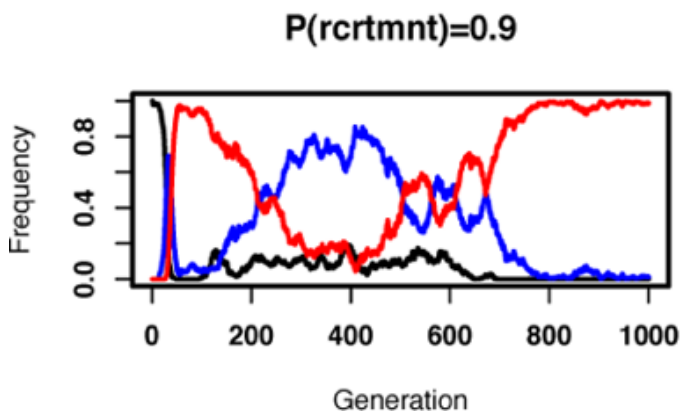


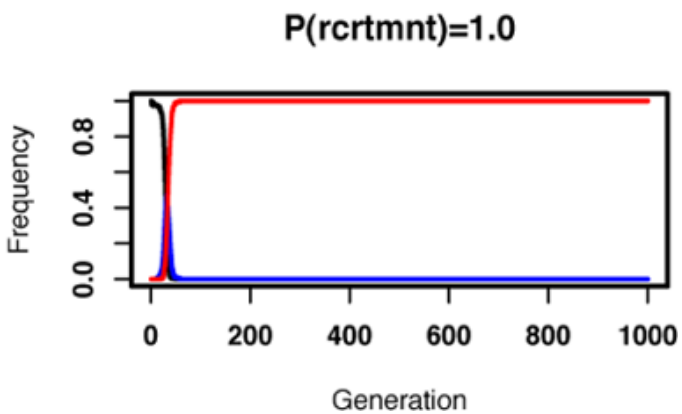




**complex**
**cross-talk**
**modular**



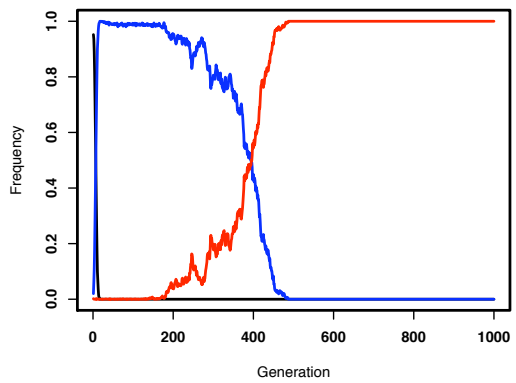



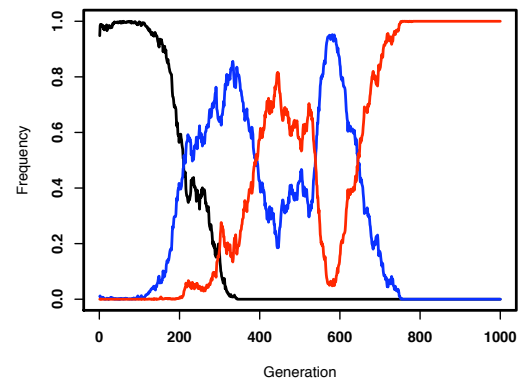



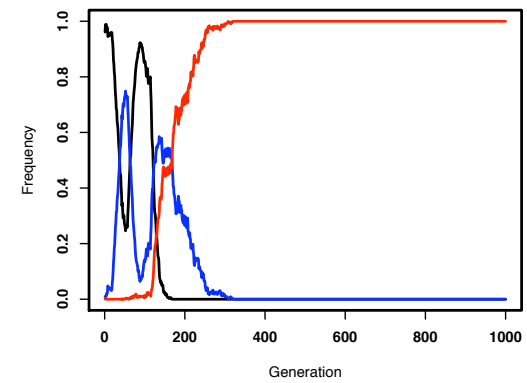
# Modularity evolution depends on initial pathway topology

**simulation 1** **simulation 2** **simulation 3**

**random initial pathways of size 6**

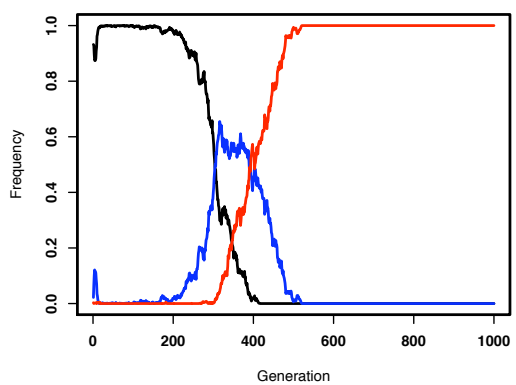


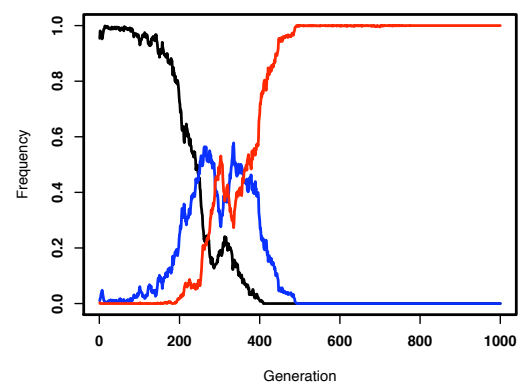



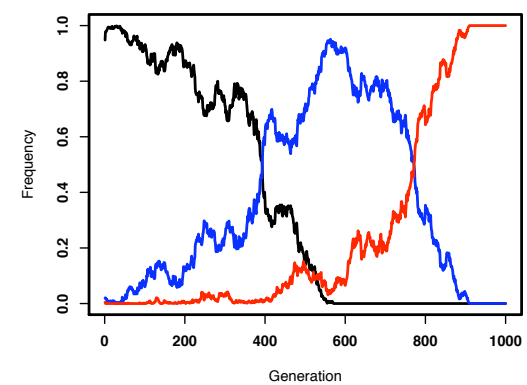




**of size 7**

Random Initial Population Of Pathways With Size 7

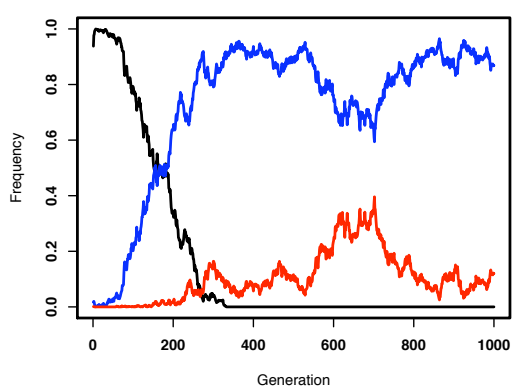


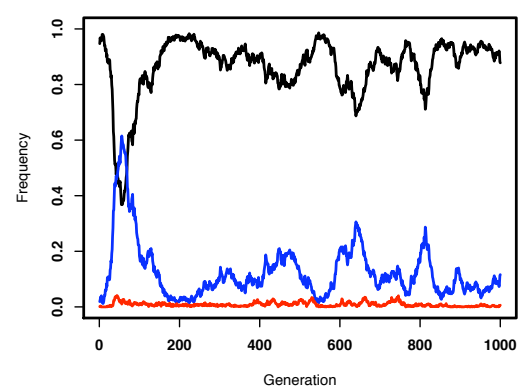



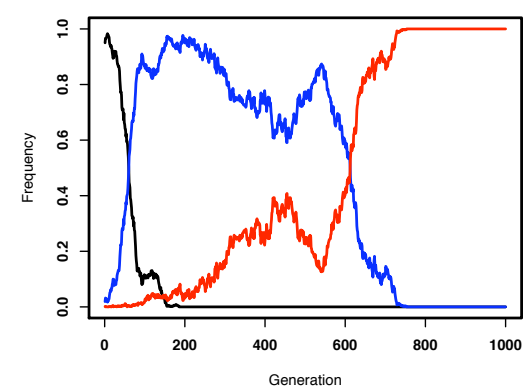




**of size 9**

Random Initial Population Of Pathways With Size 9



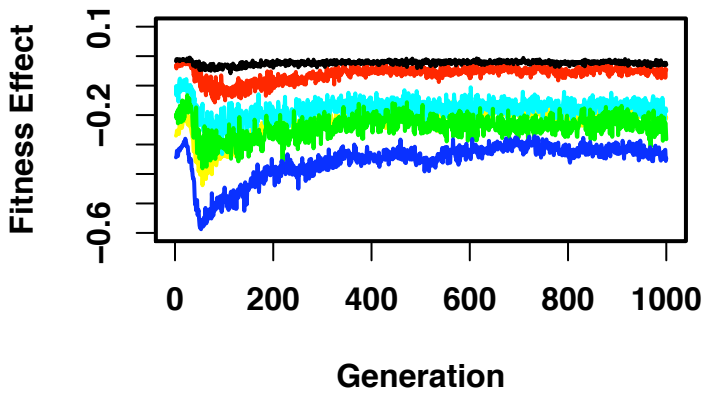



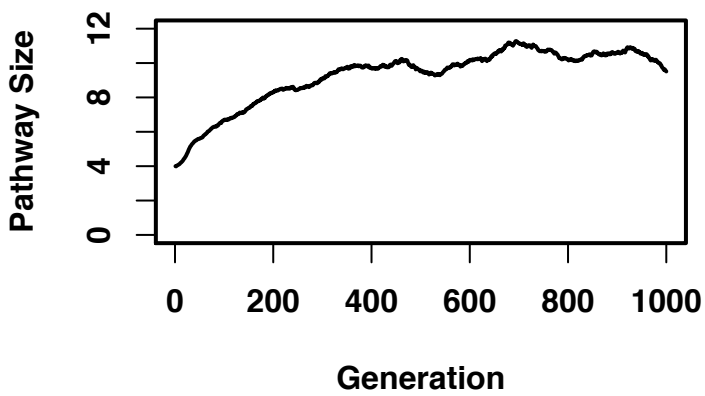



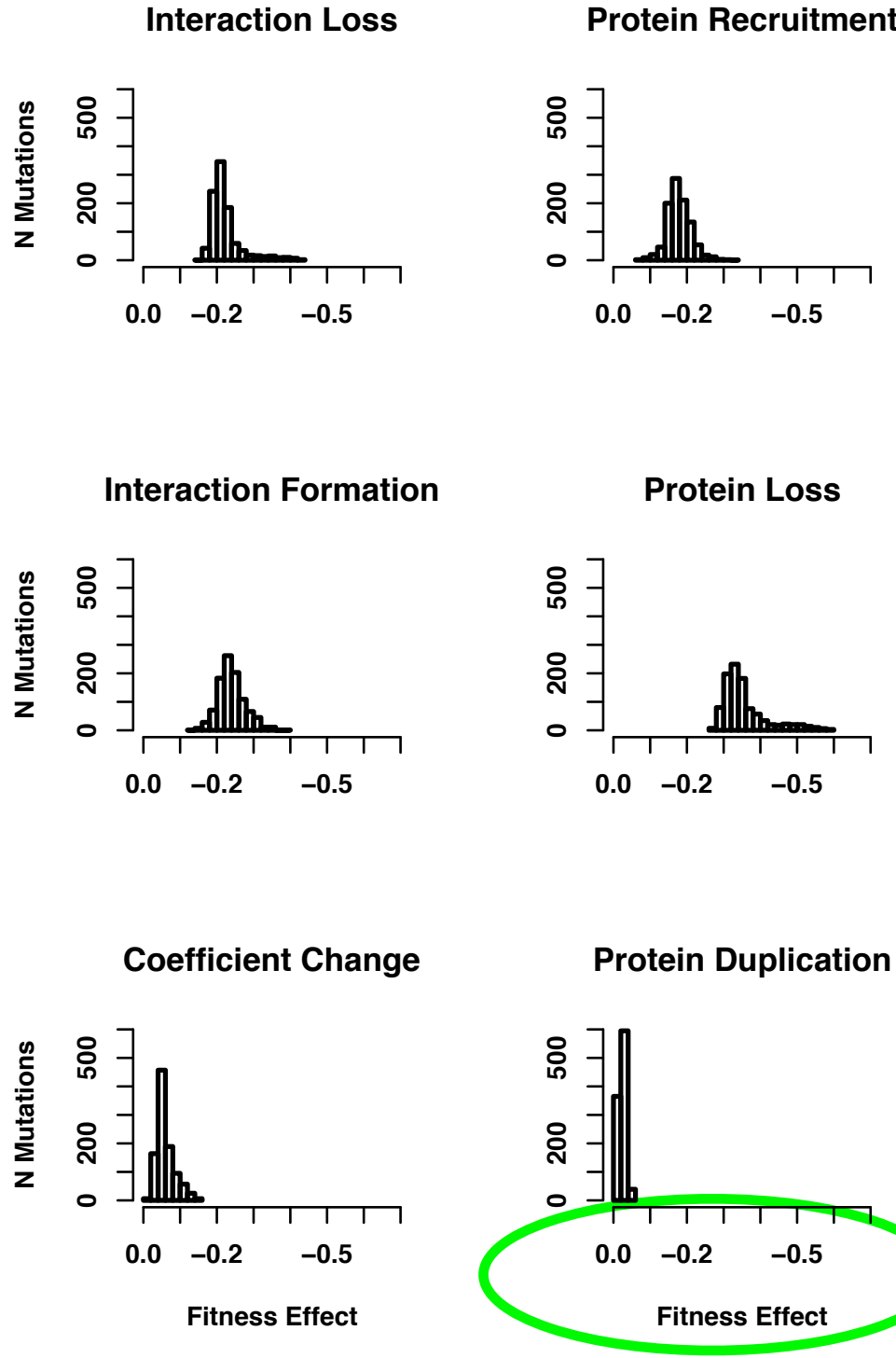

P(rcrtmnt) = 1.0

# Duplications and pathway growth

Protein duplications drive pathway evolution...

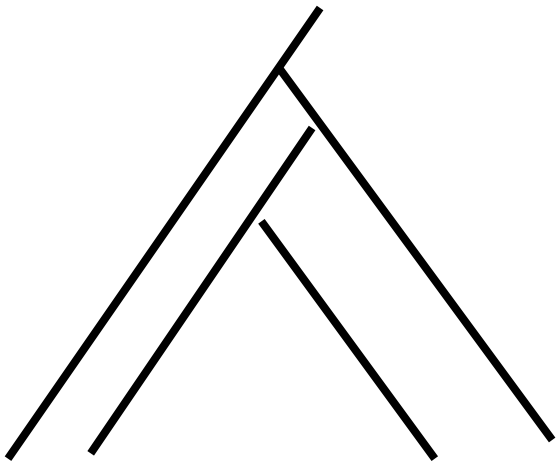


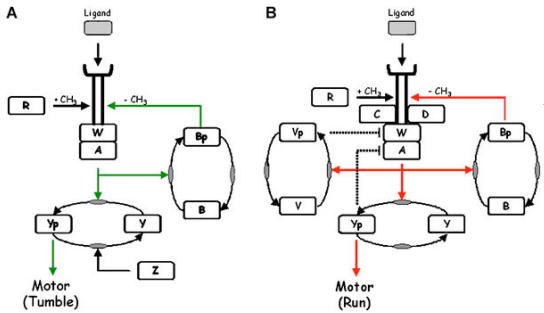



# New Answers To Old Questions Through Study of Evolution

**Natural Evolution**





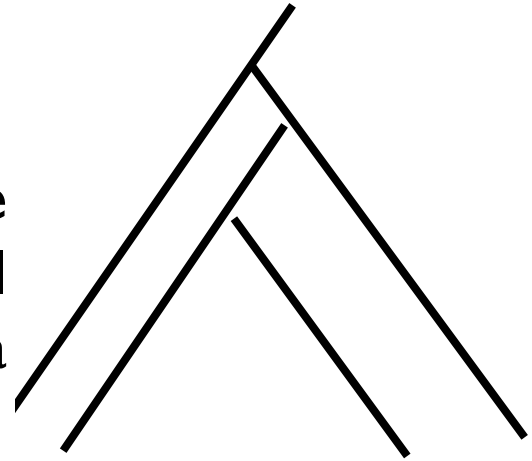

Modularity emerges readily under simple evolutionary processes without any specific selective pressure.

Determinants of modularity are the relevant rates of different mutational events and the initial location of a pathway in topology space

Approach extendable to study

- Complexity
- Modularity
- Robustness
- Evolvability

***In Silico* Evolution**



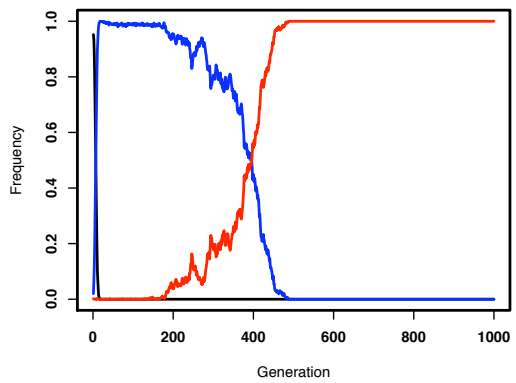
Soyer OS, Bonhoeffer S, PNAS, 2006, 103(44)

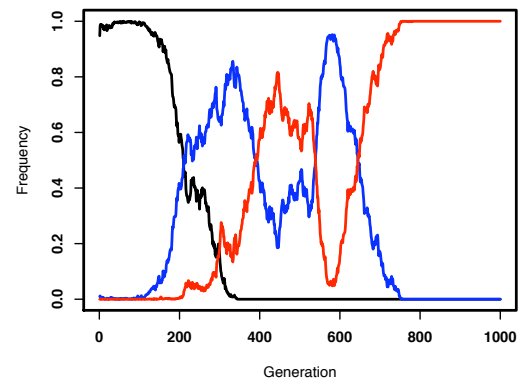
Soyer OS, BMC Evolutionary Biology, *in print*

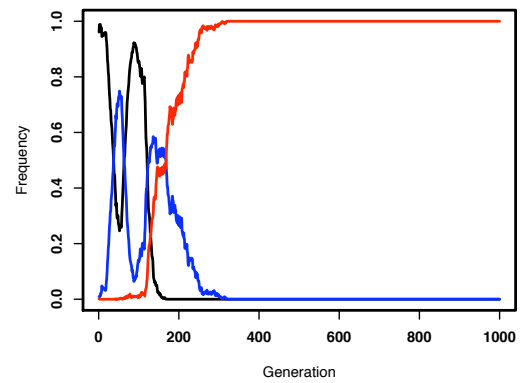
# Modularity evolution depends on pathway topology

**simulation 1** | **simulation 2** | **simulation 3**

**random initial pathways of size 6**

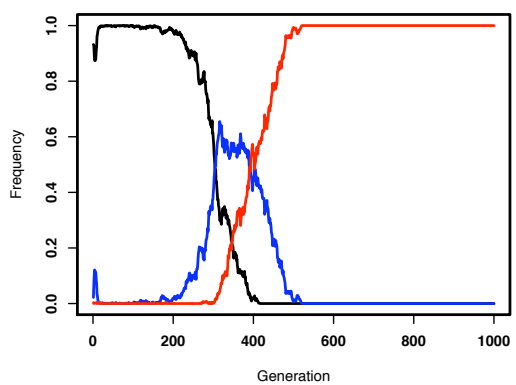


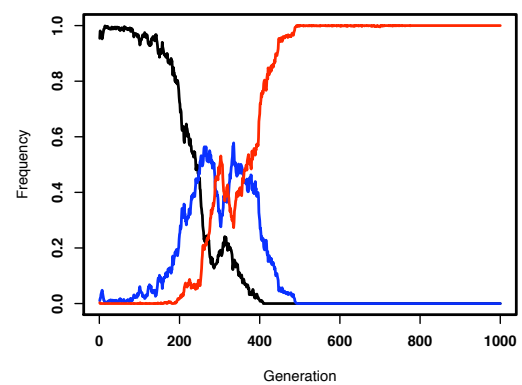



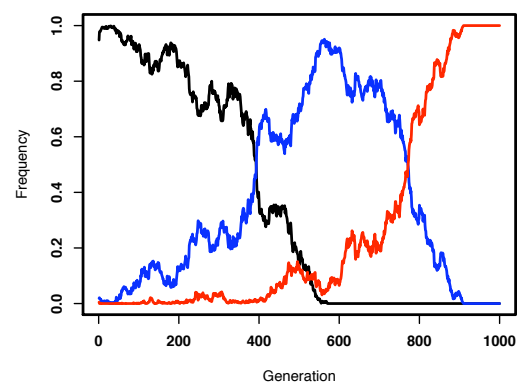




**of size 7**

Random Initial Population Of Pathways With Size 7

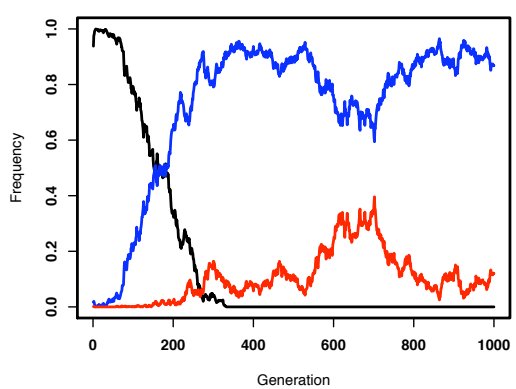


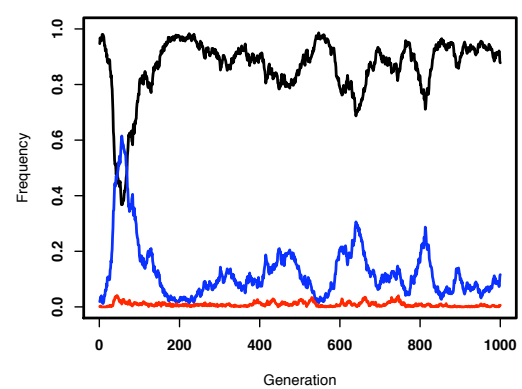



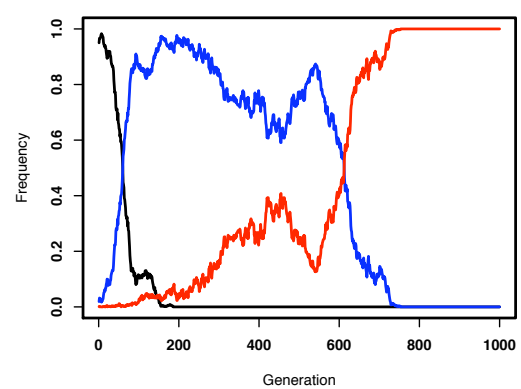




**of size 9**

Random Initial Population Of Pathways With Size 9

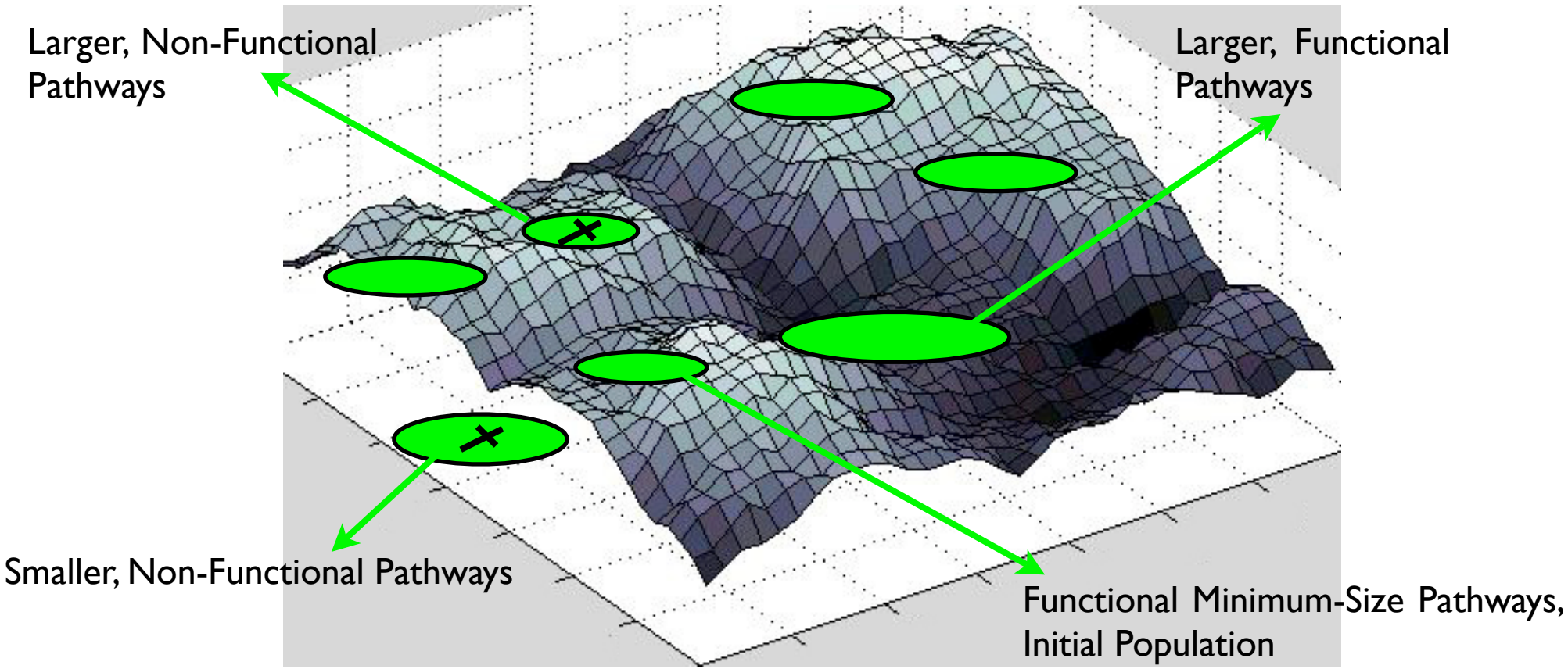



P(rcrtmnt) = 1.0

# Pathway Evolution: A random walk in topology space




**How does the topology space look like?**

# Topology space is big

| | Ligand | P1* | P2* | P3* | P4* |
|---|---|---|---|---|---|
| P1↔P1* | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| P2↔P2* | 0.0 | 0.9 | 0.0 | 0.0 | -0.5 |
| P3↔P3* | 0.0 | 0.0 | 0.7 | 0.0 | 0.0 |
| P4↔P4* | 0.0 | 0.0 | 0.0 | -0.7 | 0.0 |

**which topologies matter?**

$$N_{params} = N_{prots}^2 - N_{prots}$$

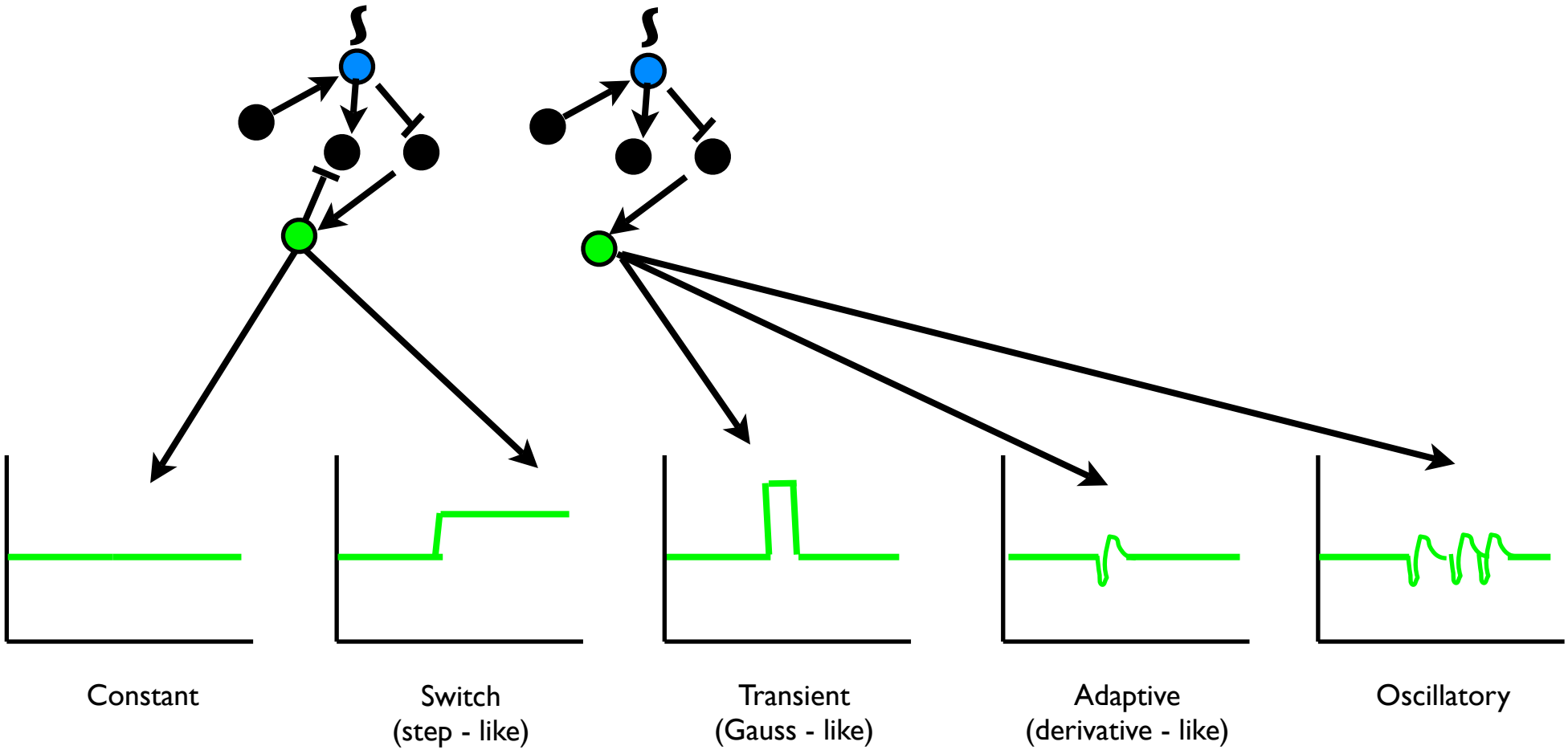
$$N_{ntwrks} = N_{values}^{N_{params}}$$

= 729 for N = 3 and values = {-1,0,1}

= 531.444 for N = 4

> $3 \times 10^9$ for N = 5

# Topology space is heterogeneous




**how do we classify topologies?**

# Topology space and pathway nature

$$\frac{\mathrm{d}[P_i]}{\mathrm{d}t} = \left[[P_i^*] \cdot \sum_j l_{ij} \cdot [P_j^*]\right] - \left[[P_i] \cdot \left(\delta_{i1} \cdot [L] + \sum_j k_{ij} \cdot [P_j^*]\right)\right]$$

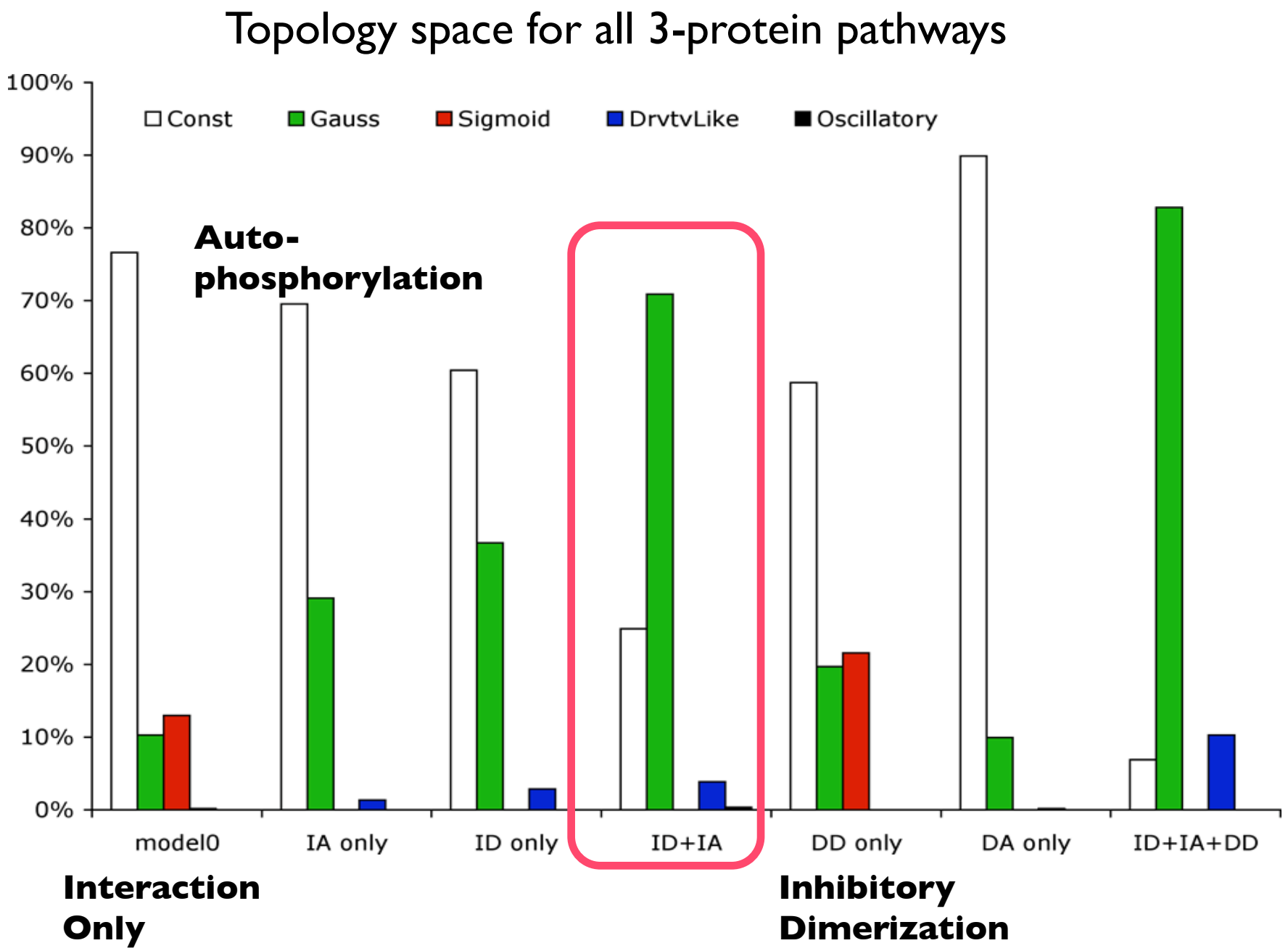
| | Ligand | P1* | P2* | P3* | P4* |
|---|---|---|---|---|---|
| P1↔P1* | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 |
| P2↔P2* | 0.0 | 0.9 | 0.0 | 0.0 | -0.5 |
| P3↔P3* | 0.0 | 0.0 | 0.7 | 0.0 | 0.0 |
| P4↔P4* | 0.0 | 0.0 | 0.0 | -0.7 | 0.0 |

$$\frac{\mathrm{d}[P_i*]}{\mathrm{dt}} = \left(k_{ii} + \sum_{j\neq i} k_{ij}[P_j*] + \delta_{i1}k_{1A}[A]\right)(1-[P_i*]) - \left(l_{ii} + \sum_{j\neq i} l_{ij}[P_j*] + \delta_{i1}l_{1A}[A]\right)[P_i*]$$

$$\frac{d[P_i]}{dt} = \sum_j \left[[P_i^*]\cdot(l_{ij}\cdot[P_j^*] + sd + \delta_{i1}\cdot dd) - [P_i]\cdot(k_{ij}\cdot[P_j^*] + sa + \delta_{i1}\cdot([L]+da)\right]$$

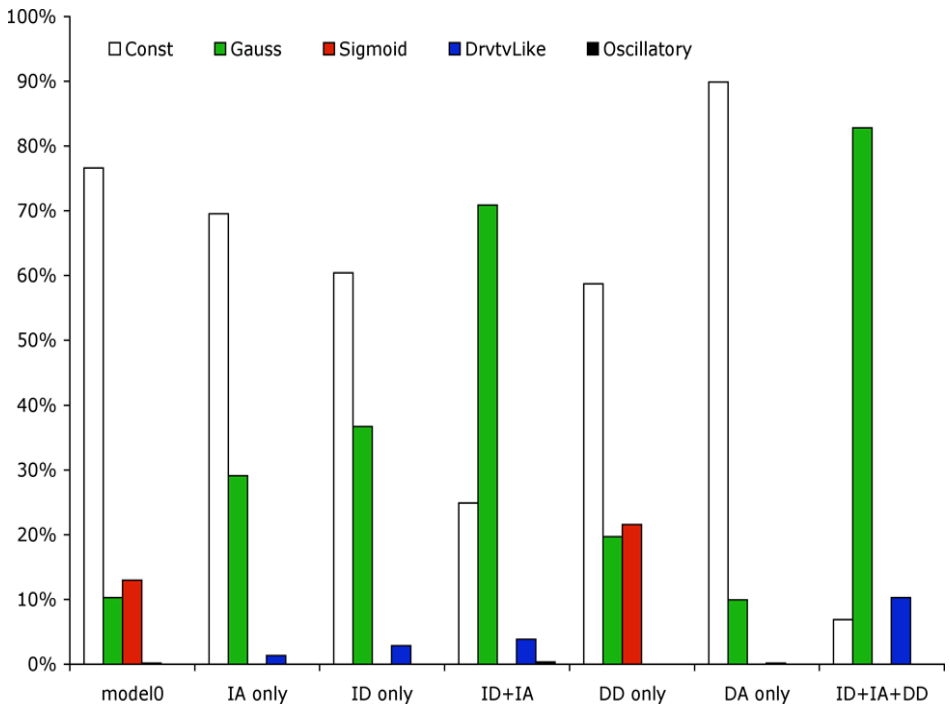
**how do we model topologies?**

# Topologies and Models: Biochemistry

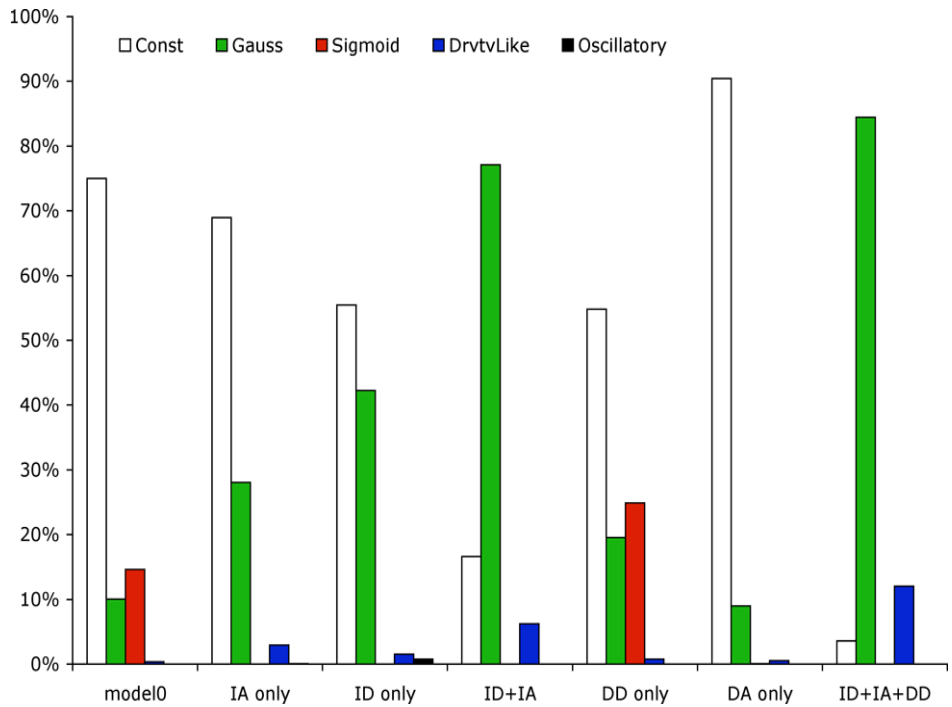


# Topologies and Models: Kinetics

Topology space for all 3-protein pathways




Binary Strength Kinetics {0 or 1}

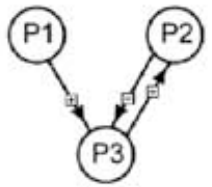



Random Strength Kinetics [0 , 1]

**Kinetics do not seem to affect overall distribution**

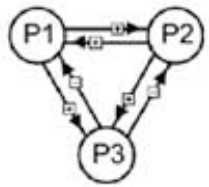
# What Determines Pathway Dynamics ? Topology or Kinetics

## Topology Nr. 19




| Model | Response Summary of 1000 Mutants | | | | | | Response |
|---|---|---|---|---|---|---|---|
| | Constant | Gauss-like | Derivative-like | Oscillatory | Bistable Switch | Integrator-like | Diversity |
| Null | 0 | 0 | 0 | 0 | 0 | 1000 | 0.034 |
| SD | 0 | 1000 | 0 | 0 | 0 | 0 | 0.034 |
| SA | 1000 | 0 | 0 | 0 | 0 | 0 | 0.039 |
| SD+SA | 2 | 998 | 0 | 0 | 0 | 0 | 0.036 |

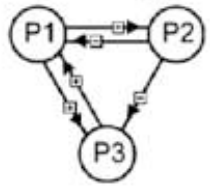
## Topology Nr. 260




| Model | Response Summary of 1000 Mutants | | | | | | Response |
|---|---|---|---|---|---|---|---|
| | Constant | Gauss-like | Derivative-like | Oscillatory | Bistable Switch | Integrator-like | Diversity |
| Null | 554 | 0 | 0 | 0 | 167 | 279 | 0.549 |
| SD | 23 | 460 | 103 | 413 | 0 | 0 | 1.391 |
| SA | 1000 | 0 | 0 | 0 | 0 | 0 | 0.039 |
| SD+SA | 19 | 979 | 2 | 0 | 0 | 0 | 0.078 |

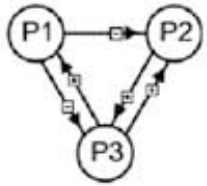
---

## Topology Nr. 291




| Model | Response Summary of 1000 Mutants | | | | | | Response |
|---|---|---|---|---|---|---|---|
| | Constant | Gauss-like | Derivative-like | Oscillatory | Bistable Switch | Integrator-like | Diversity |
| Null | 4 | 996 | 0 | 0 | 0 | 0 | 0.039 |
| SD | 0 | 306 | 305 | 388 | 0 | 0 | 1.585 |
| SA | 2 | 998 | 0 | 0 | 0 | 0 | 0.036 |
| SD+SA | 1 | 348 | 651 | 0 | 0 | 0 | 0.954 |

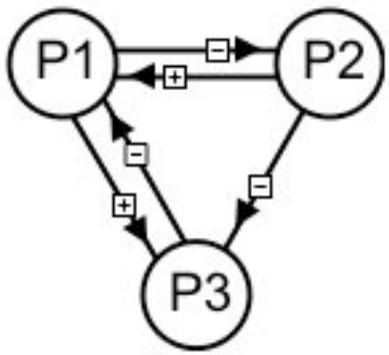
## Topology Nr. 526




| Model | Response Summary of 1000 Mutants | | | | | | Response |
|---|---|---|---|---|---|---|---|
| | Constant | Gauss-like | Derivative-like | Oscillatory | Bistable Switch | Integrator-like | Diversity |
| Null | 1000 | 0 | 0 | 0 | 0 | 0 | 0.039 |
| SD | 8 | 468 | 153 | 359 | 0 | 0 | 1.468 |
| SA | 1000 | 0 | 0 | 0 | 0 | 0 | 0.039 |
| SD+SA | 4 | 991 | 5 | 0 | 0 | 0 | 0.081 |

**Answer depends on topology and biochemistry**

# Specific Topologies

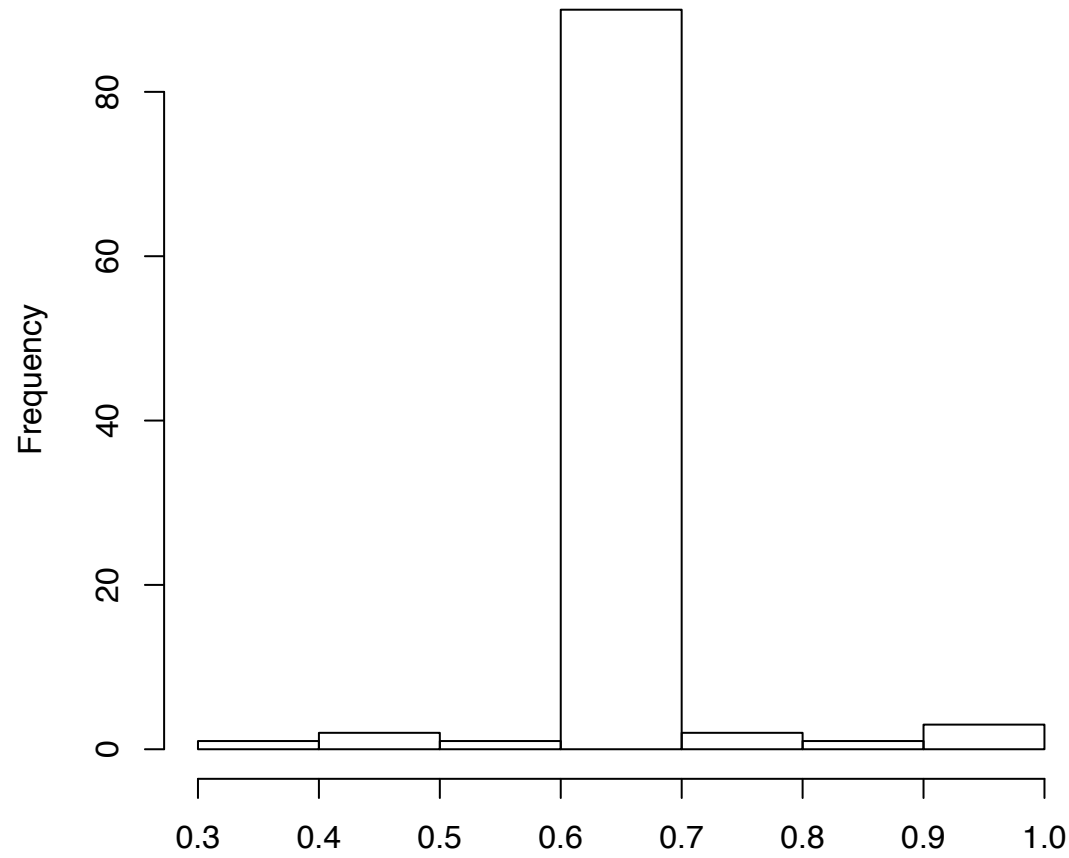



Solution of ODEs indicate neutral stability with respect to protein 3

$$P_i \underset{\sum_j l_{ij}[P_j^*]}{\overset{\sum_j k_{ij}[P_j^*]}{\rightleftharpoons}} P_i^*$$

Pi + Pi* = const.

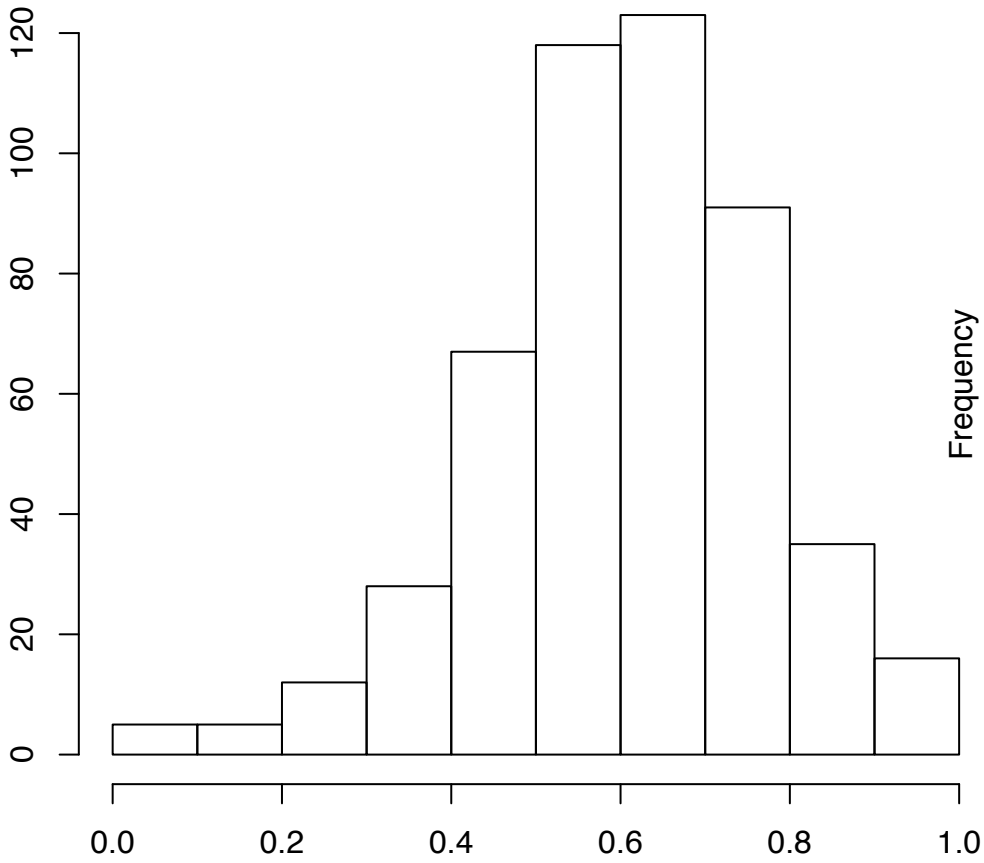
Steady state concentration of active Protein 3



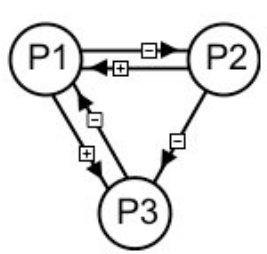

In numerical simulations, protein 3 has specific SS value with changing active fraction of proteins at start of simulation.

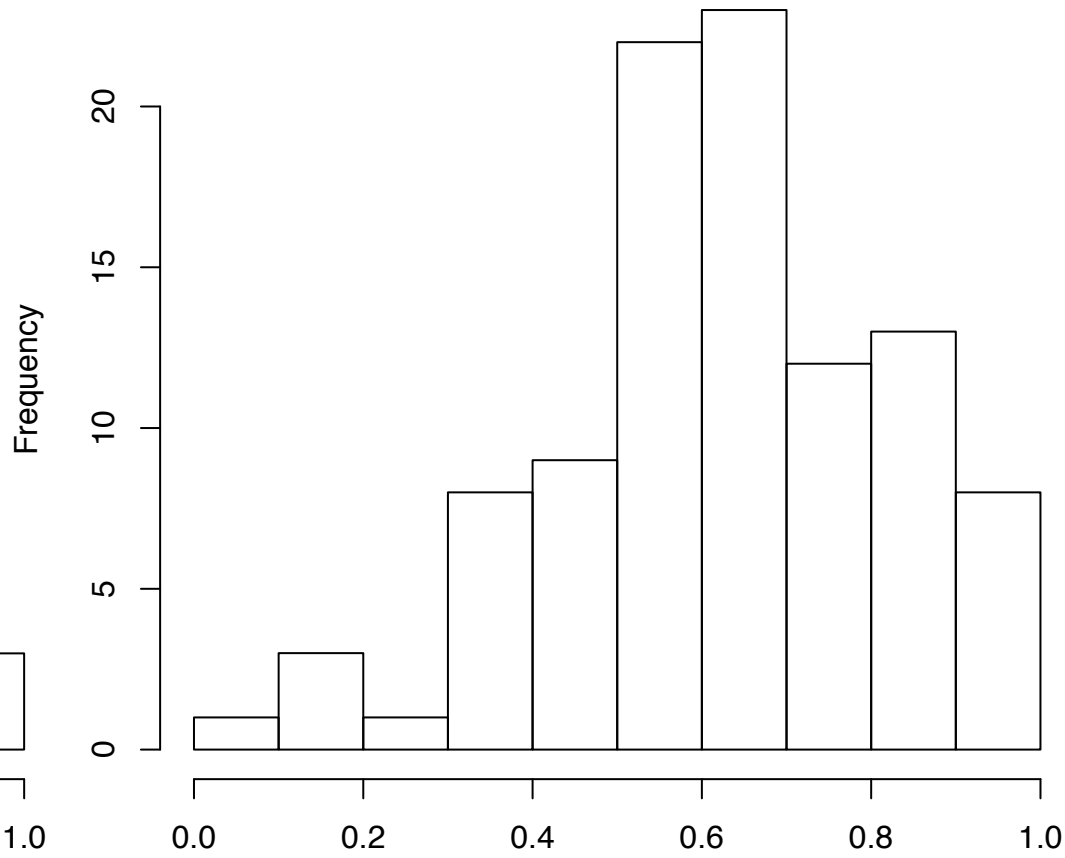
# Specific Topologies

Steady state concentration of active Protein 3



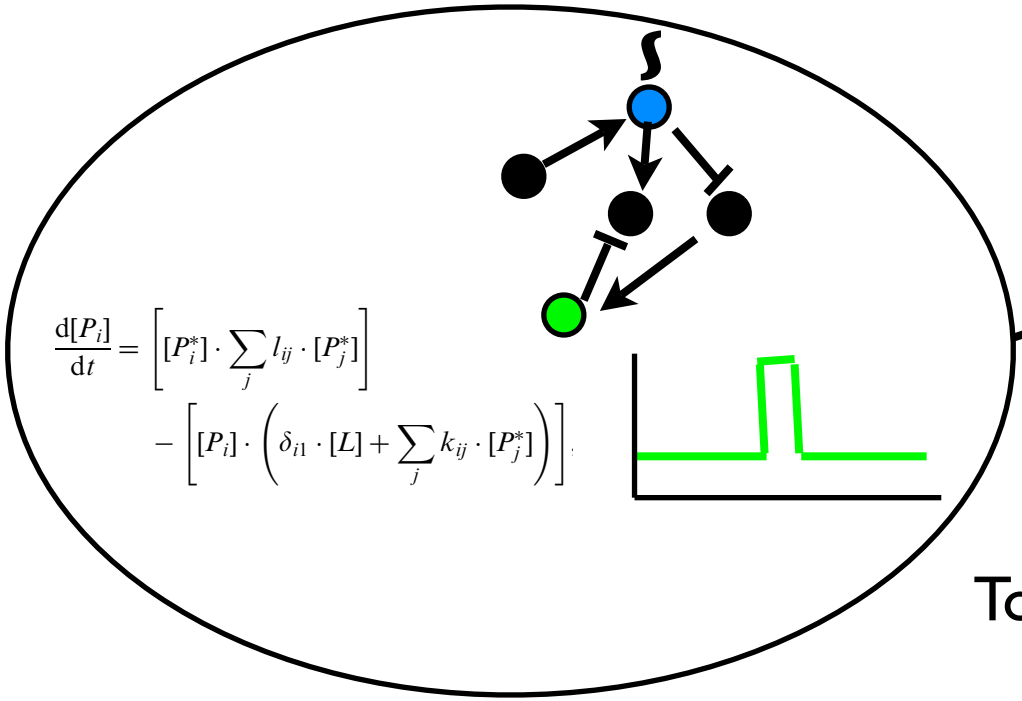

changing coefficients.

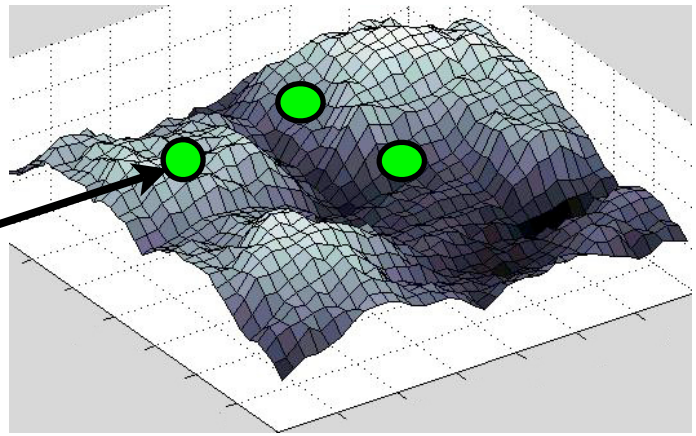



changing total protein concentrations.

# Topologies, Biochemistry, and Evolution



Topology dynamics relation is complex

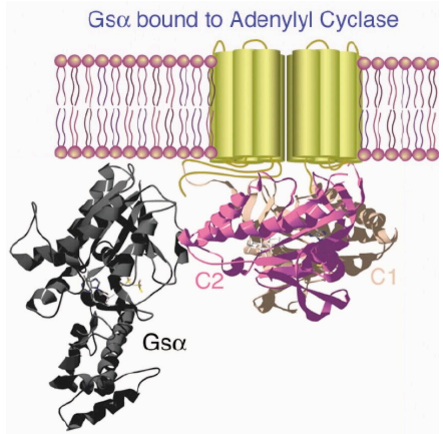
Biochemical processes, even those treated as negligable, can have important effects on pathway dynamics

Dimerization, auto phosphorylation might be important as determinants of dynamic behavior and robustness

Soyer OS, Salathe M, Bonhoeffer S, JTB, 2006, 238

# Big Picture

System Specific
Understanding



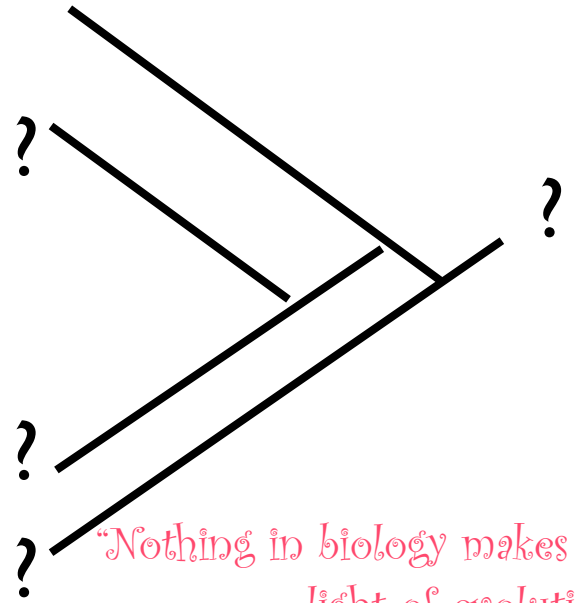

realistic models

toy models

theory

experiments

- From Systems to Behavior
- Global Properties
- Design Principles ??




"Nothing in biology makes sense except in light of evolution"
Theodosius Dobzhansky

**EVOLUTIONARY SYSTEMS BIOLOGY**

# The Microsoft Research - University of Trento Centre for Computational and Systems Biology

Theoretical Biology Group at ETH, Zürich

Richard Goldstein
MRC, UK