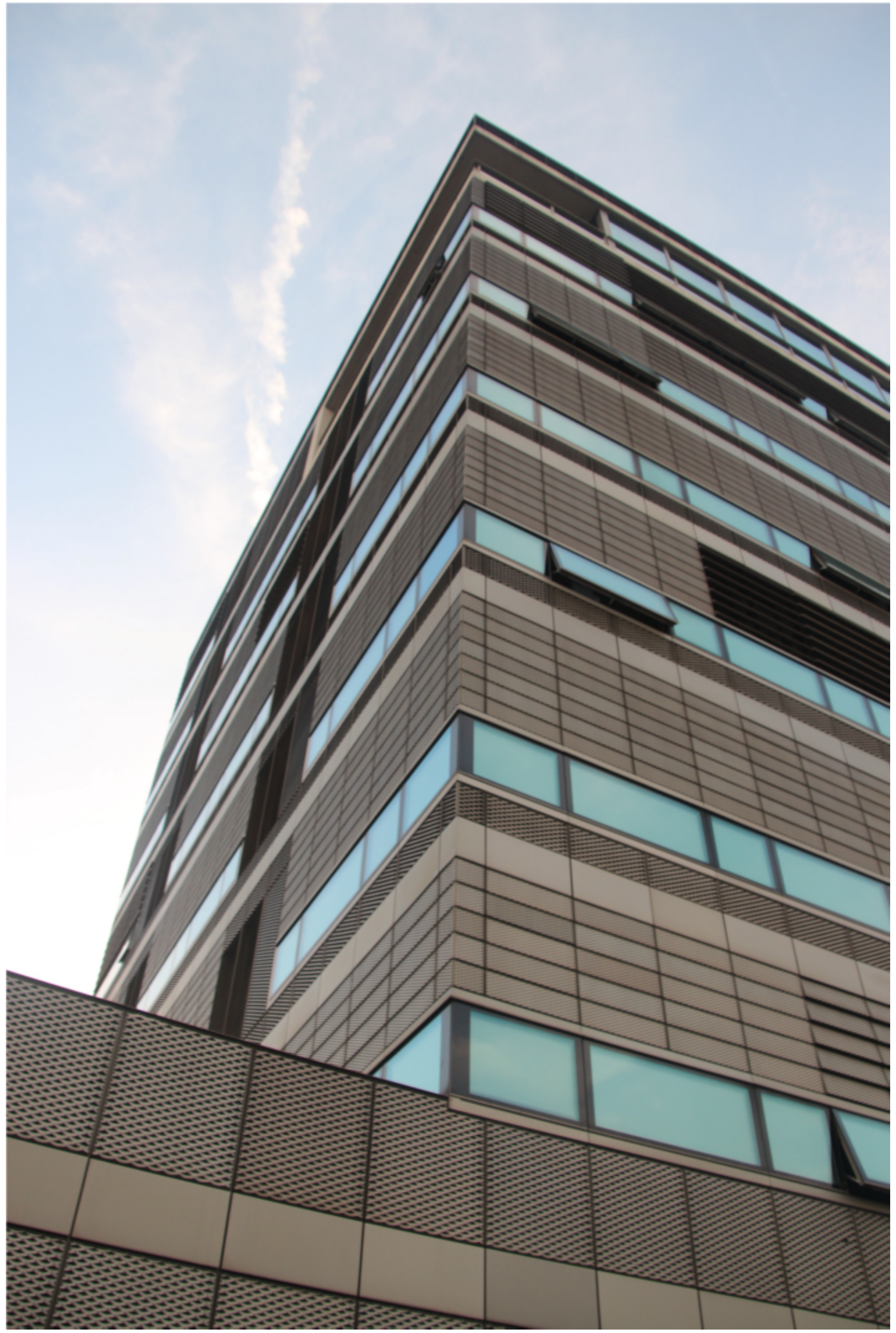


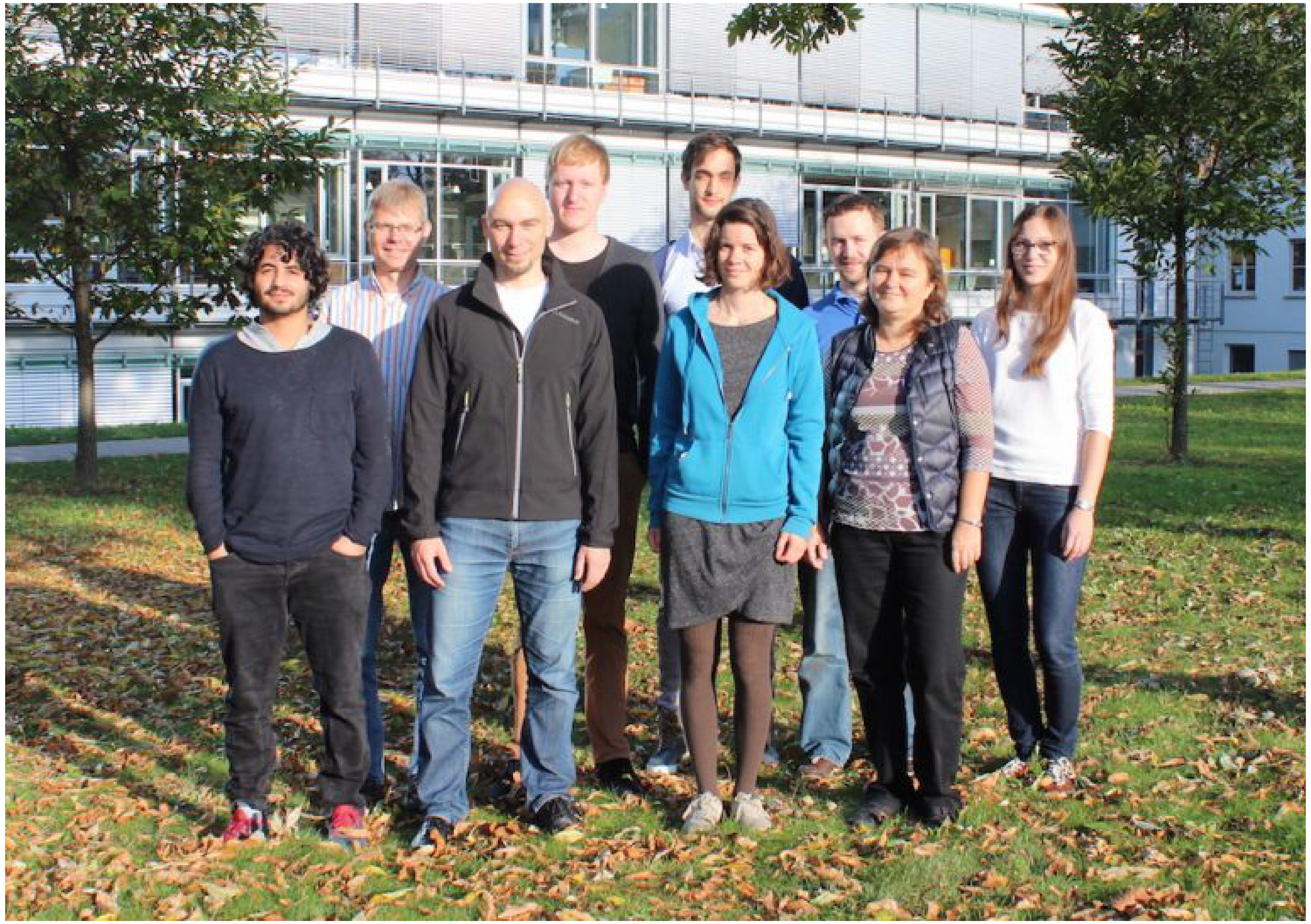
Information biology – How to apply Shannon's information theory to biology

35C3 Leipzig

Jürgen Pahle
https://lab.pahle.org

27th December 2018

Pahle Lab @ BioQuant
Heidelberg University



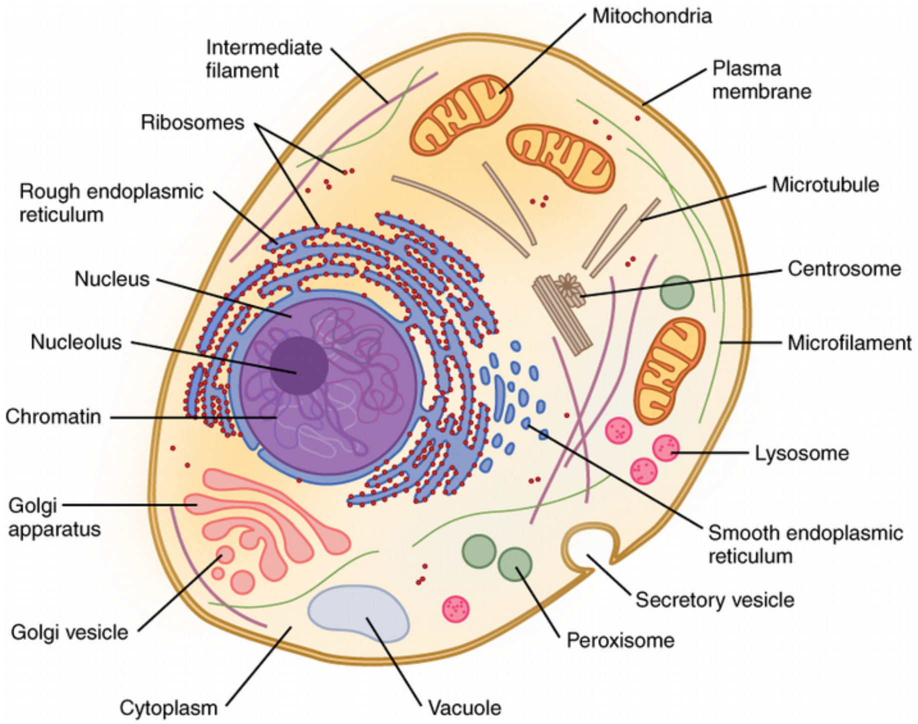
Biological Information Processing Group

BioQuant
MODEL base of LIFE

Funded by
BIOMS

- Cellular communication and information processing is everywhere.
- How do cells manage all these computations with relatively unreliable components (proteins, instead of transistors) and (intrinsic) fluctuations in molecular numbers?
- If signalling is impaired severe diseases can be the result, e.g. cancer or epilepsy.
- Cellular signalling pathways have been studied on a molecular level in detail but surprisingly little conceptual work has been done...

How-to (Biochemical modelling)



- **Compartments** (nucleus, cytosol, ...)
- **Biochemical species** (proteins, enzymes, ions, ...)
- **Reactions** (who reacts with whom?, substrates, products)
- **Kinetics** (velocity of reactions)

Mathematical model

$$\frac{\mathrm{d}G_\alpha}{\mathrm{d}t} = k_1 + k_2 \cdot G_\alpha - \frac{k_3 \cdot PLC \cdot G_\alpha}{(K_4 + G_\alpha)} - \frac{k_5 \cdot [Ca^{2+}] \cdot G_\alpha}{(K_6 + G_\alpha)} \qquad G_\alpha(t_0) = 0.01\ nmol$$

$$\frac{\mathrm{d}PLC}{\mathrm{d}t} = k_7 \cdot G_\alpha - \frac{k_8 \cdot PLC}{(K_9 + PLC)} \qquad PLC(t_0) = 0.01\ nmol$$

$$\frac{\mathrm{d}[Ca^{2+}]}{\mathrm{d}t} = k_{10} \cdot G_\alpha - \frac{k_{11} \cdot [Ca^{2+}]}{(K_{12} + [Ca^{2+}])} \qquad [Ca^{2+}](t_0) = 0.01\ nmol$$

Simulation:

How does the system change over time?

Analysis of the model:

Which parts influence the behavior most?

Which states are stable (steady state, oscillations)?

"All models are wrong but some are useful"

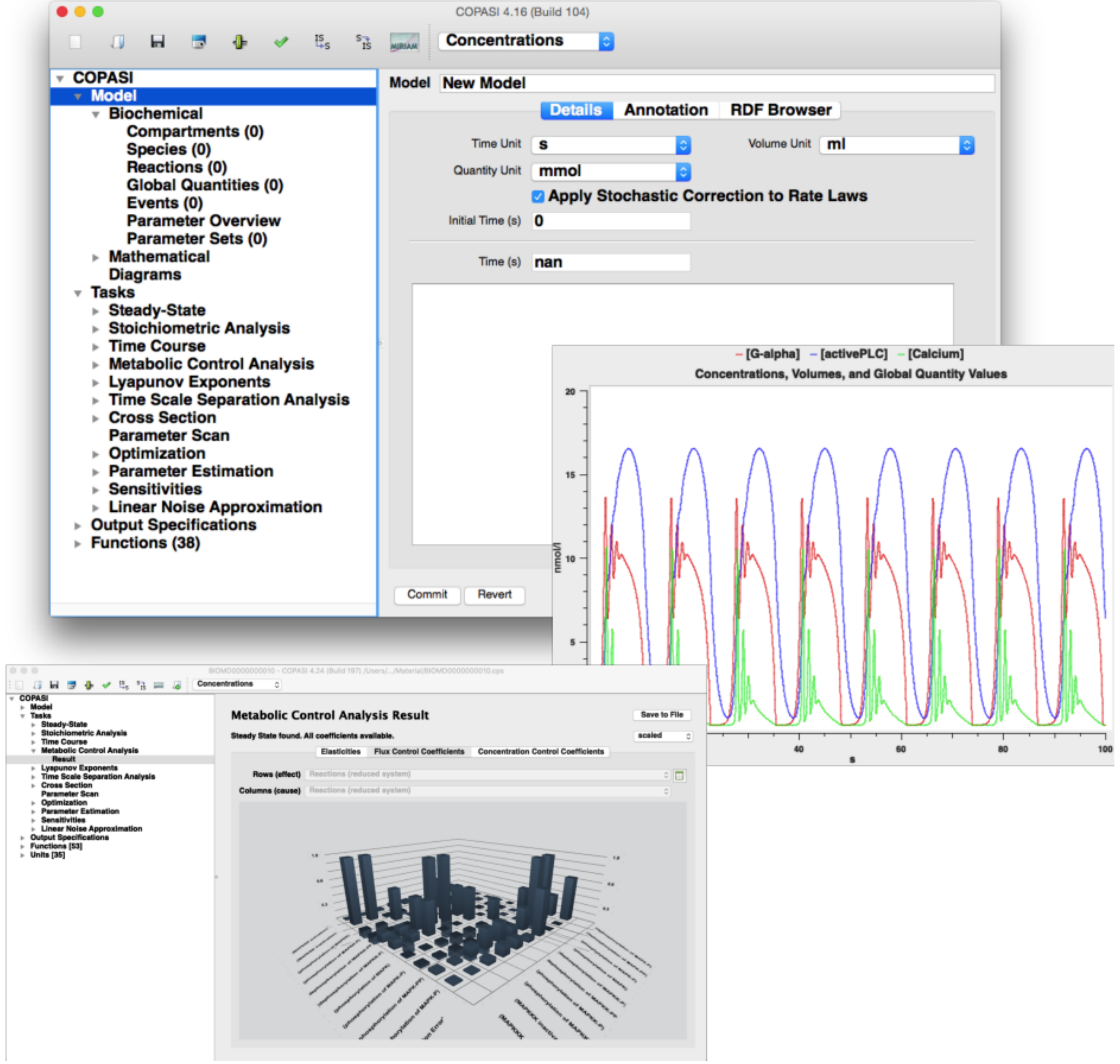
Box, G.E.P. (1979) ”Robustness in the strategy of scientific model building” in Robustness in Statistics (R.L. Launer and G.N. Wilkinson, Eds.), Academic Press

"[..] the practical question is how wrong do they have to be to not be useful"

Box, G.E.P. & Draper, N.R. (1987). Empirical Model-Building and Response Surfaces. Wiley. pp. 74

http://www.copasi.org

- COPASI (COmplex PAthway SImulator)
- Stand-alone software for the modelling, simulation and analysis of biochemical networks
- "Tool kit" with a variety of different methods:
  - Deterministic, **stochastic and hybrid simulation methods**
  - Metabolic Control Analysis, Elementary Flux Mode Analysis, Sensitivity Analysis, Cross-sections, etc.
  - Parameter Scanning, Optimization, Parameter Fitting
  - User-friendly **GUI**, runs under Mac, Linux, Windows and **command line version**
  - Non-commercial, freely available, open-source (**Artistic license**). Can be used via APIs.
  - Reads and writes the Systems Biology Markup Language (**SBML**), and allows annotations etc.



CoRC – Copasi R Connector

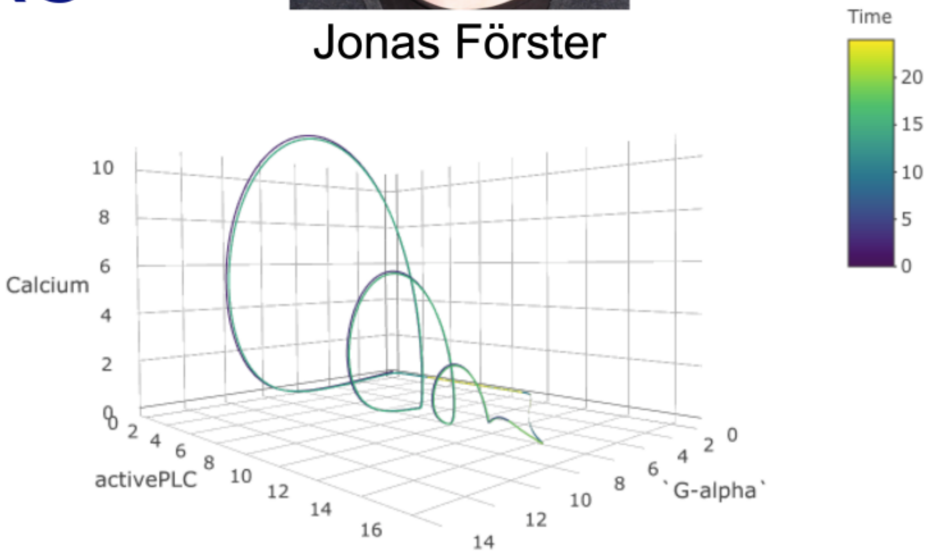
jpahle.github.io/CoRC

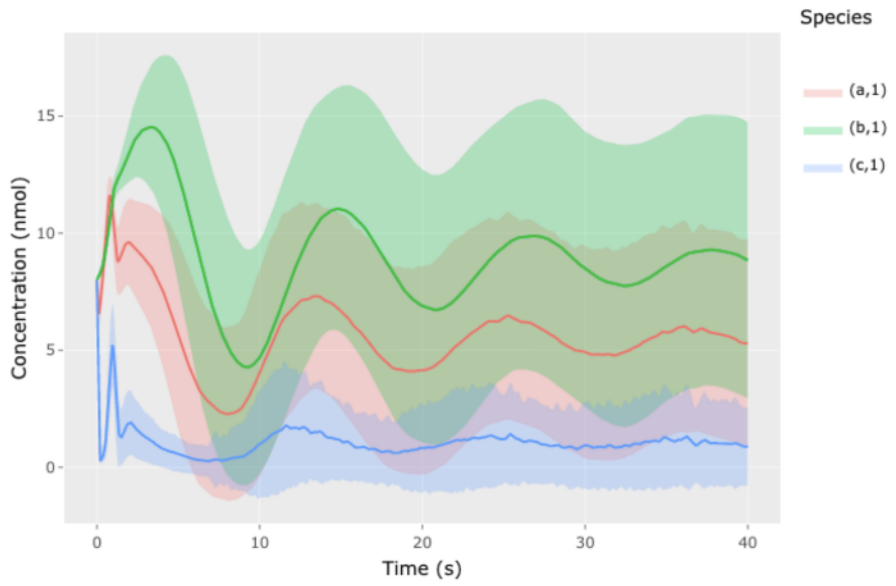
Jonas Förster

CoRC connects the

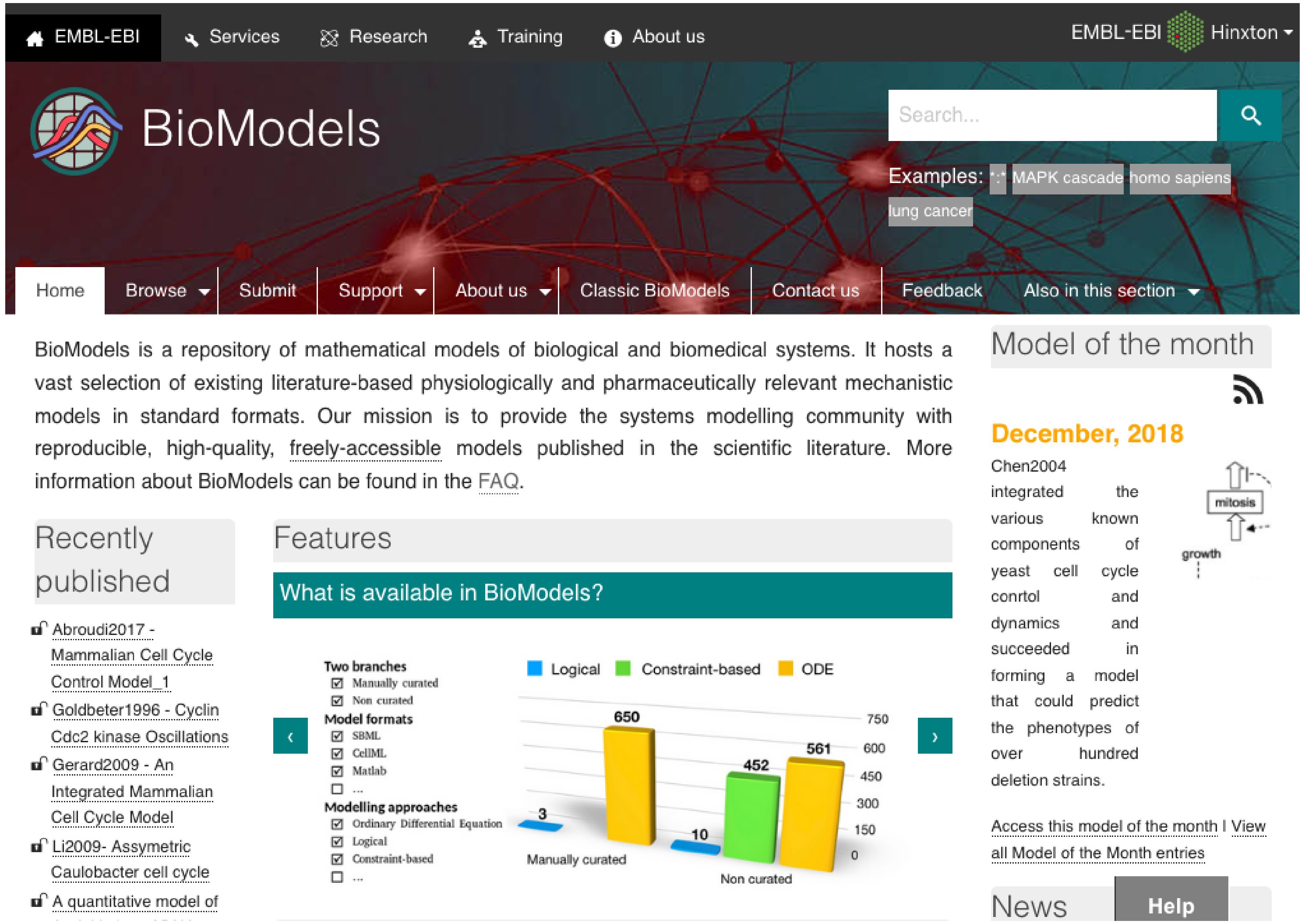
- Complex Pathway Simulator COPASI (copasi.org) and the
- (statistical) programming environment R (r-project.org).
- Free and open-source

```
loadSBML("http://www.ebi.ac.uk/biomodels-main/download?
mid=BIOMD0000000329")
setTimeCourseSettings(24, intervals = 10000)
timecourse <-  runTimeCourse(method = list(method = "directMethod",
                                           use_random_seed = T,
                                           random_seed = 1))$result
```





Biomodels.net

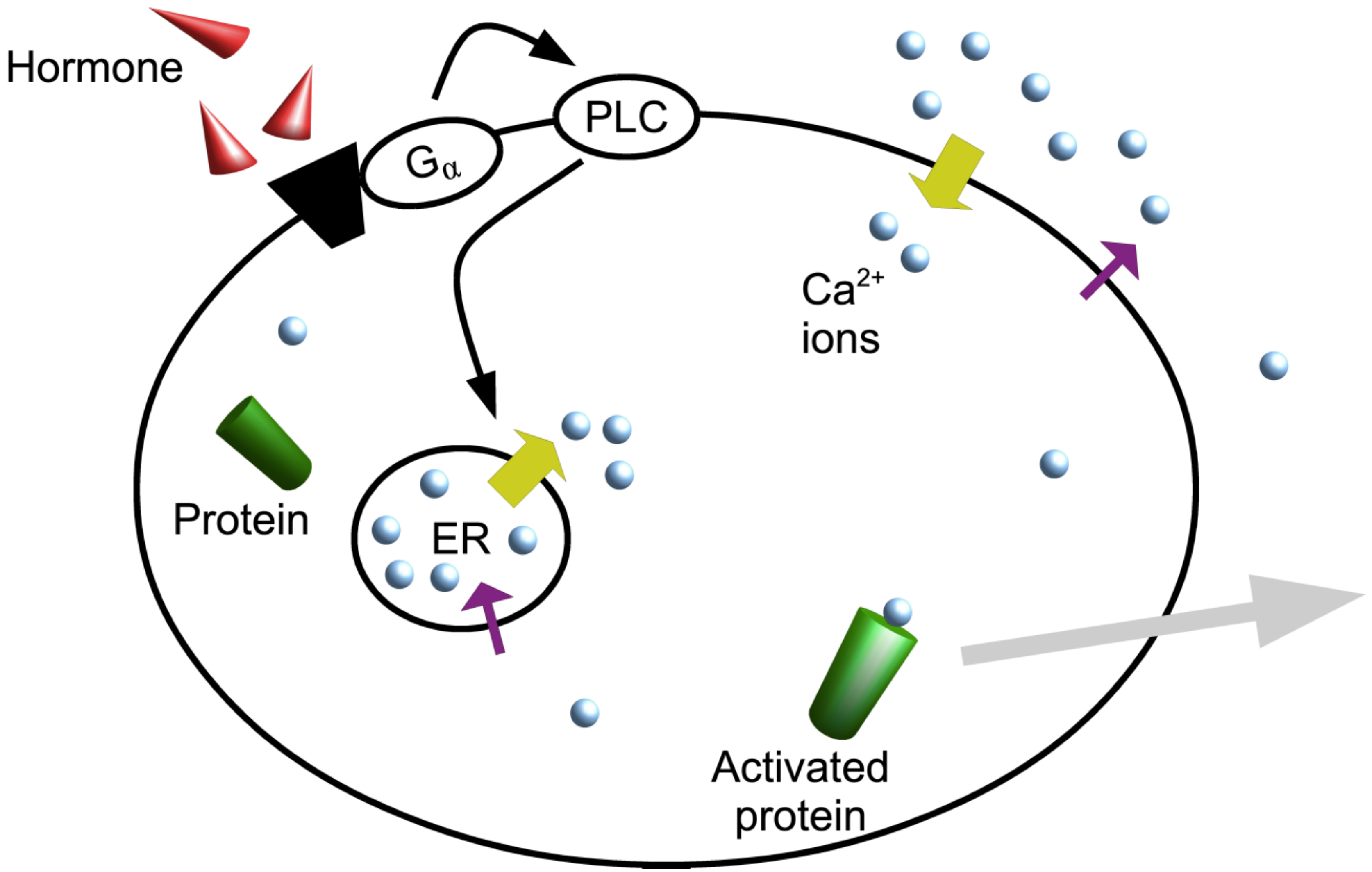


Repository for published biochemical models

- 727 manually curated models
- 7901 non-curated models

Calcium oscillation model: Biomodel 329

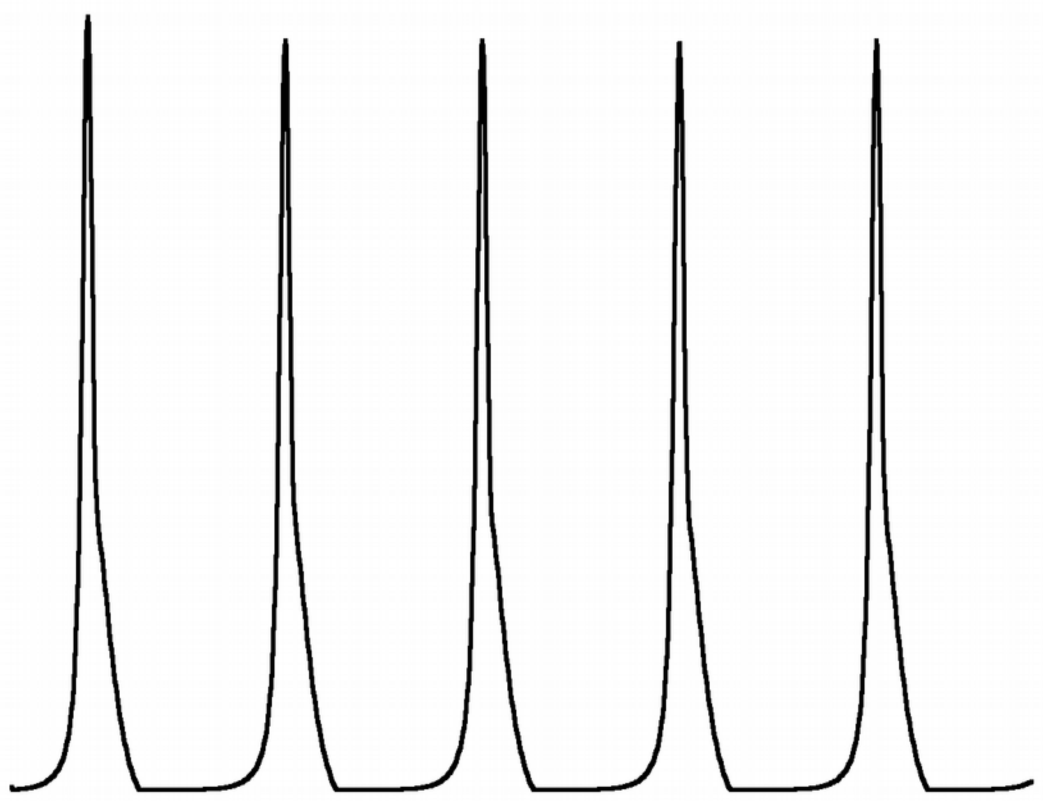
Signal transduction via Ca^{2+} ions



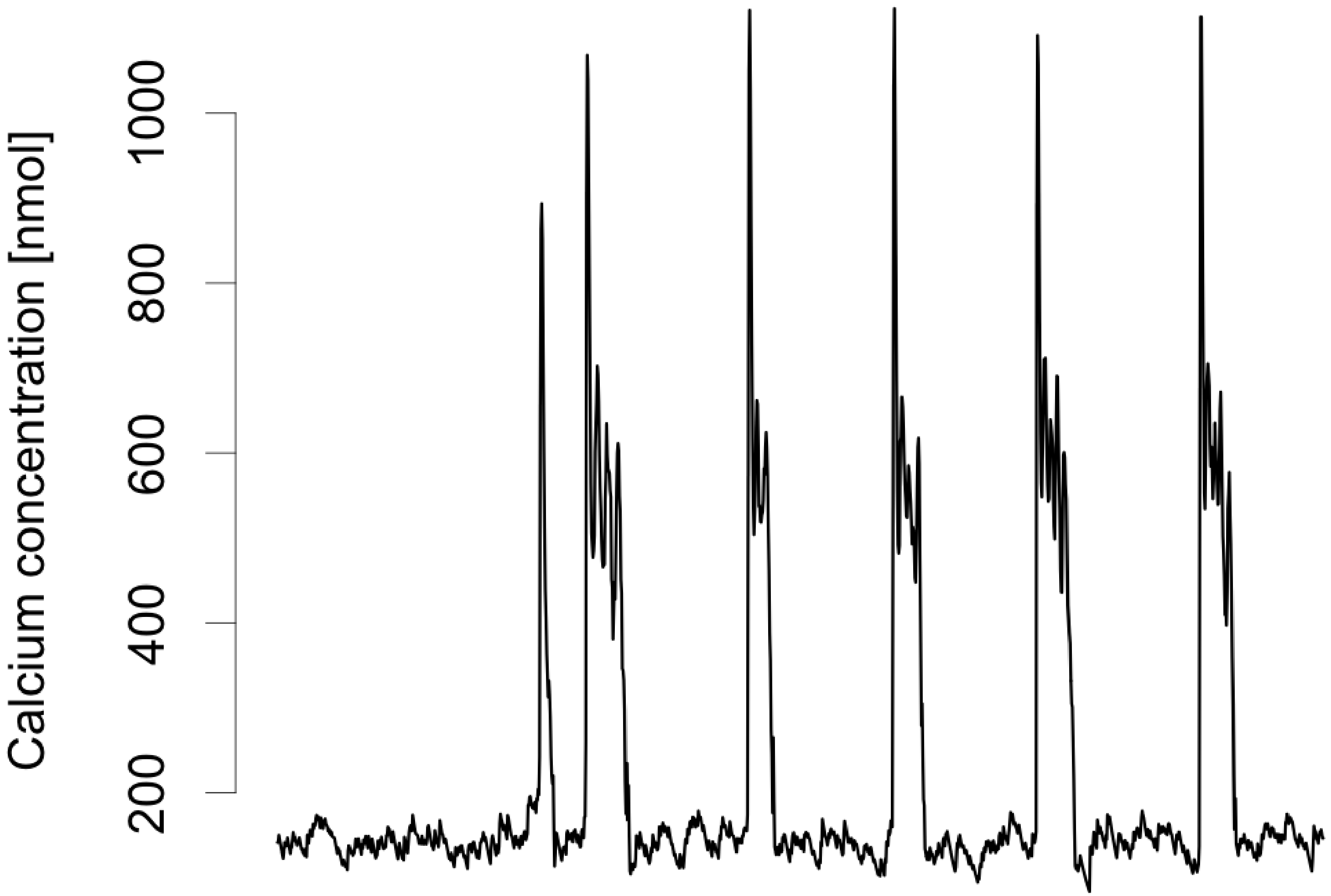
- **Movement**: muscle contraction
- **Learning**: long-term potentiation in neurons
- **Regulation** of metabolism etc.
- **Secretion** of neurotransmitters and others
- **Fertilization**
- **...**

Calcium dynamics (simulated deterministically)

Spiking



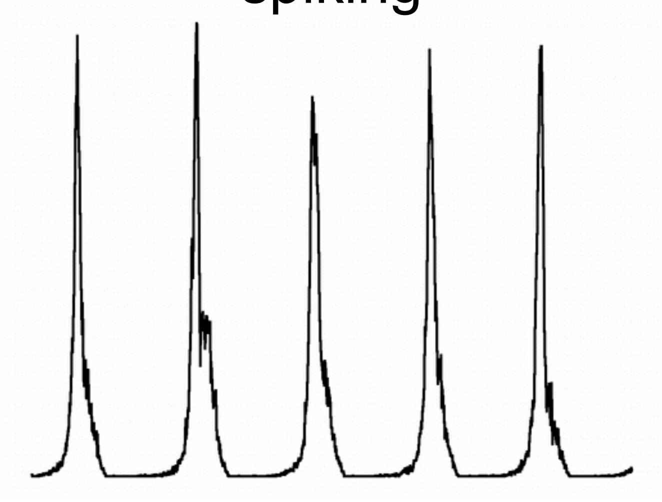
Experimental time series



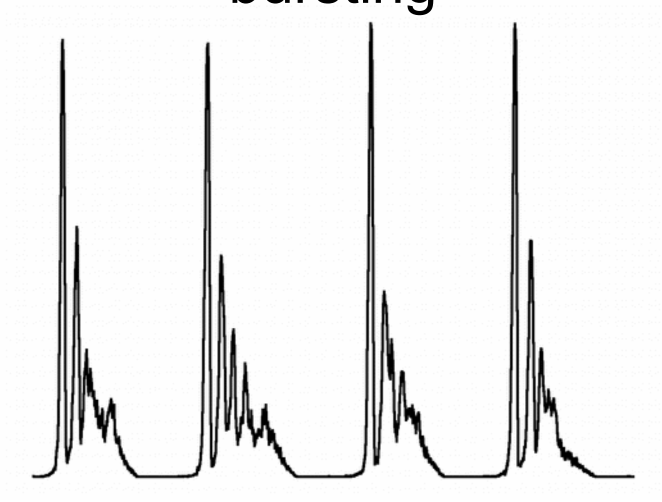
Calcium concentration oscillations in single rat hepatocytes stimulated with ATP (1.5 µM).
(data from: Pahle et al. (2008), *BMC Bioinformatics* **9**:139, doi:10.1186/1471-2105-9-139)

Gillespie algorithm

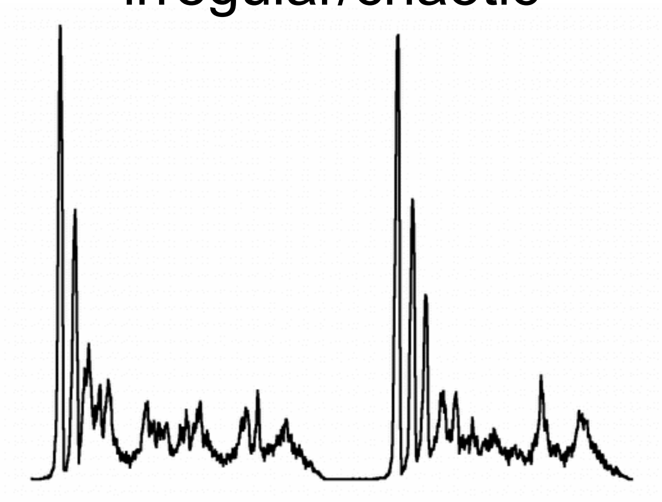
spiking



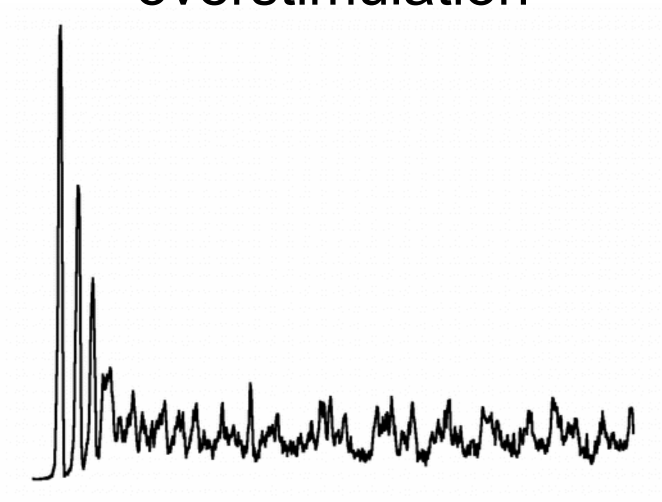
bursting



irregular/chaotic



overstimulation



“Gillespie St” by Jürgen Pahle CC BY-SA 4.0

Monte Carlo simulation algorithms

D.T. Gillespie (1976) *J. Comp. Phys.* **22**(4):403, doi:10.1016/0021-9991(76)90041-3

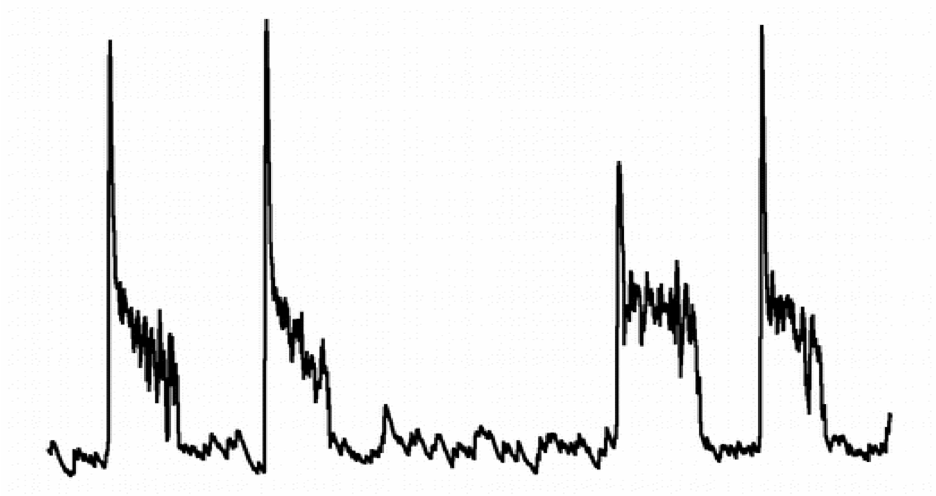
and extensions thereof...

Signal transduction via Calcium

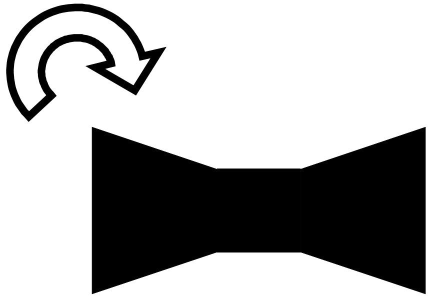
Hormones (angiotensin II, vasopressin etc.)

Nucleotides (ATP, UTP)

Ca^{2+}-signal



calcium code?



bow tie structure

Target proteins (calmodulin, phosphorylase b kinase, etc.)

Transcription factors (NF-κB, etc.)

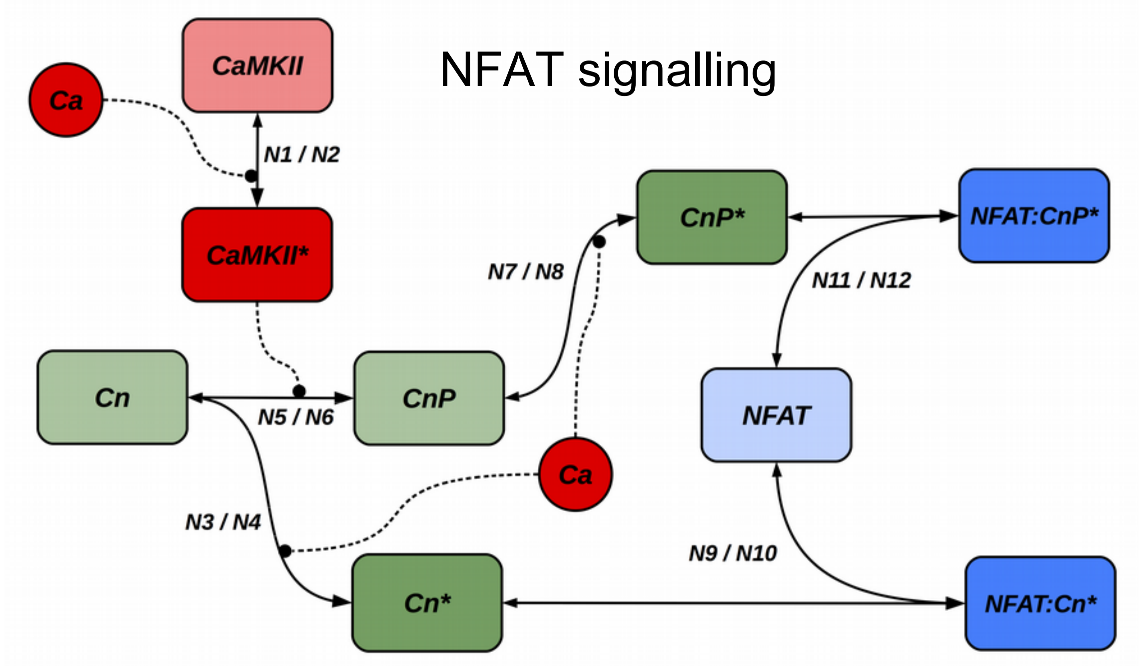
The calcium code

- Intracellular targets sense only the Ca^{2+}-concentration inside the cell (plus possibly other signalling pathways, “cross-talk”)
- How is the information from different hormones encoded in the calcium signal?
- How is the calcium signal decoded again at the target proteins?

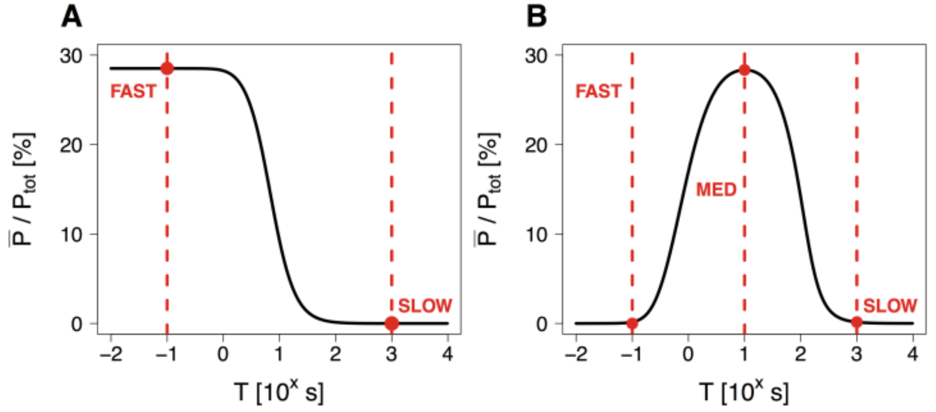
The calcium code

- Intracellular targets sense only the Ca^{2+}-concentration inside the cell (plus possibly other signalling pathways, “cross-talk”)
- How is the information from different hormones encoded in the calcium signal?
- How is the calcium signal decoded again at the target proteins?
- Specific information seems to be encoded in the
  - amplitude (AM)
  - frequency (FM)
  - duration
  - waveform

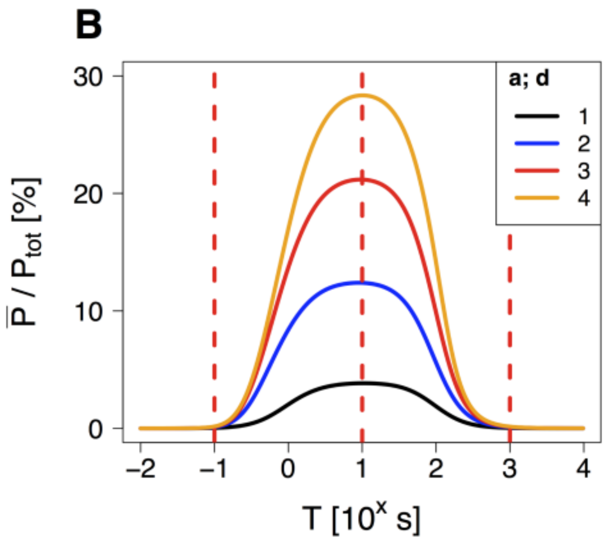
Band-pass activation

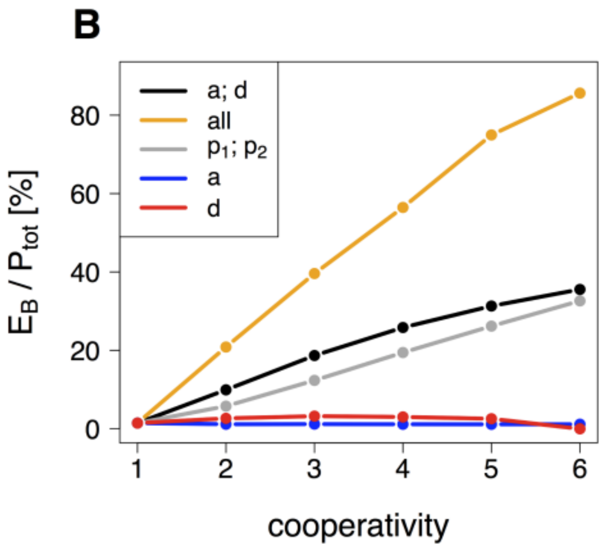


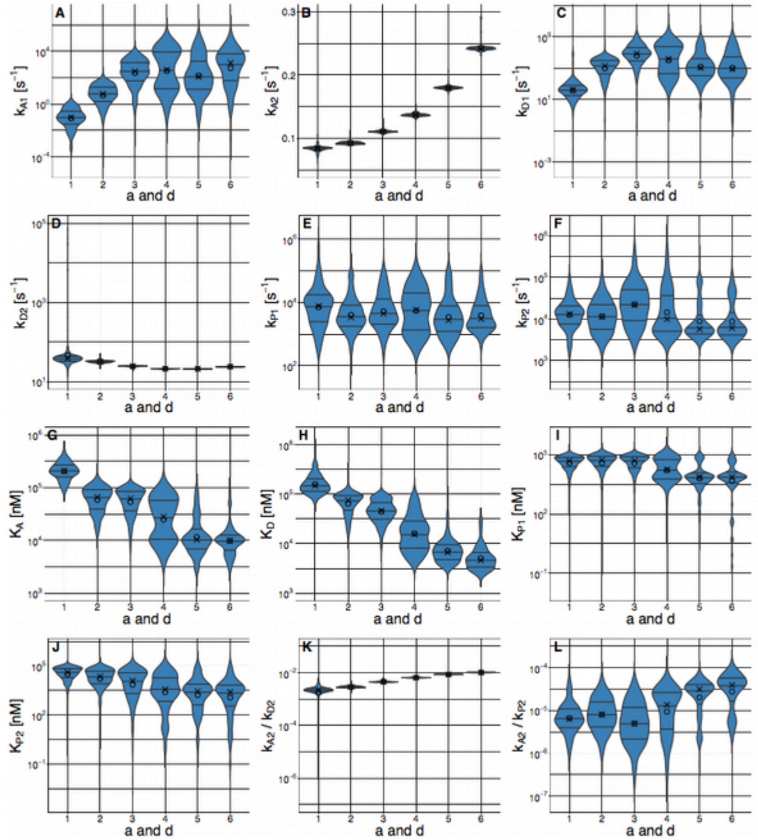
NFAT gets only activated by calcium oscillations of a certain frequency.



Arne Schoch







Signalling speeds

Telekom Broadband ~ 25 – 1000 Mbit/s

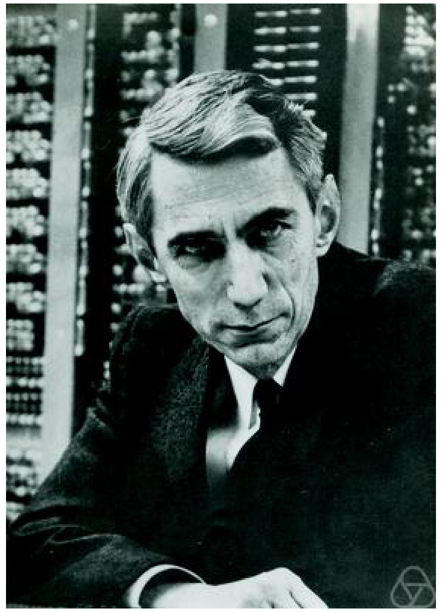
4G / LTE ~ 150 – 1200 Mbit/s

Biological signalling

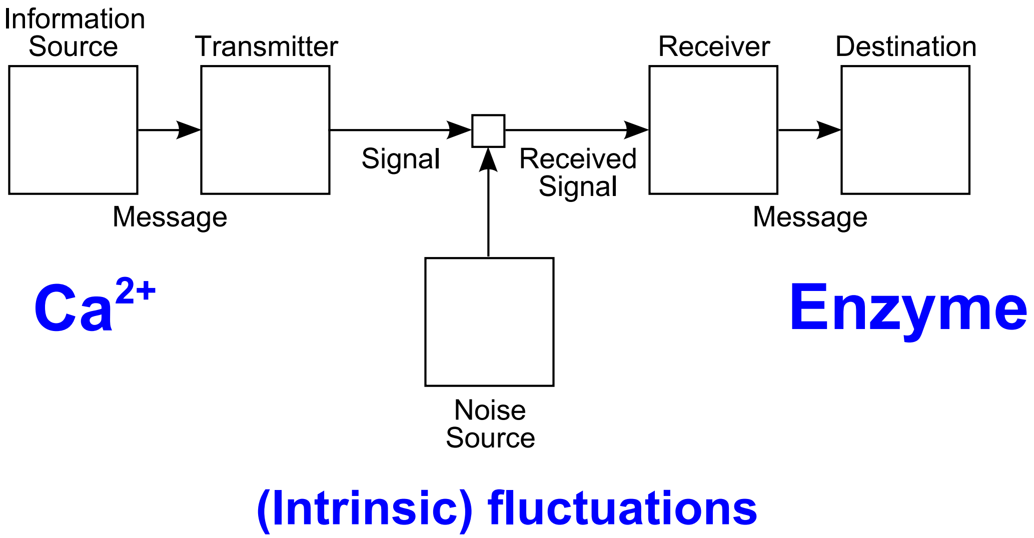
(Intracellular) calcium signalling operates at roughly 0.39 bit/s

Information theory

- Claude E. Shannon (1916-2001)
  "Mathematical Theory of Communication" (1948)
- Information theory can answer questions about limits of faithful information transfer over a given (noisy) channel etc.



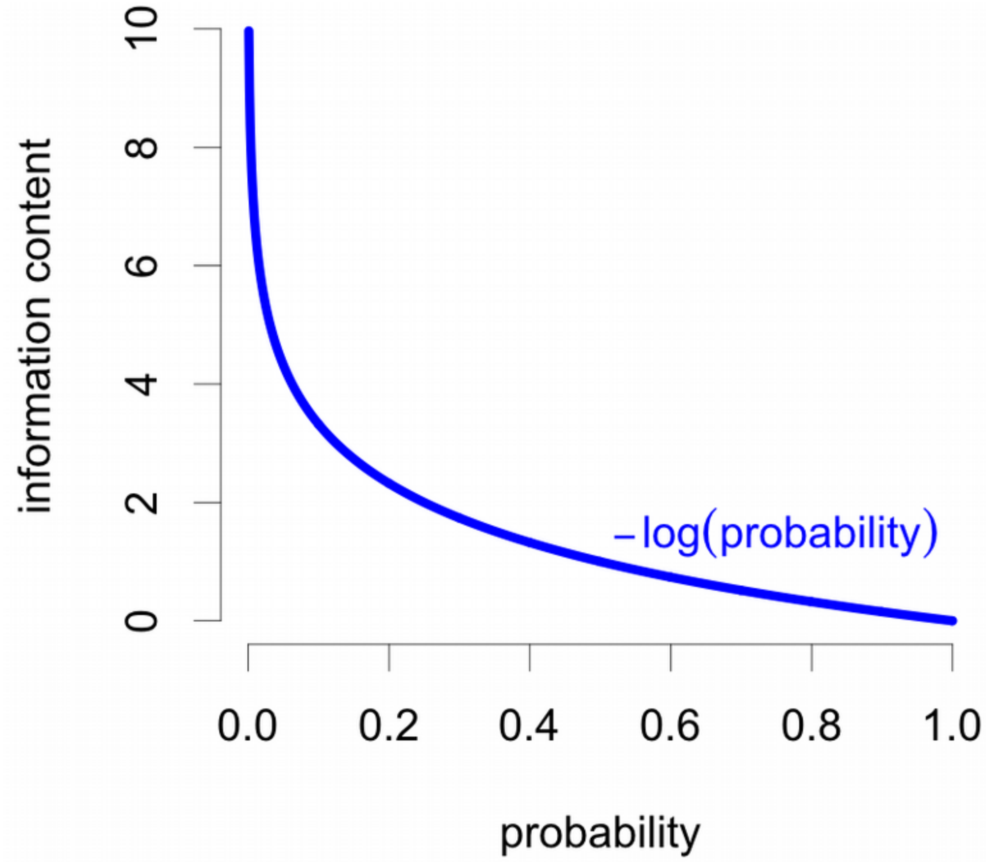
"Shannon, Claude" by
Konrad Jacobs CC BY-SA 2.0 DE



"Shannon communication system" by
Wanderingstan, public domain

Information theory 101:
How to quantify information?

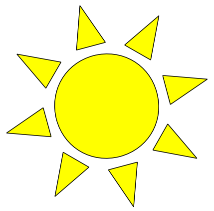
Information content of
an event
→ Uncertainty / Surprisal
(negative log of probability)



Average uncertainty of all possible
events → Entropy

Information = **Decrease** in uncertainty

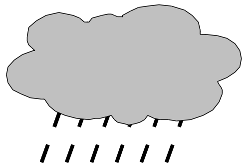
Weather example



50%

Probability(sunny) = ½

→ Uncertainty(sunny) = 1.0



50%

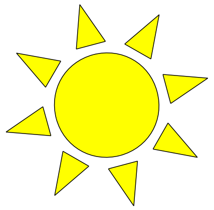
Probability(rainy) = ½

→ Uncertainty(rainy) = 1.0

On average (entropy of the weather)

→ 1.0 [bit/day]

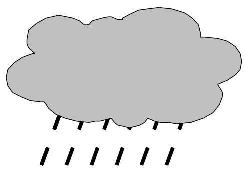
Weather example



100%

Probability(sunny) = 1.0

→ Uncertainty(sunny) = 0.0



0%

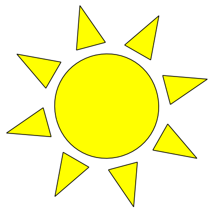
Probability(rainy) = 0

→ Uncertainty(rainy) = 0.0 per convention

On average (entropy of the weather)

→ 0.0 [bit/day]

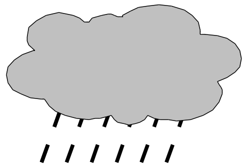
Weather example



80%

Probability(sunny) = 0.8

→ Uncertainty(sunny) = 0.32



20%

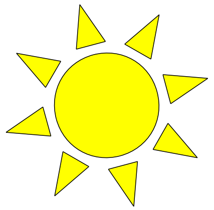
Probability(rainy) = 0.2

→ Uncertainty(rainy) = 2.32

On average (entropy of the weather)

→ 0.64 [bit/day]

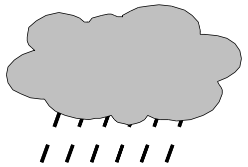
Weather example (Leipzig)



266/365

Probability(sunny) = 0.73

→ Uncertainty(sunny) = 0.32



99/365

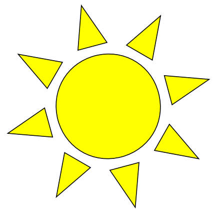
Probability(rainy) = 0.27

→ Uncertainty(rainy) = 1.88

On average (entropy of the weather in Leipzig)

→ 0.84 [bit/day]

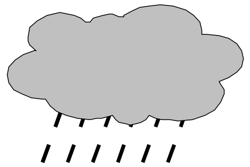
Weather example



p%

Probability(sunny) = p

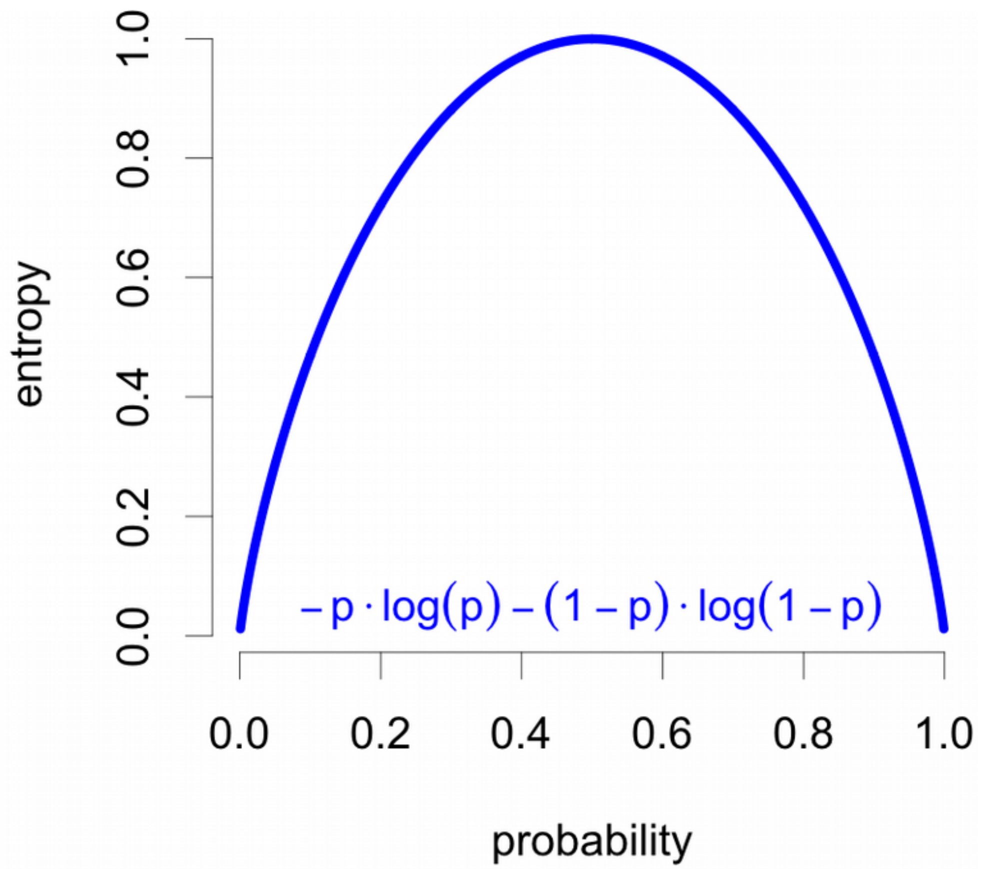
→ Uncertainty(sunny) = $-\log_2(p)$



(1-p)%

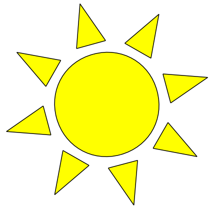
Probability(rainy) = 1-p

→ Uncertainty(rainy) = $-\log_2(1-p)$



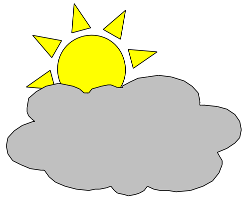
On average (entropy of the weather)
[bit/day]

Weather example



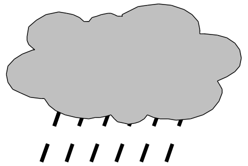
Probability(sunny) = 0.25

→ Uncertainty(sunny) = 2.0



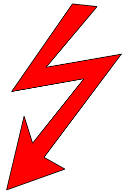
Probability(cloudy) = 0.25

→ Uncertainty(cloudy) = 2.0



Probability(rainy) = 0.25

→ Uncertainty(rainy) = 2.0



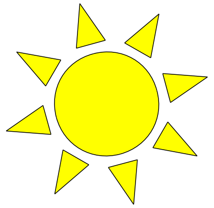
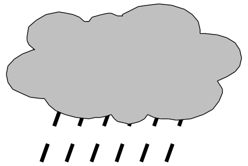
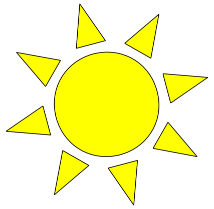
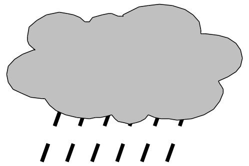
Probability(thunderstorm) = 0.25

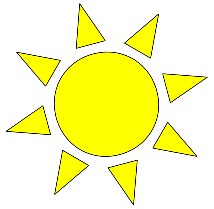
→ Uncertainty(thunderstorm) = 2.0

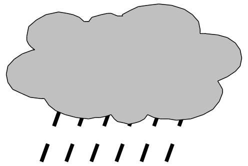
On average (entropy of the weather)

→ 2.0 [bit/day]

Weather dynamics

		
	0.75	0.25
	0.5	0.5

 ⅔

 ⅓

Markov process

Andrey A. Markov (1856-1922)

- Markov process can not remember former states, future is only dependent on the current state.
- Markovian modeling is used in a variety of fields:
  - Communication: Telephone system (Hidden Markov models)

  - Hard disks (error correction)

  - Language recognition
  - PageRank algorithm of Google
  - Biological modeling: Population dynamics, etc.
  - Games of chance (chutes and ladders)

Information/Entropy-rate

The information gained by observing tomorrow's weather, when the today's weather is known:

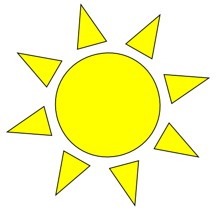
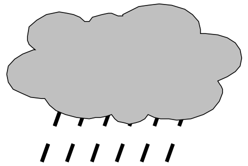
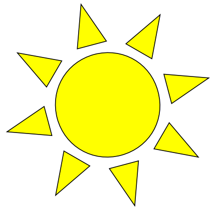
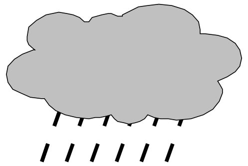
Entropy(tomorrow's weather | today's weather)
→ conditional probabilities

In our example:

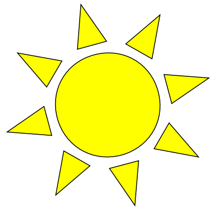
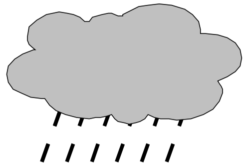
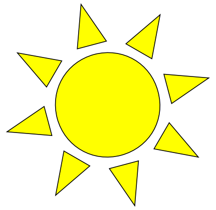
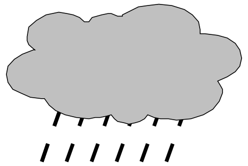
Entropy(tomorrow's weather) = 0.92 [bit/day]

Entropy(tomorrow's weather | today's weather) = 0.87 [bit/day]

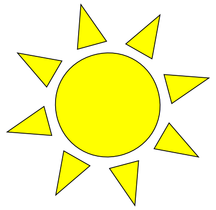
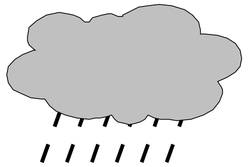
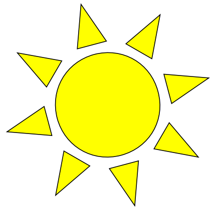
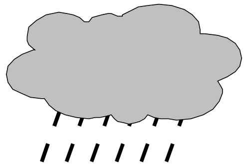
Weather dynamics

			
	high	0.9	0.1
	low	0.6	0.4
	high	1	0
	low	0	1

Weather dynamics (Leipzig)

26.12.2018, 10:00: 1029 hPa → high			
	high	0.9	0.1
	low	0.6	0.4
	high	1	0
	low	0	1

Weather dynamics (Leipzig)

26.12.2018, 10:00: 1029 hPa → high			
	high	0.9	0.1
	low	0.6	0.4
	high	1	0
	low	0	1

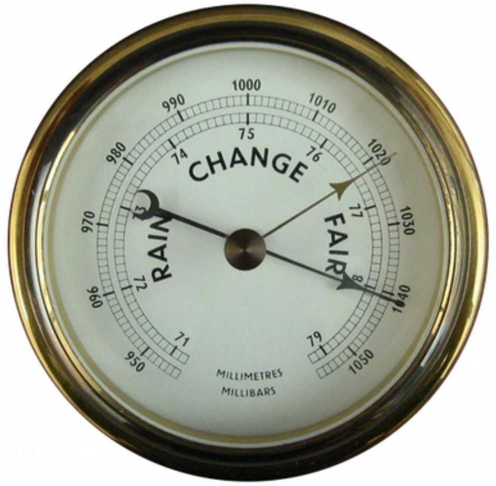
Information provided by the barometer

Information =

Uncertainty (without barometer)

minus

Uncertainty (with barometer)



Assumption Probability(high) = Probability(low) = 0.5

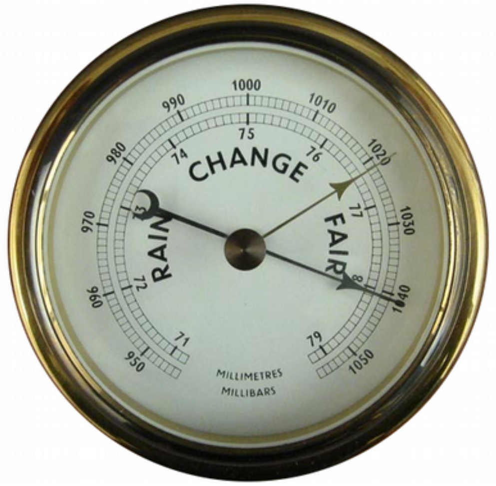
Information provided by the barometer

Information =

Uncertainty (without barometer)

minus

Uncertainty (with barometer)



Assumption Probability(high) = Probability(low) = 0.5

Information = 0.39 [bit/day]

Coming back to calcium

Enzyme dynamics is influenced by calcium
How much of the uncertainty about the enzyme dynamics is taken away, if we know the calcium signal?

→ information transferred
from calcium to target enzyme

Transfer Entropy

Quantifies the information transferred by calculating how much uncertainty is lost (or information gained) about a dynamic stochastic process, when the value of the driving signal is known

Kullback-Leibler-form

T. Schreiber (2000), *Phys. Rev.* **85**(2), 461

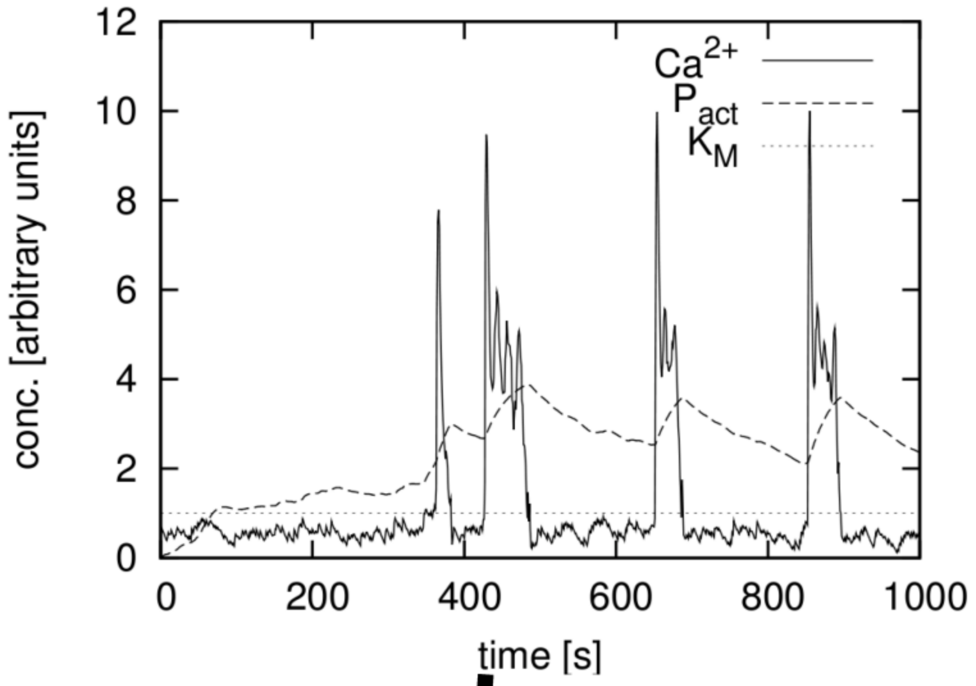
$$T_{J\rightarrow I} = \sum p\left(i_{n+1}, i_n^{(k)}, j_n^{(l)}\right) \log\left(\frac{p\left(i_{n+1} | i_n^{(k)}, j_n^{(l)}\right)}{p\left(i_{n+1} | i_n^{(k)}\right)}\right)$$

Practical problems:

- Signals are not discrete:
  Sums → integrals, probabilities → Probability densities
- Probability densities are not known → Estimation (kernel density, Kraskov, rank-based histogram,...)
- Confounding variables → Additional conditioning, influence of history length

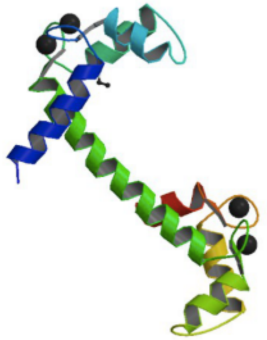
Coupled protein activation

stochastic coupling to a simulated protein



$$\frac{d[\mathrm{P}_{\mathrm{act}}]}{dt} = \frac{k_{\mathrm{act}} \cdot [\mathrm{Ca}^{2+}]^p}{K_M^p + [\mathrm{Ca}^{2+}]^p} \cdot [\mathrm{P}_{\mathrm{inact}}] - k_{\mathrm{inact}} \cdot [\mathrm{P}_{\mathrm{act}}]$$

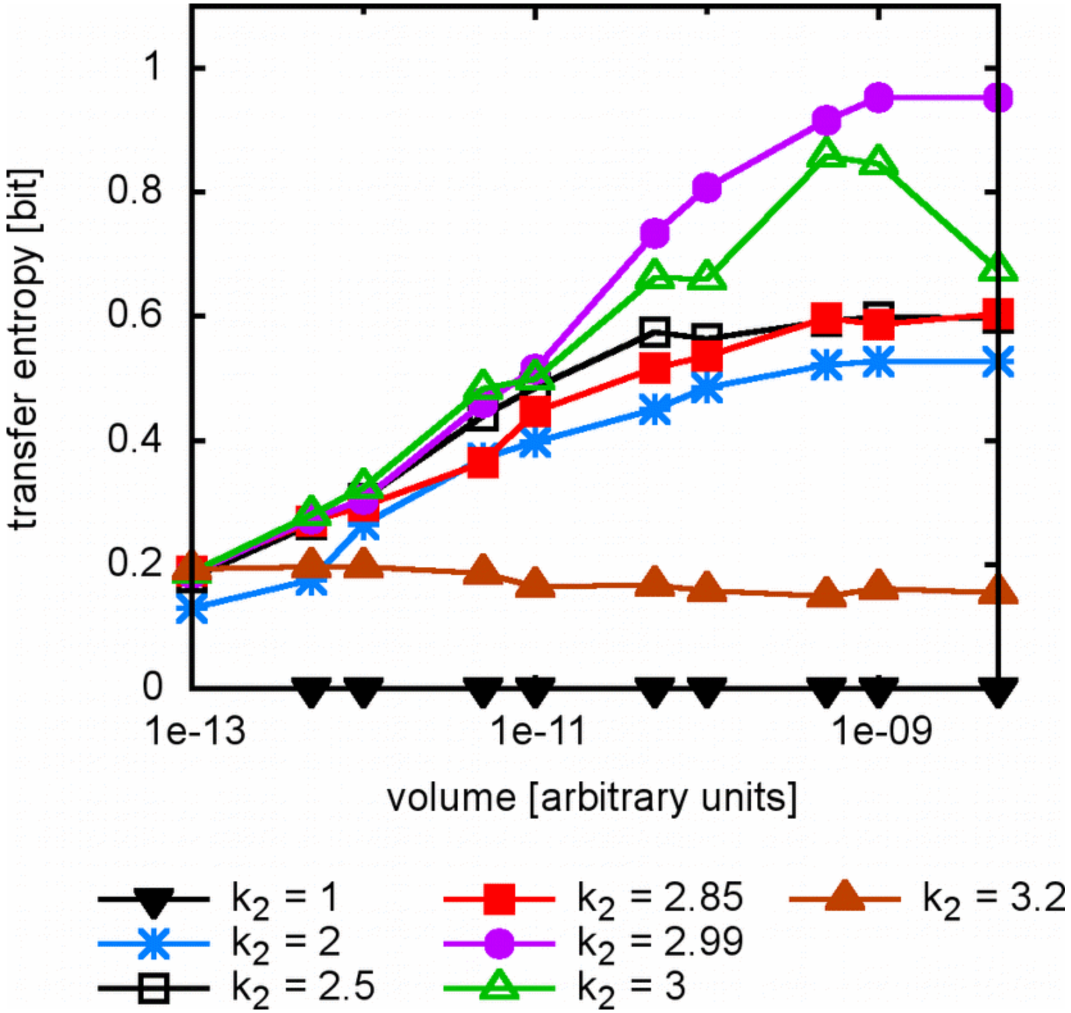
e.g. calmodulin



estimation of transfer entropy
(box-assisted) kernel-density estimation
using rectangular or Gaussian kernels
rank-based adaptive histogram method

"Calmodulin structure" by Chattopadhyaya et al. PUBMED:1474585 at RCSB PDB

Information transfer increases with system size

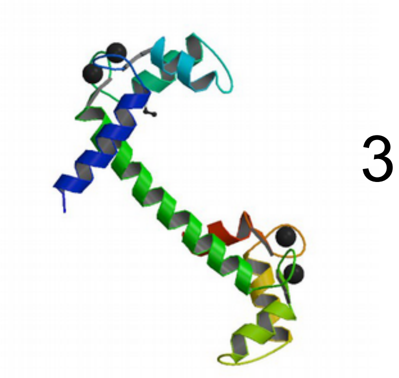


Rate of information transfer increases with increasing system size (particle number)
Bursting: max. rate ~ 0.6 bit/sample

Slight increase in transfer entropy (TE) from spiking to more complex bursting oscillations
In the case of a (elevated) steady state drop to very low values

k_2	Dynamics	TE
1	under-stimulation	0.00
2	spiking	0.52
2.5	Bursting	0.59
2.85	Bursting	0.60
2.99	elevated oscillations	0.95
3.2	elevated steady-state	0.15

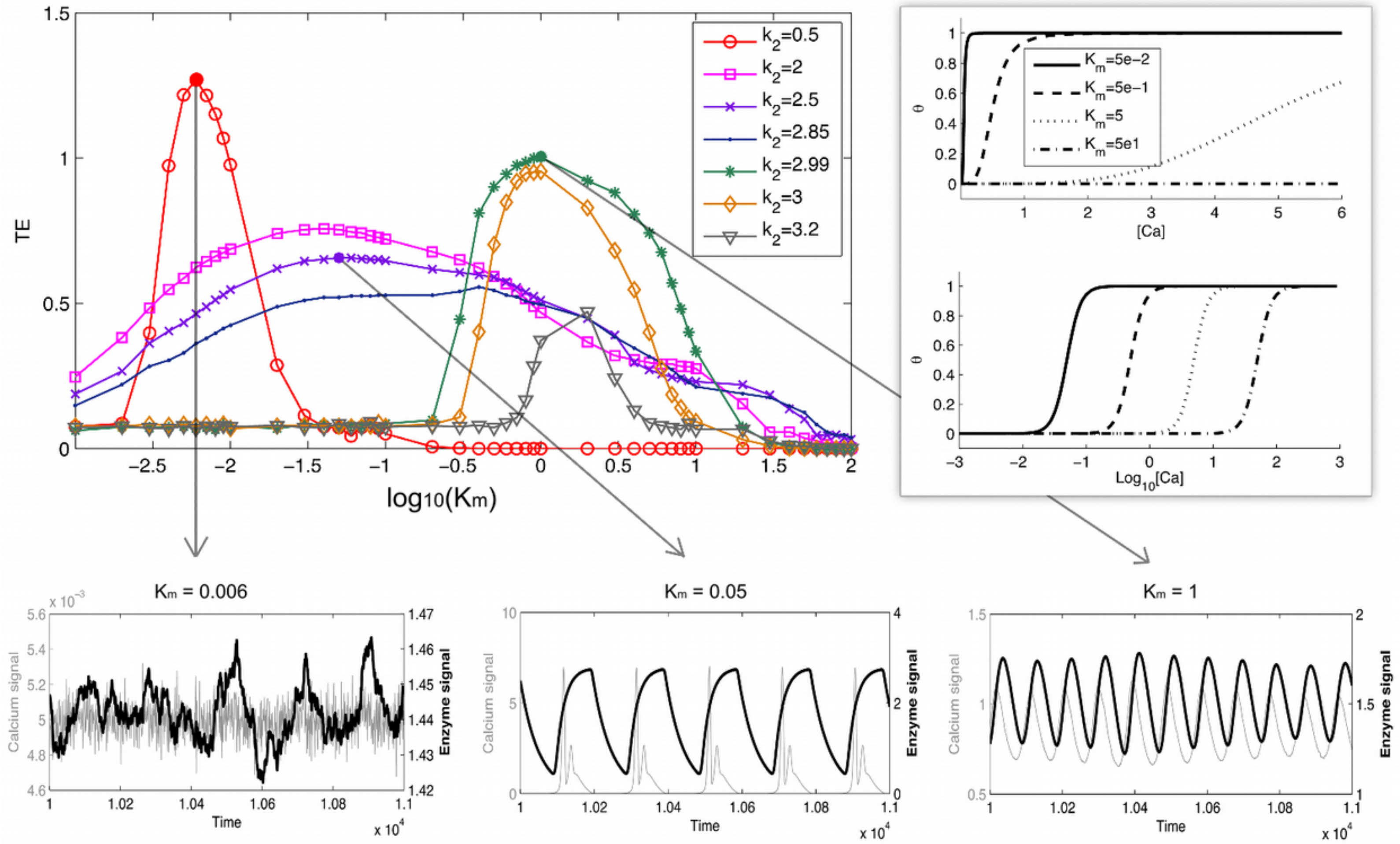
Proteins can be tuned to certain characteristics of the calcium input



3D structure

Dynamic behaviour of protein

Which feature of the calcium input influences the activation of a protein most?
→ Differential regulation of proteins!

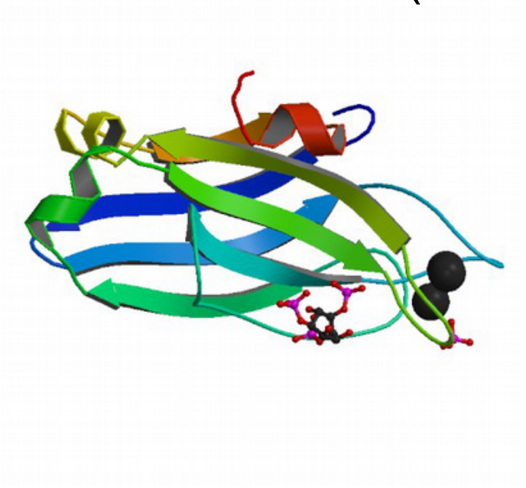


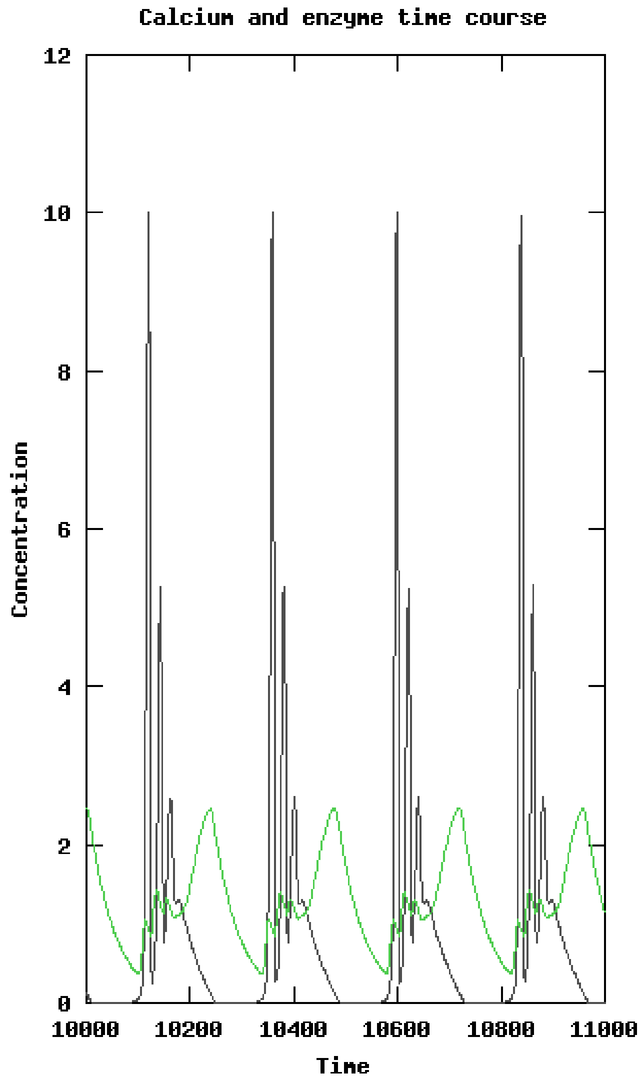
Activation of differently parameterised calmodulin-like proteins by different calcium input oscillations

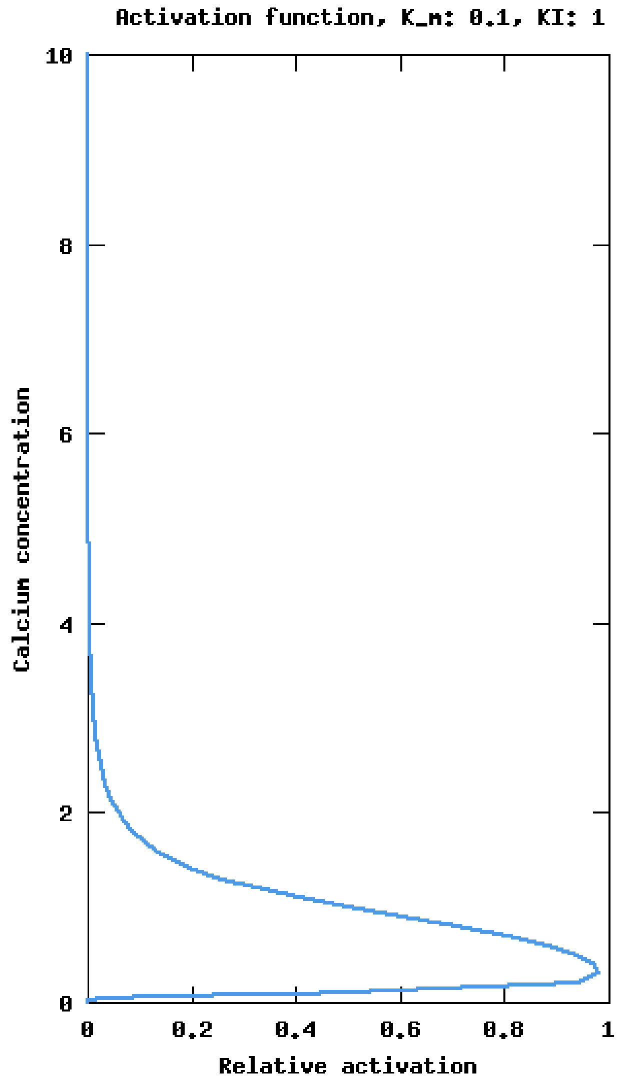
"Calmodulin structure" by Chattopadhyaya et al. PUBMED:1474585 at RCSB PDB

More complex proteins allow higher information transfer (additional inhibitory mechanism)

e.g. Protein kinase C (PKC-α)



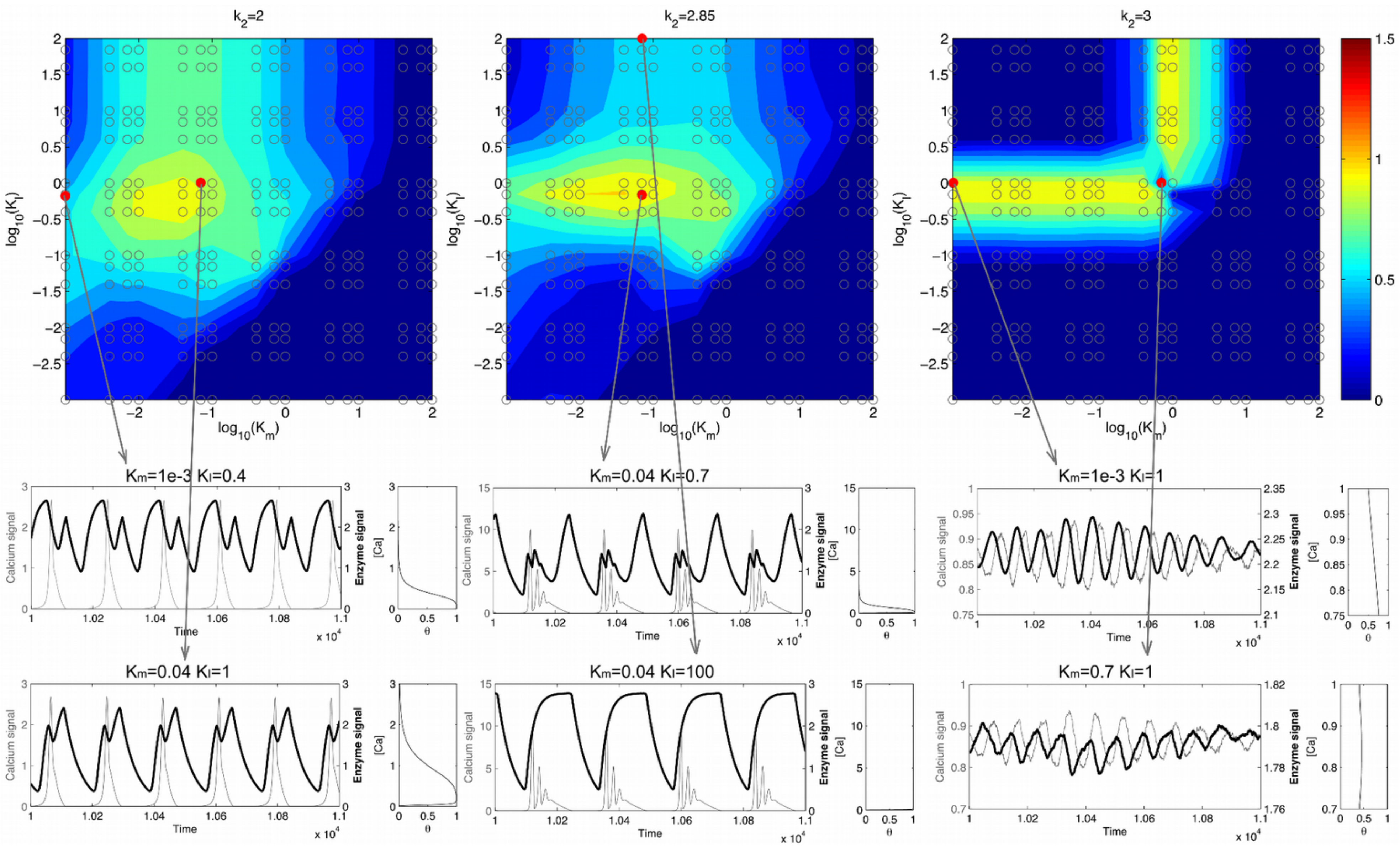




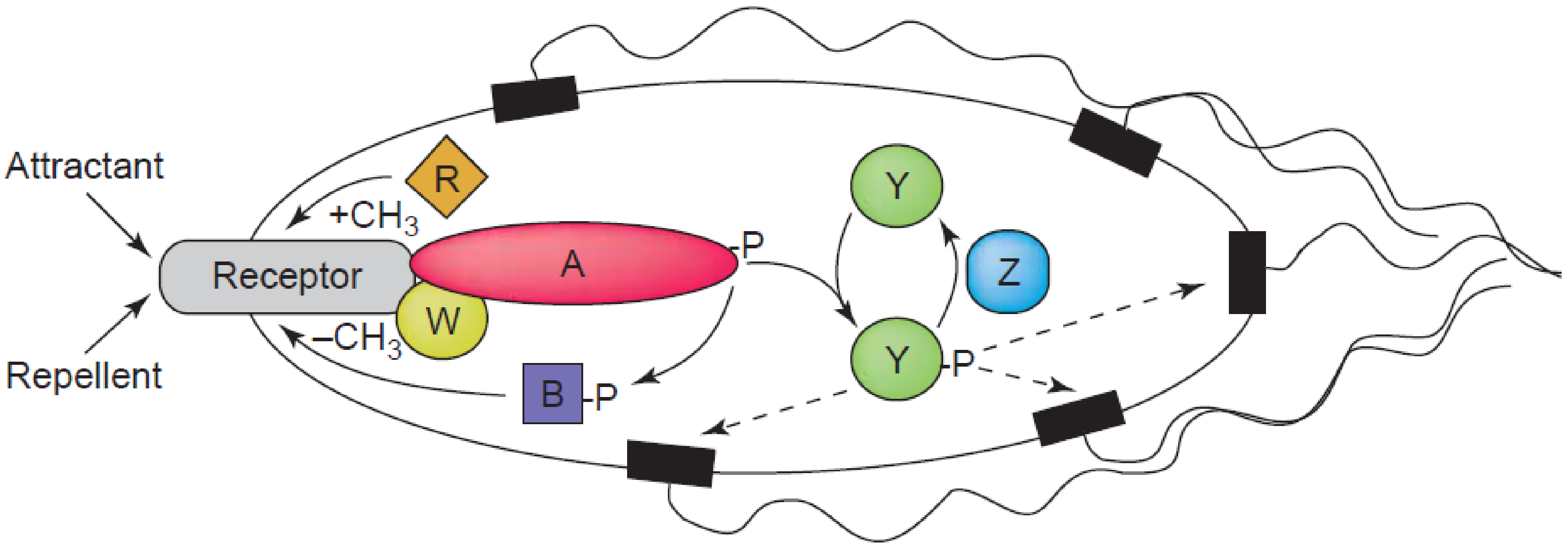
$$\frac{d[\mathrm{P_{act}}]}{dt} = \frac{k_{\mathrm{act}} \cdot [\mathrm{Ca}^{2+}]^p}{K_M^p + [\mathrm{Ca}^{2+}]^p} \cdot \frac{K_I^p}{[\mathrm{Ca}^{2+}]^p + K_I^p} \cdot [\mathrm{P_{inact}}] - k_{\mathrm{inact}} \cdot [\mathrm{P_{act}}]$$

"PKC structure" by Verdaguer et al. PUBMED:10562545 at RCSB PDB

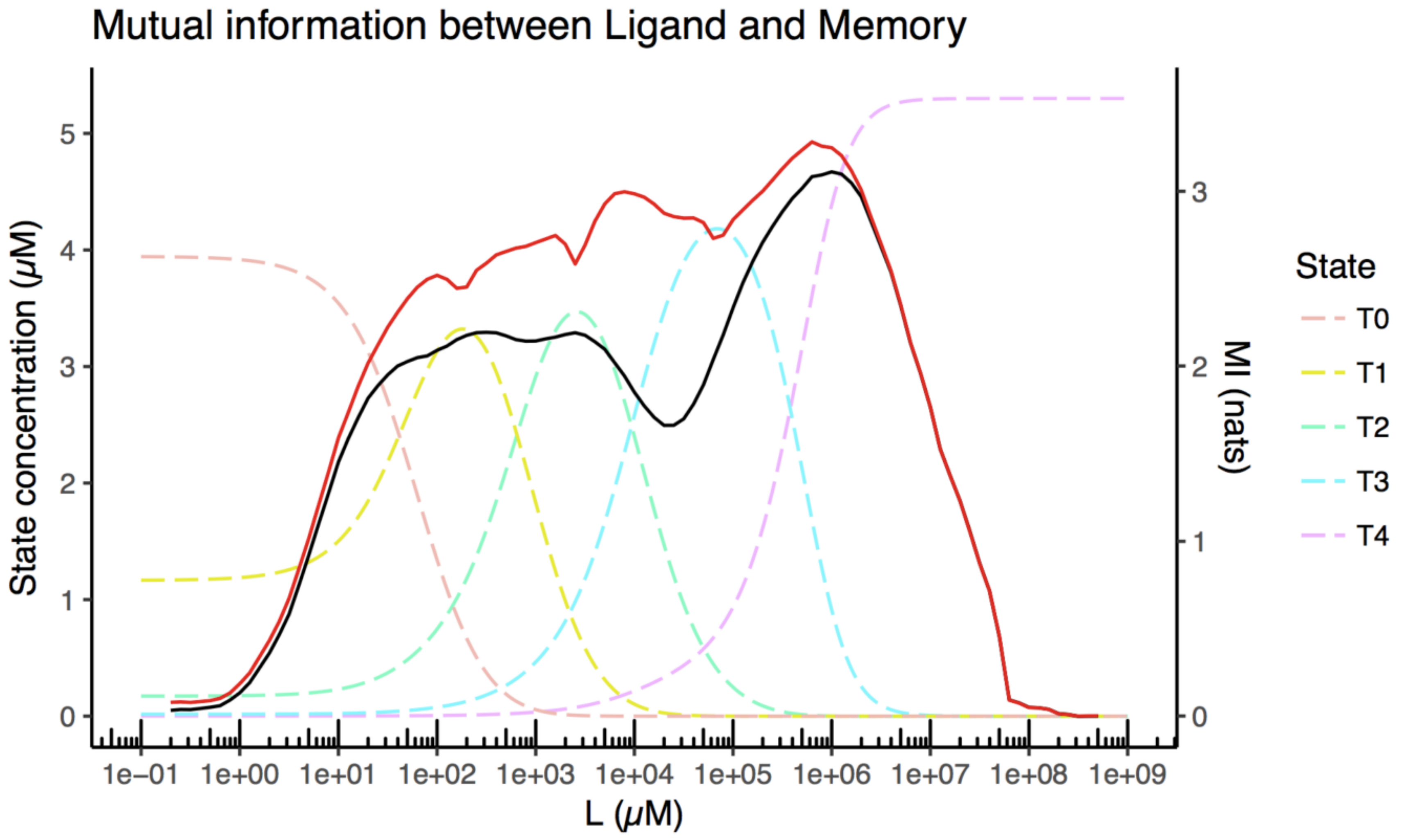
More complex proteins allow higher information transfer (additional inhibitory mechanism)



Memory and information flow in *E. coli* chemotaxis



E. coli memory:
Internal representation of outside nutrient (ligand) levels

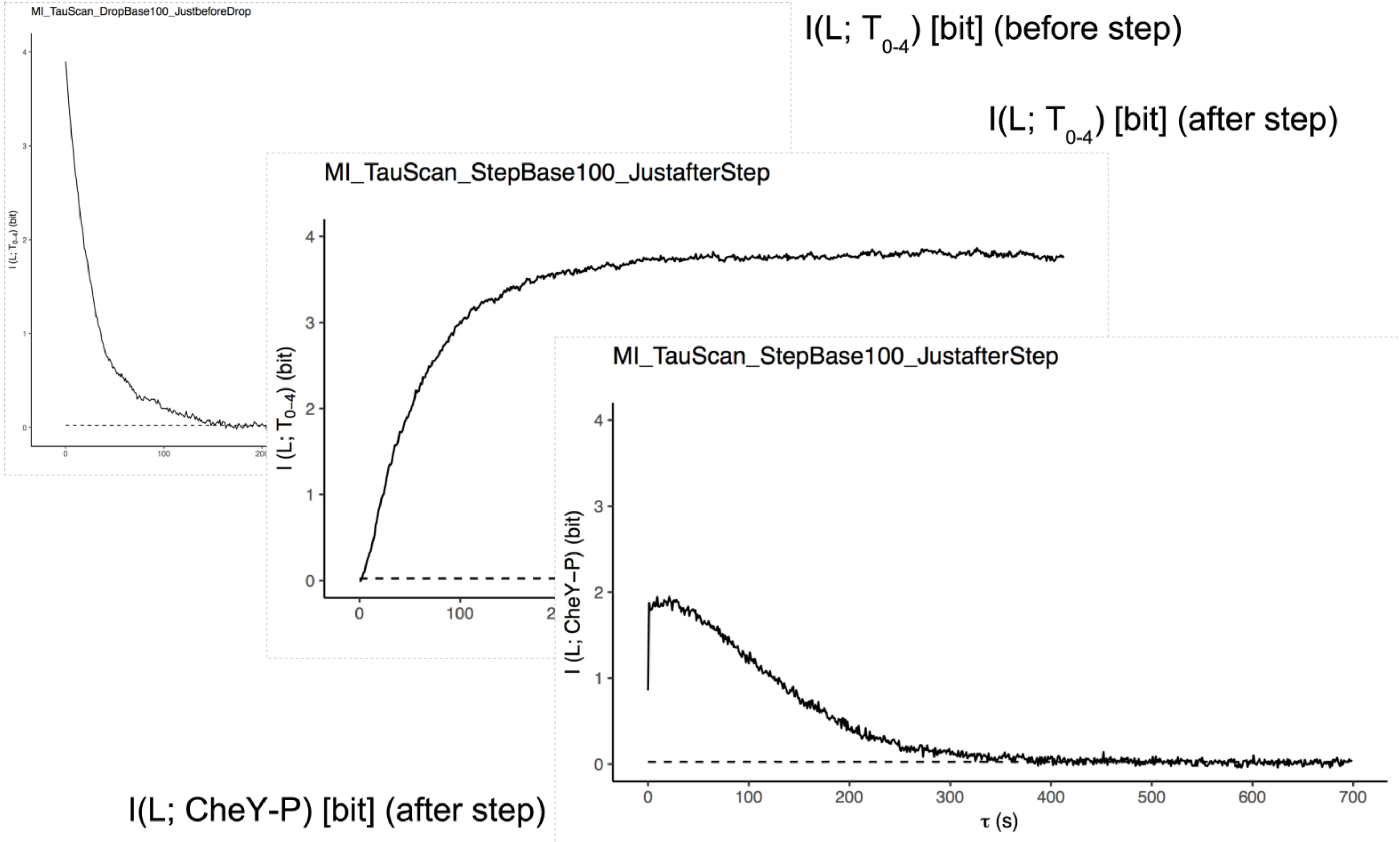


Memory is in the methylation levels (T0 – T4) of the receptors

Time- scales of memory building / forgetting / adaptation

$I(L; T_{0\text{-}4})$ [bit] (before step)

$I(L; T_{0\text{-}4})$ [bit] (after step)



$I(L; CheY\text{-}P)$ [bit] (after step)

Conclusion

- Information theory can be applied in biology. For instance, this allows to quantify the performance of cellular signalling systems under different conditions.
- Targets of signalling systems can be tuned to the input, such that communication channels are effectively turned on or off depending on the shape of the input.
- Study what features of biological signals carry information (that is relevant for a disease and how to target these) during information transfer (communication in space) and memory (communication in time).
- Estimating and interpreting information-theoretic functionals can be a tricky business.

Acknowledgements

BIP group @ Uni Heidelberg: Arne Schoch, Martin Zauser, Sarah Kaspar, Irina Surovtsova, Jonas Förster, …

Yujiang Wang, Aaròn Jimenez Vázqez Jiménez

Copasi team: Pedro Mendes (University of Connecticut, USA), Ursula Kummer (Uni Heidelberg), Sven Sahle (Uni Heidelberg), Frank Bergmann (Uni Heidelberg), Stefan Hoops (University of Virginia, USA) ...

Anne K. Green (Warwick, UK) and C. Jane Dixon (Leicester, UK)

Funding: BIOMS

Questions?

Contact details

Dr. Jürgen Pahle
Biological Information Processing Group
BIOMS / BioQuant
Heidelberg University

Links

lab.pahle.org (BIP group website)

www.copasi.org (biochemical modelling and simulation software)

jpahle.github.io/CoRC (scripting for COPASI)

OscillatorGenerator (R-package for the generation of artifical oscillatory input signals)

Literature

J. Pahle, A.K. Green, C.J. Dixon and U. Kummer (2008) Information transfer in signaling pathways: a study using coupled simulated and experimental data. *BMC Bioinformatics* **9**:139, doi:10.1186/1471-2105-9-139

A. Schoch and J. Pahle (2018) Requirements for band-pass activation of Ca^{2+}-sensitive proteins such as NFAT. Biophysical Chemistry in press, doi:10.1016/j.bpc.2018.10.005

S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes and U. Kummer (2006) COPASI – a COmplex PAthway SImulator. *Bioinformatics* **22**(24):3067, doi:10.1093/bioinformatics/btl485

F.T. Bergmann, S. Hoops, B. Klahn, U. Kummer, P. Mendes, J. Pahle and S. Sahle (2017) COPASI and its Applications in Biotechnology. *Journal of Biotechnology* **261**:215-220, doi:10.1016/j.jbiotec.2017.06.1200

J. Pahle. Biochemical simulations: stochastic, approximate stochastic and hybrid approaches (2009) *Briefings in Bioinformatics* **10**(1):53-64, doi:10.1093/bib/bbn050

P.H.G.M. Willems, J. Pahle, X.L. Stalpers, D. Mugahid, A. Nikolaew, W.J.H. Koopman and U. Kummer (2015) PKC-mediated inhibitory feedback of the cholecystokinin 1 receptor controls the shape of oscillatory Ca^{2+} signals. *FEBS Journal* **282**(11):2187, doi:10.1111/febs.13267