V12: Structure Learning in Bayesian Networks II

Roughly speaking, there are 3 approaches to learning
without a prespecified structure:

(1) constraint-based structure learning

Finds a model that best explains the dependencies/independencies in the data.

(2) **Score-based** stucture learning (today)

We define a hypothesis space of potential models and a scoring function
that measures how well the model fits the observed data.
Our computational task is then to find the highest-scoring network.

(3) **Bayesian model averaging** methods

Generates an ensemble of possible structures.

Bayesian Structure Scores

There are 2 obvious choices of scoring functions:

- Maximum likelihood parameters (V11)
- Bayesian scores (today)

Bayesian Score

The main principle of the Bayesian approach is that whenever we have **uncertainty** over anything, we should place a **distribution** over it.

Here, we are uncertain both about the **structure** and **parameters** of the network.

We define a **structure prior** P(G) that puts a prior probability on different graph structures, and
a **parameter prior** $P(\theta_G|G)$ that puts a probability on different choices of parameters once the graph is given.

Bayesian Score

By Bayes rule, we have

$$P(G|D) = \frac{P(D|G)P(G)}{P(D)}$$

The denominator P(D) is simply a normalizing factor that does not help distinguishing between different structures.

Thus, we define the **Bayesian score** as

$$score_B(G:D) = \log P(D|G) + \log P(G)$$

The ability to ascribe a prior P(G) over structures gives us a way of preferring some structures over others.

E.g. we could penalize dense structures more than sparse ones.

However, the structure-prior term is almost irrelevant compared to the first term.

Bayesian Score

The first term P(D|G) takes into consideration our uncertainty over the parameters:

$$P(D|G) = \int_{\theta_G} P(D|\theta_G, G) P(\theta_G|G) d\theta_G$$

$P(D|\theta_G, G)$ is the **likelihood** of the data given the network $\langle G, \theta_G \rangle$.

$P(\theta_G|G)$ is our prior distribution over different parameter values for the network G.

P(D | G) is called the **marginal likelihood** of the data given the structure.

The maximum likelihood score would return the maximum of this function.

In contrast, the marginal likelihood is the average value of this function.

Why does the Bayesian score avoid overfitting ?

Bayesian Score

For this we examine the **sensitivity** of the likelihood to the particular choice of parameters.

In V11, we discussed that the maximal likelihood is overly „optimistic“ in its evaluation of the score:

it evaluates the likelihood of the training data using the best parameter values for the given data.

This is unrealistic.

Bayesian Score

The Bayesian approach tells us that, although the choice of parameter $\hat{\theta}$ is the most likely given the training set D, it is not the only choice.

The posterior over parameters provides us with a range of parameter choices, along with a measure how likely each of them is.

By integrating $P(D|\theta_G, G)$ over the different choices of parameters θ_G ,

$$P(D|G) = \int_{\theta_G} P(D|\theta_G, G)P(\theta_G|G)d\theta_G$$

we are measuring the expected likelihood, averaged over different possible choices of θ_G.

Thus, we are being more conservative.

Bayesian Score

Using the chain rule for probabilities, we can rewrite the marginal likelihood as

$$P(D|G) = \prod_{m=1}^{M} P(\xi[m]|\xi[1], \ldots, \xi[m-1], G)$$

Each of the terms in this product is the probability of the m‘th instance using the parameters learned from the first m – 1 instances (using Bayesian estimation).

Thus, the m‘th instance provides us with one data point for testing the ability of our model to predict a new data instance, based on the model learned from the previous one.

The Bayesian score does not depend on the order of instances.

Bayesian Score

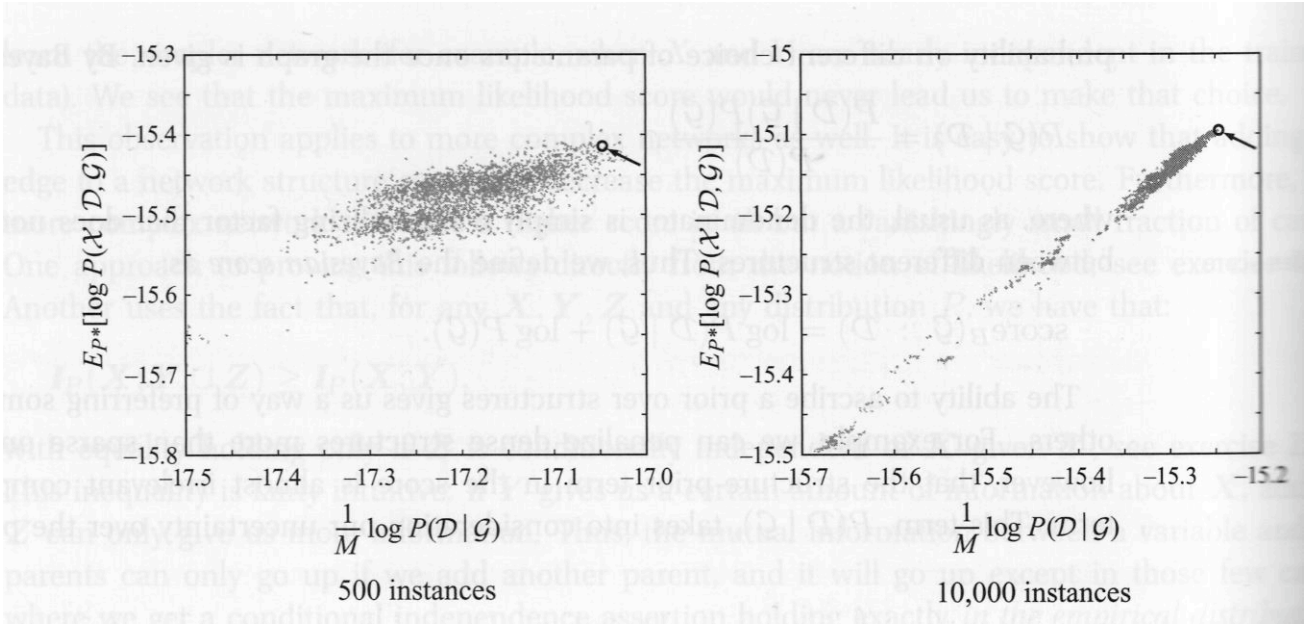
This suggests that

$$\frac{1}{M}\log P(D|G) \approx E_P[\log P(X|G,D)]$$

is an estimator for the average log-likelihood of a new sample X from the distribution P*.

In practice, it turns out that for reasonable sample sizes this is indeed a fairly good estimator of the ability of a model to generalize to unseen data.

Bayesian Score



Comparison of the average log-marginal-likelihood per sample in training data (x-axis) to the expected log-likelihood of new samples from the underlying distribution (y-axis).

Each point corresponds to a network structure, the true network is marked by a circle.

Marginal likelihood for a single variable

We start by examining how to compute the marginal likelihood for simple cases.

Consider a simple binary random variable X, and assume
that we have a prior distribution $Dirichlet(\alpha_1, \alpha_0)$ over X (see V10, p.20).

Consider a data set D with M[1] heads and M[0] tails.

Then the maximum likelihood value given D is

$$P(D|\hat{\theta}) = \left(\frac{M[1]}{M}\right)^{M[1]} \cdot \left(\frac{M[0]}{M}\right)^{M[0]}$$

Marginal likelihood for a single variable

Now let us consider the marginal likelihood.

For this, we need to compute the probability P(X[1], …, X[M])
of the data given our prior.

We can do this e.g. using the chain rule:

$$P(x[1], \ldots, x[M]) = P(x[1]) \cdot P(x[2]|x[1]) \cdot \cdots \cdot P(x[M]|x[1], \ldots, x[M-1])$$

Recall from V10 (p.14) that if we used a Beta prior, then

$$P(x[m+1] = H|x[1], \ldots, x[m]) = \frac{M^m[1] + \alpha_1}{m + \alpha}$$

where $M^m[1]$ is the number of heads in the first *m* examples.

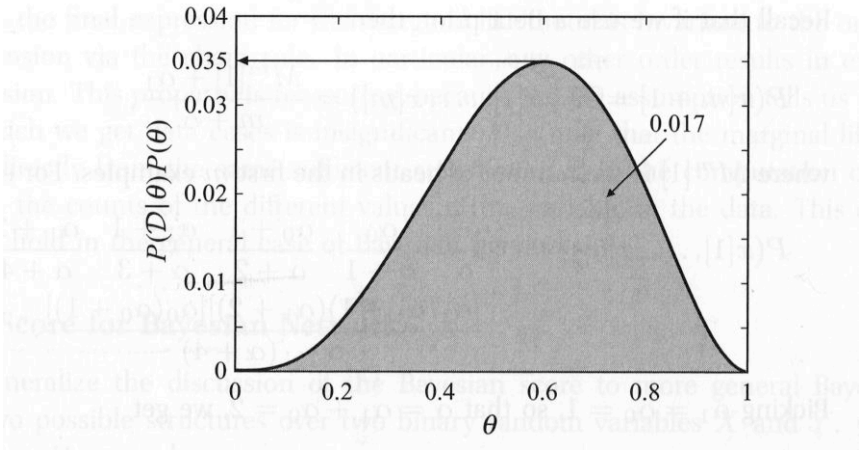
Marginal likelihood for a single variable

For example, if D = ⟨ H, T, T, H, H⟩,

$$P(x[1],\ldots,x[5]) = \frac{\alpha_1}{\alpha}\cdot\frac{\alpha_0}{\alpha+1}\cdot\frac{\alpha_0+1}{\alpha+2}\cdot\frac{\alpha_1+1}{\alpha+3}\cdot\frac{\alpha_1+2}{\alpha+4}$$

$$= \frac{\alpha_1(\alpha_1+1)(\alpha_1+2)\alpha_0(\alpha_0+1)}{\alpha\ldots(\alpha+4)}$$

Picking e.g. $\alpha_1 = \alpha_0 = 1$ so that $\alpha = \alpha_1 + \alpha_0 = 2$, we get $\frac{1\cdot2\cdot3\cdot1\cdot2}{2\cdot3\cdot4\cdot5\cdot6} = \frac{12}{720} = 0.017$



This is significantly lower than the likelihood

$$\left(\frac{3}{5}\right)^3\cdot\left(\frac{2}{5}\right)^2 = \frac{108}{3125} \approx 0.035$$

Thus, a model using maximum-likelihood parameters ascribes a much higher probability to the sequence than does the marginal likelihood.

Marginal likelihood for a single variable

In general, for a binomial distribution with a Beta prior, we have

$$P(x[1], \ldots, x[M]) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + M)} \cdot \frac{\Gamma(\alpha_1 + M[1])}{\Gamma(\alpha_1)} \cdot \frac{\Gamma(\alpha_0 + M[1])}{\Gamma(\alpha_0)}$$

A similar formula holds for a multinomial distribution over the space $x^1, \ldots, x^k$ with a Dirichlet prior with hyperparameters $\alpha_1, \ldots, \alpha_k$:

$$P(x[1], \ldots, x[M]) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + M)} \cdot \prod_{i=1}^{k} \frac{\Gamma\left(\alpha_i + M\left[x^i\right]\right)}{\Gamma(\alpha_i)}$$

Bayesian score for Bayesian networks

We now generalize the discussion of the Bayesian score to more general BNs.

Consider 2 possible structures of 2 binary random variables X and Y.

G_0 is the graph with no edges. Here, we have

$$P(D|G_0) = \int_{\Theta_X \times \Theta_Y} P(D|\Theta_X, \Theta_Y, G_0) P(\Theta_X, \Theta_Y|G_0) d[\Theta_X, \Theta_Y]$$

The likelihood term $P(D|\Theta_X, \Theta_Y, G_0)$ can be written as a product of terms,
one involving Θ_X and the observations of X in the data,
and the other involving Θ_Y and the observations of Y in the data.

Bayesian score for Bayesian networks

If we also assume parameter independence, that is that $P(\Theta_X, \Theta_Y | G_0)$ decomposes as a product $P(\Theta_X | G_0)\ P(\Theta_Y | G_0)$ then we can simplify the integral

$$P(D|G_0) = \int_{\Theta_X \times \Theta_Y} P(D|\Theta_X, \Theta_Y, G_0) P(\Theta_X, \Theta_Y | G_0) d[\Theta_X, \Theta_Y]$$

$$= \left(\int_{\Theta_X} P(\Theta_X | G_0) \prod_m P(x[m] | \Theta_X, G_0) d\Theta_X \right) \left(\int_{\Theta_Y} P(\Theta_Y | G_0) \prod_m P(y[m] | \Theta_Y, G_0) d\Theta_Y \right)$$

Since we assumed parameter independence, we wrote the integral over a product of independent functions as the product of the integrals of the functions.

Each of the 2 integrals is the marginal likelihood of a single variable.

Bayesian score for Bayesian networks

Now consider the network $G_{X \to Y}$ = (X → Y)

If we again assume parameter independence, we can decompose this integral into a product of 3 integrals, each over a single parameter family.

$$P(D|G_{X\to Y}) = \left(\int_{\Theta_X} P(\Theta_X|G_{X\to Y}) \prod_m P(x[m]|\Theta_X, G_{X\to Y}) d\Theta_X \right)$$

$$\left(\int_{\Theta_{Y|x^0}} P\left(\Theta_{Y|x^0}|G_{X\to Y}\right) \prod_{m:x[m]=x^0} P\left(y[m]\left|\Theta_{Y|x^0}, G_{X\to Y}\right.\right) d\Theta_{Y|x^0} \right)$$

$$\left(\int_{\Theta_{Y|x^1}} P\left(\Theta_{Y|x^1}|G_{X\to Y}\right) \prod_{m:x[m]=x^1} P\left(y[m]\left|\Theta_{Y|x^1}, G_{X\to Y}\right.\right) d\Theta_{Y|x^1} \right)$$

Each term can be written using the closed form solution, see Dirichlet prior (p.13).

Bayesian score for Bayesian networks

Comparing the marginal likelihood of the two structures G_0 and $G_{X \to Y}$, we see that the term that corresponds to X is similar in both.

In fact, the terms $P(x[m]|\Theta_X, G_{X \to Y})$ and $P(x[m]|\Theta_X, G_0)$ both make the same predictions given the parameter values.

Thus, if we choose the prior $P(\Theta_X|G_0)$ to be the same as $P(\Theta_X|G_{X \to Y})$, the first terms in the marginal likelihood of both structures are identical.

Under this assumption, the difference between the two marginal likelihoods is due to the difference between the marginal likelihood of the observations of Y and the marginal likelihoods of the observations of Y when we partition our examples based on the observed value of X.

Let us now consider what effect the complexity of the network has.

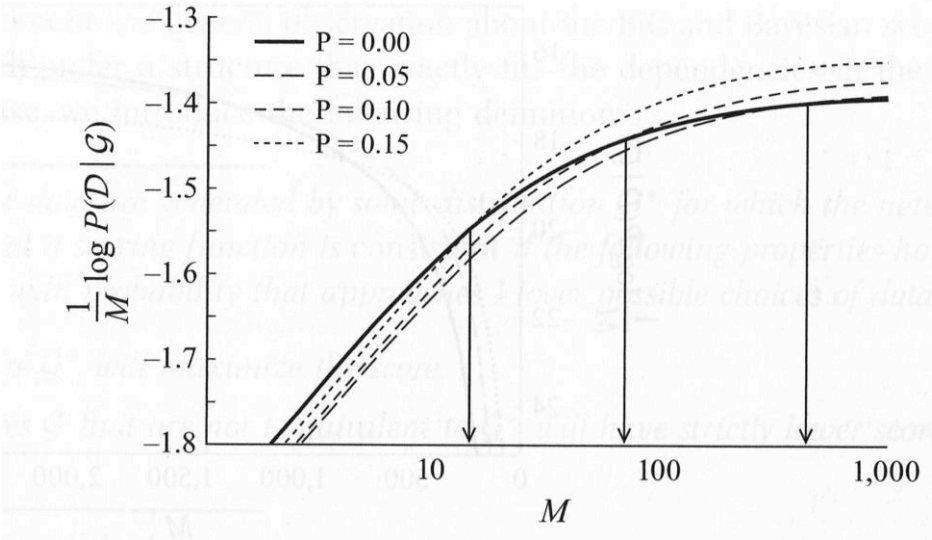
Bayesian score for Bayesian networks

We consider an idealized experiment where the empirical distribution is such that $P(x^1) = 0.5$ and $P(y^1 \mid x^1) = 0.5 + p$ and $P(y^1 \mid x^0) = 0.5 - p$.

p is a free parameter.
Larger values of p imply stronger dependence between X and Y.

The marginal distributions of X and Y are the same regardless of the value of p.
Thus, the score of the empty graph G_0 does not depend on p.

But the score of the structure $G_{X \to Y}$ depends on p. The figure illustrates how the scores change depending on the number of training samples.
As we get more data, the Bayesian score prefers the structure $G_{X \to Y}$ where X and Y are dependent.



Bayesian score for Bayesian networks

But if the dependence becomes weaker (smaller *p* values), more data are needed to establish this preference.

Proposition: Let G be a network structure and let P(Θ_G | G) be a parameter prior satisfying global parameter independence. Then

$$P(D|G) = \prod_i \int_{\Theta_{X_i|Pa_{X_i}}} \prod_m P\left(x_i[m]\middle|pa_{X_i}[m],\Theta_{X_i|Pa_{X_i}}, G\right) P\left(\Theta_{X_i|Pa_{X_i}}|G\right) d\Theta_{X_i|Pa_{X_i}}.$$

The Bayesian score is biased towards simpler structures.

But as it gets more data, it is willing to recognize that a more complex structure is necessary. It appears to trade off the fit to data with model copmlexity.

Practical example of parameter learning: Stochastic Dynamics simulations of a photosynthetic vesicle

I Introduction: prelude photosynthesis

II Process view and geometric model of a chromatophore vesicle

Tihamér Geyer & V. Helms (Biophys. J. 2006a, 2006b)

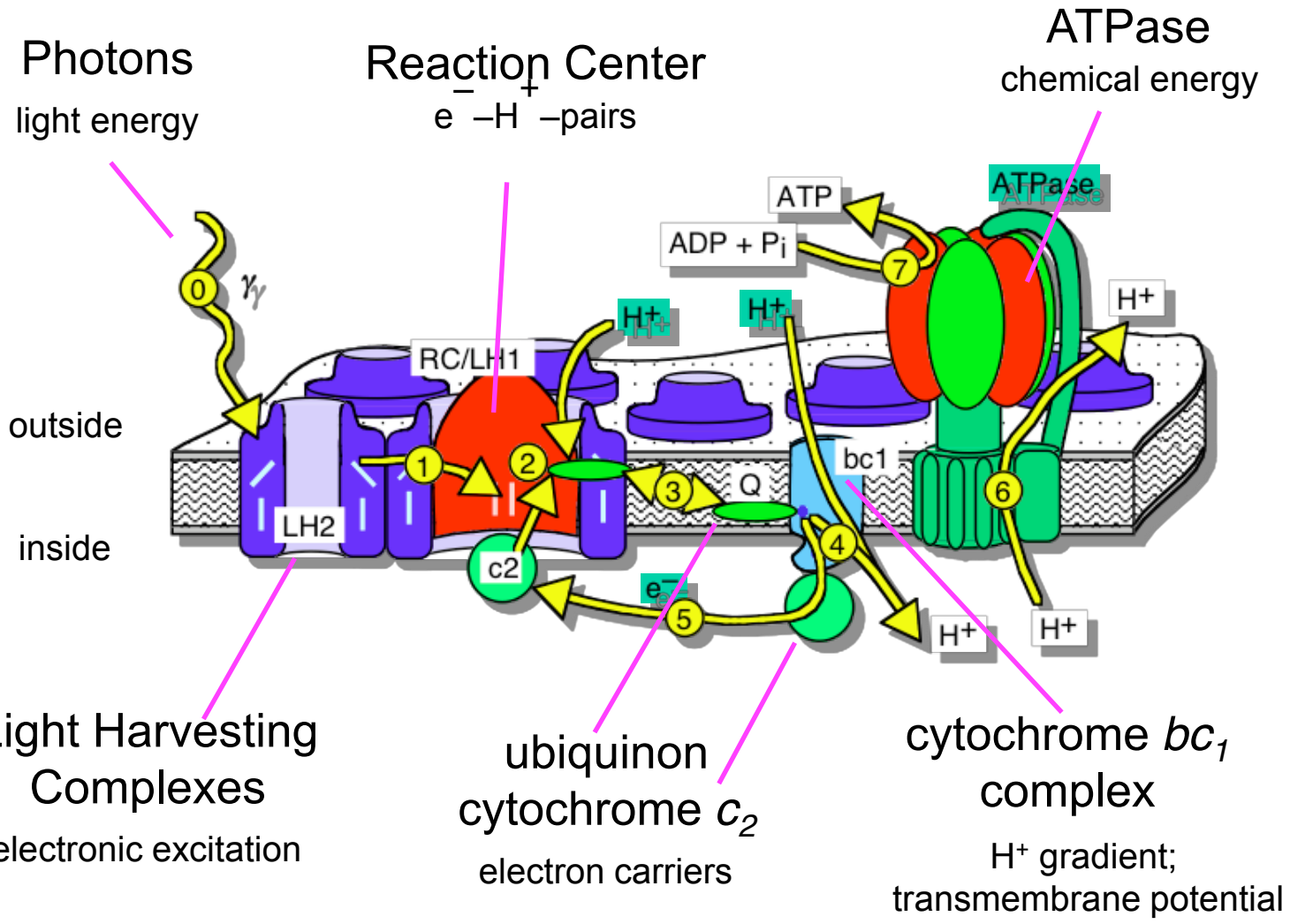
III Stochastic dynamics simulations

T. Geyer, Florian Lauck & V. Helms (J. Biotechnol. 2007)

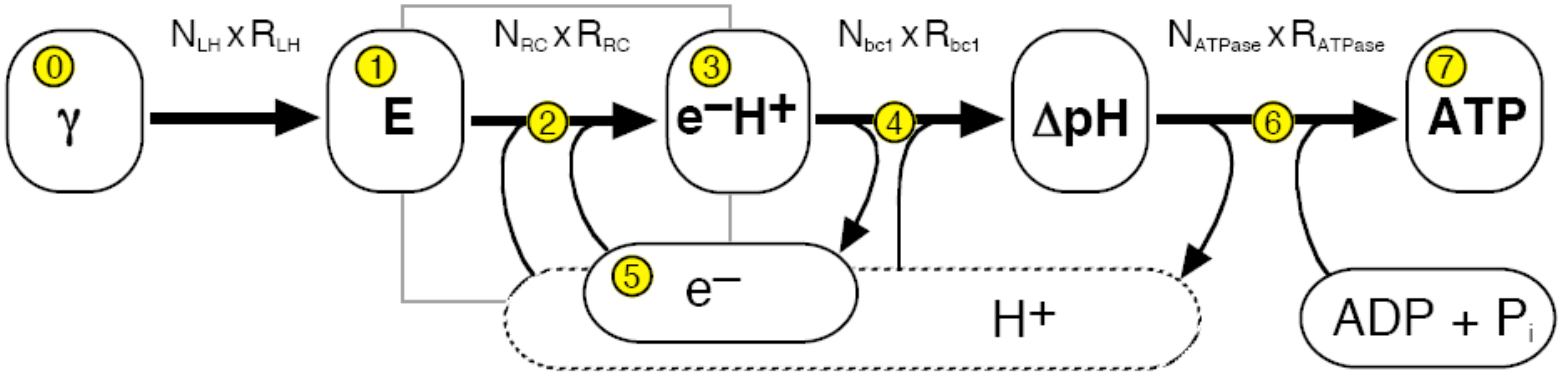
IV Parameter fit through evolutionary algorithm

T. Geyer, X. Mol, S. Blaß & V. Helms (PLoS ONE 2010)

Bacterial Photosynthesis 101



Viewing the photosynthetic apparatus as a conversion chain

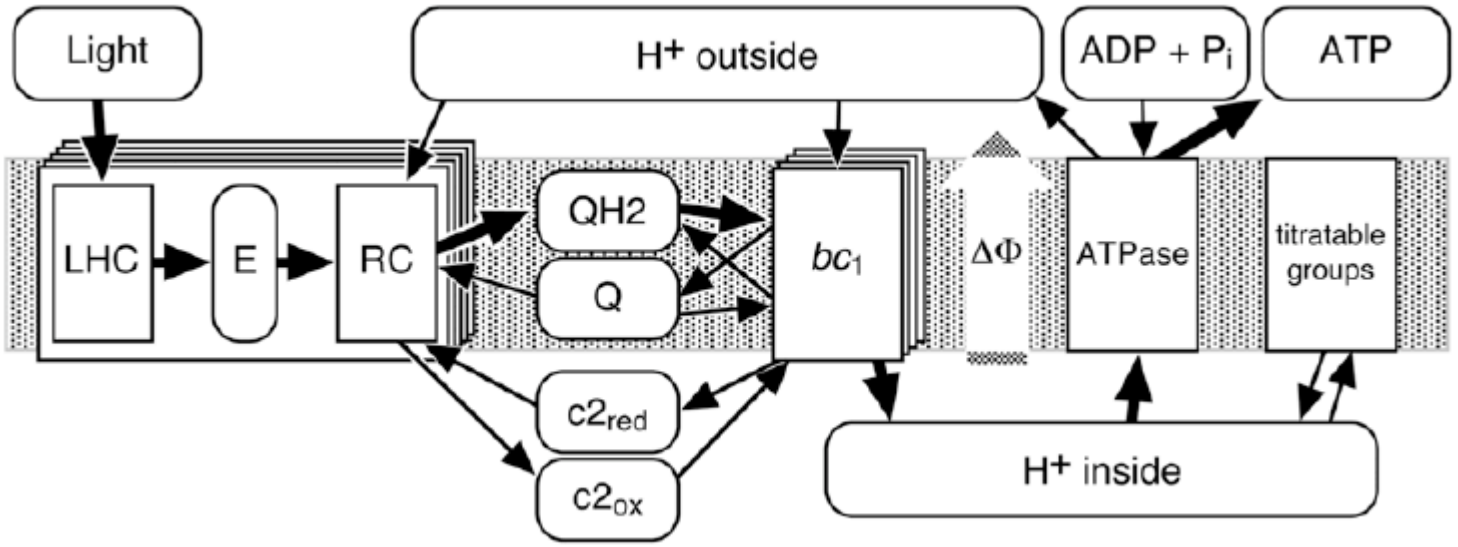


Thick arrows : path through which the photon energy is converted into chemical energy stored in ATP via the intermediate stages (rounded rectangles).

Each conversion step takes place in parallely working proteins.
Their number N times the conversion rate of a single protein R determines the total throughput of this step.

γ : incoming photons collected in the LHCs
E : excitons in the LHCs and in the RC
e^-H^+ electron–proton pairs stored on the quinols
e^- for the electrons on the cytochrome c_2
pH : transmembrane proton gradient
H^+ : protons outside of the vesicle (broken outine of the respective reservoir).

Stochastic dynamics simulations: Molecules & Pools model



Round edges: **pools** for metabolite molecules

Rectangles: protein machines are modeled explicitly as multiple copies

fixed set of parameters

integrate rate equations with stochastic algorithm

Include stochastic effects

(Consequence1) $\rightarrow$ modeling of reactions as continuous fluxes of matter is no longer correct.

(Consequence2) Significant **stochastic fluctuations** occur.

To study the stochastic effects in biochemical reactions, stochastic formulations of chemical kinetics and Monte Carlo computer simulations have been used.

Daniel Gillespie (J Comput Phys 22, 403 (1976); J Chem Phys 81, 2340 (1977)) introduced the exact **Dynamic Monte Carlo (DMC)** method
that connects the traditional chemical kinetics and stochastic approaches.

Basic outline of the direct method of Gillespie

(Step i) generate a list of the components/species and define the initial distribution at time $t = 0$.

(Step ii) generate a list of possible events E_i (chemical reactions as well as physical processes).

(Step iii) using the current component/species distribution, prepare a probability table $P(E_i)$ of all the events that can take place.
Compute the total probability

$$P_{tot} = \sum P(E_i)$$

$P(E_i)$: probability of event E_i .

(Step iv) Pick two random numbers r_1 and $r_2 \in [0...1]$ to decide which event E_μ will occur next and the amount of time τ after which E_μ will occur.

Resat et al., J.Phys.Chem. B 105, 11026 (2001)

Basic outline of the direct method of Gillespie

Using the random number r_1 and the probability table,
the event E_μ is determined by finding the event that satisfies the relation

$$\sum_{i=1}^{\mu-1} P(E_i) < r_1 P_{tot} \leq \sum_{i=1}^{\mu} P(E_i)$$

The second random number r_2 is used to obtain the amount of time τ between the reactions

$$\tau = -\frac{1}{P_{tot}} \ln(r_2)$$

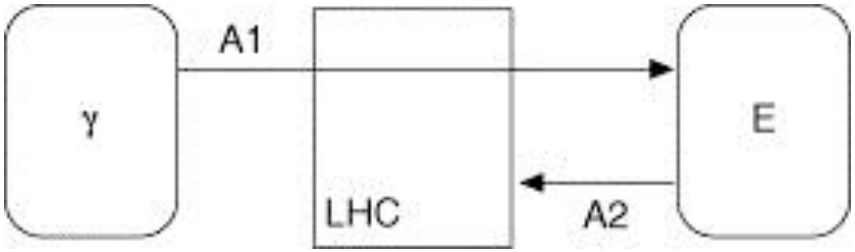
As the total probability of the events changes in time, the time step between occurring steps varies.

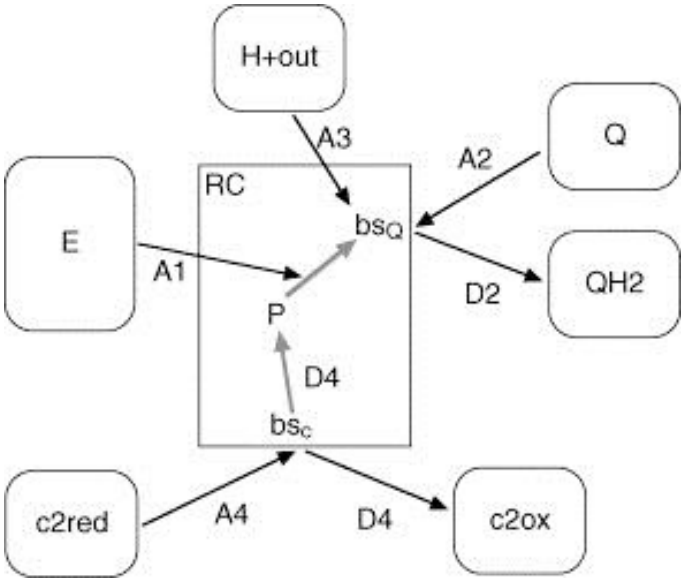
Steps (iii) and (iv) are repeated at each step of the simulation.

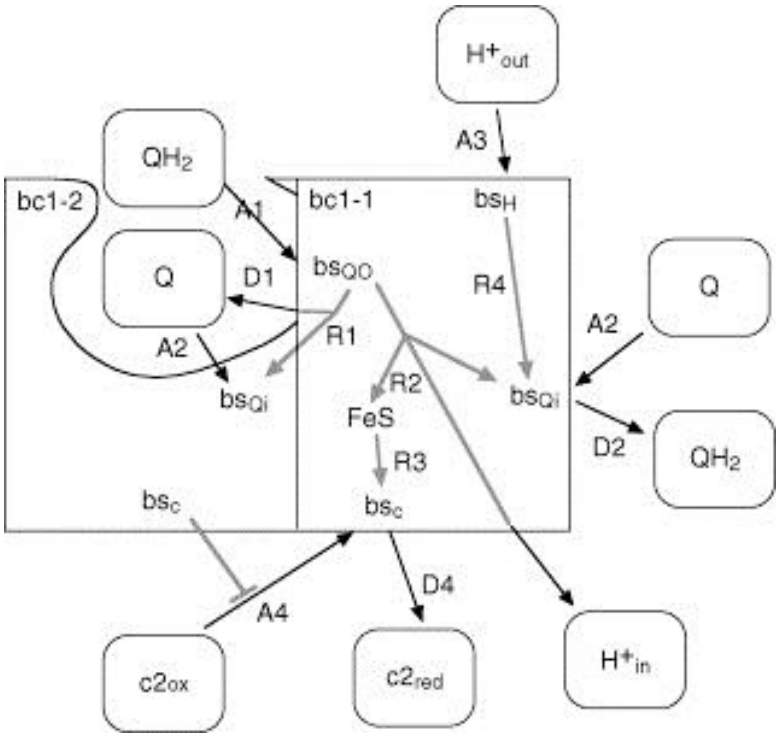
The necessary number of runs depends on the inherent noise of the system and on the desired statistical accuracy.

Resat et al., J.Phys.Chem. B 105, 11026 (2001)

reactions included in stochastic model of chromatophore







Stochastic simulations of a complete vesicle

Model vesicle: 12 LH1/RC-monomers
1-6 bc_1 complexes
1 ATPase

120 quinones
20 cytochrome c_2

integrate rate equations with:

- Gillespie algorithm (associations)

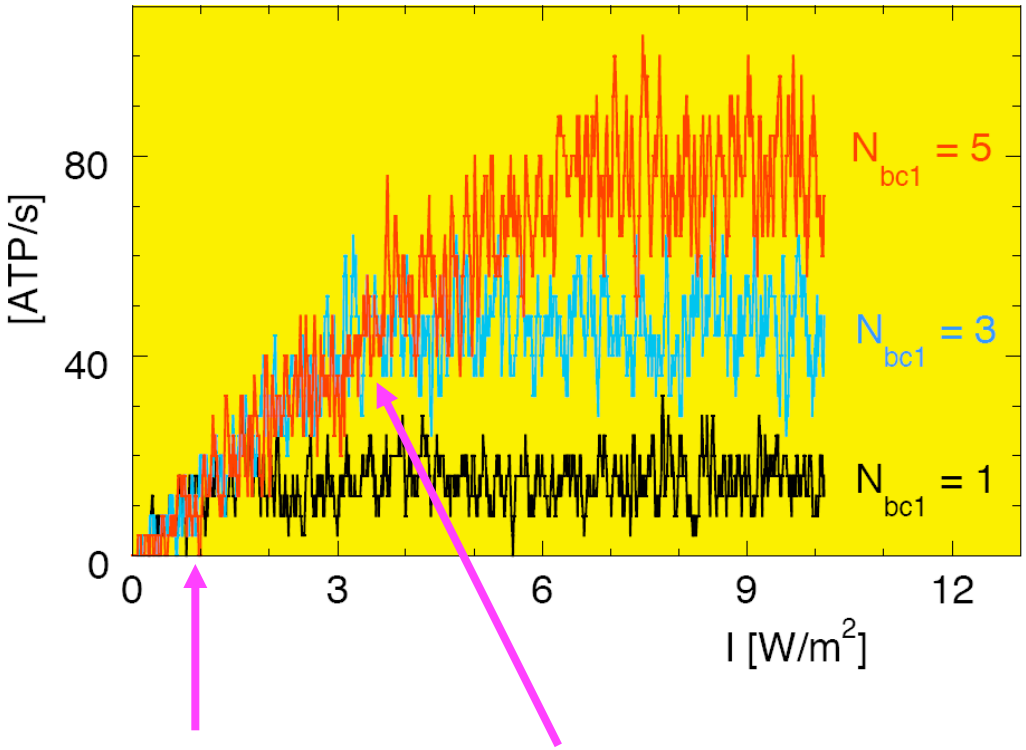
- Timer algorithm (reactions); 1 random number determines when reaction occurs

simulating 1 minute real time requires 1.5 minute on one opteron 2.4 GHz proc

simulate increase of light intensity (sunrise)

during 1 minute,
light intensity is slowly
increased from 0 to 10 W/m^2
(quasi steady state)

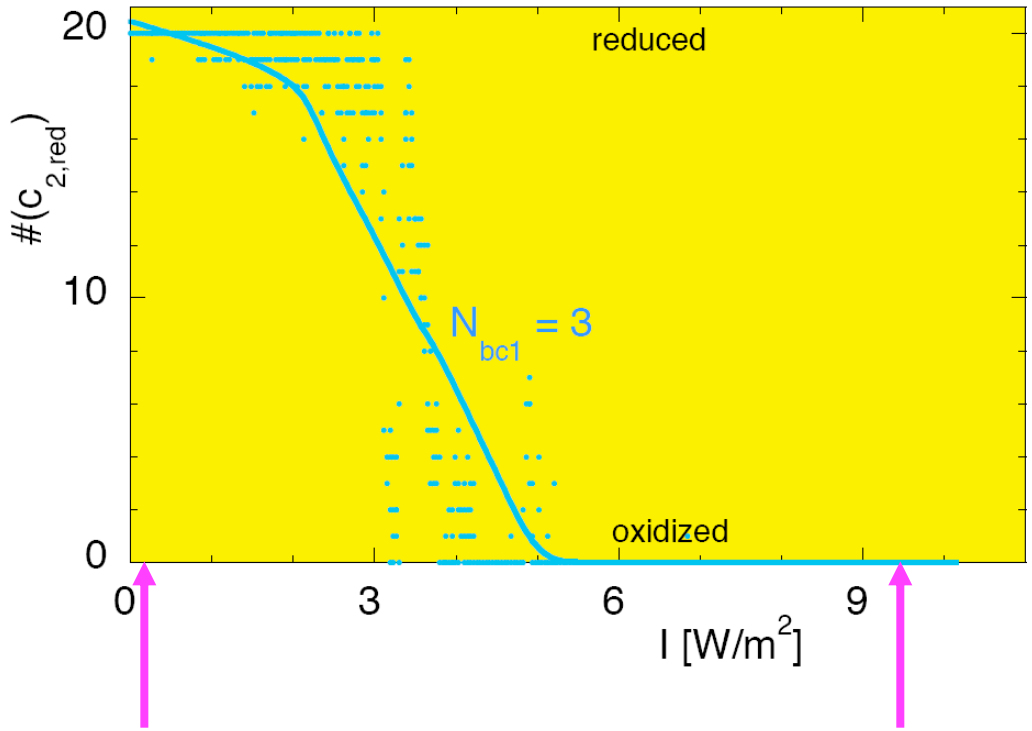
→ there are two regimes
- one limited by available light
- one limited by bc_1 throughput



low light intensity:
linear increase of
ATP production
with light intensity

high light intensity:
saturation is reached
the later the higher the
number of bc1 complexes

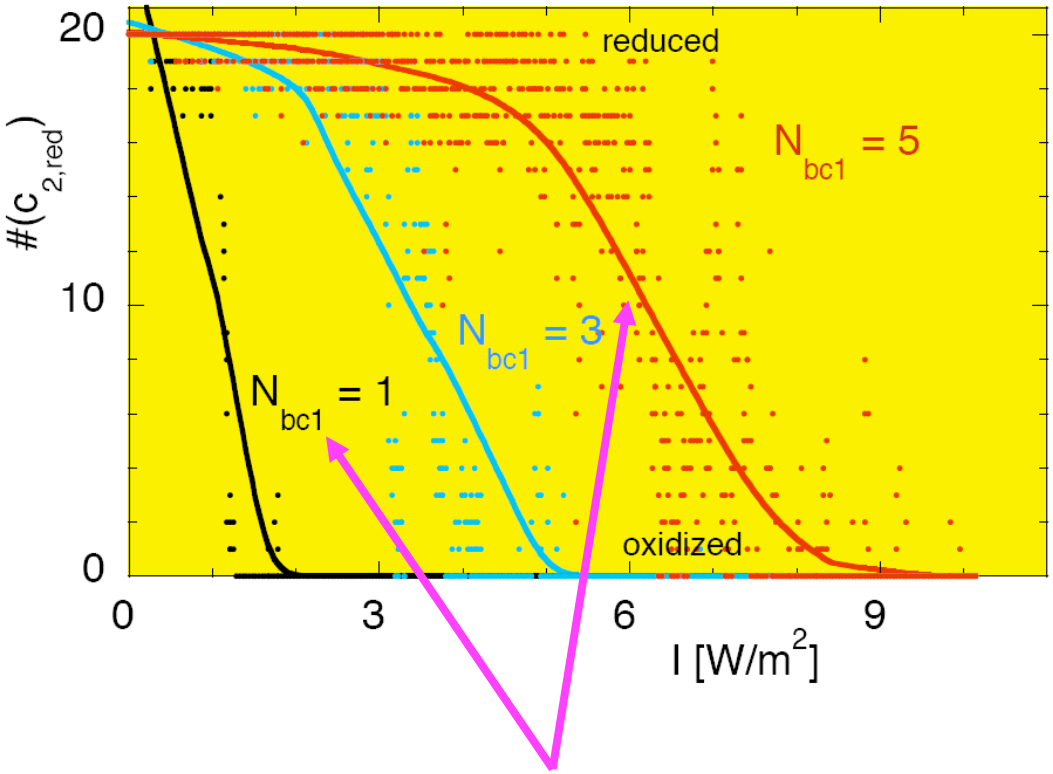
oxidation state of cytochrome c_2 pool



low light intensity:
all 20 cytochrome c_2 are reduced by bc_1

high light intensity
RCs are faster than bc_1, c_2s wait for electrons

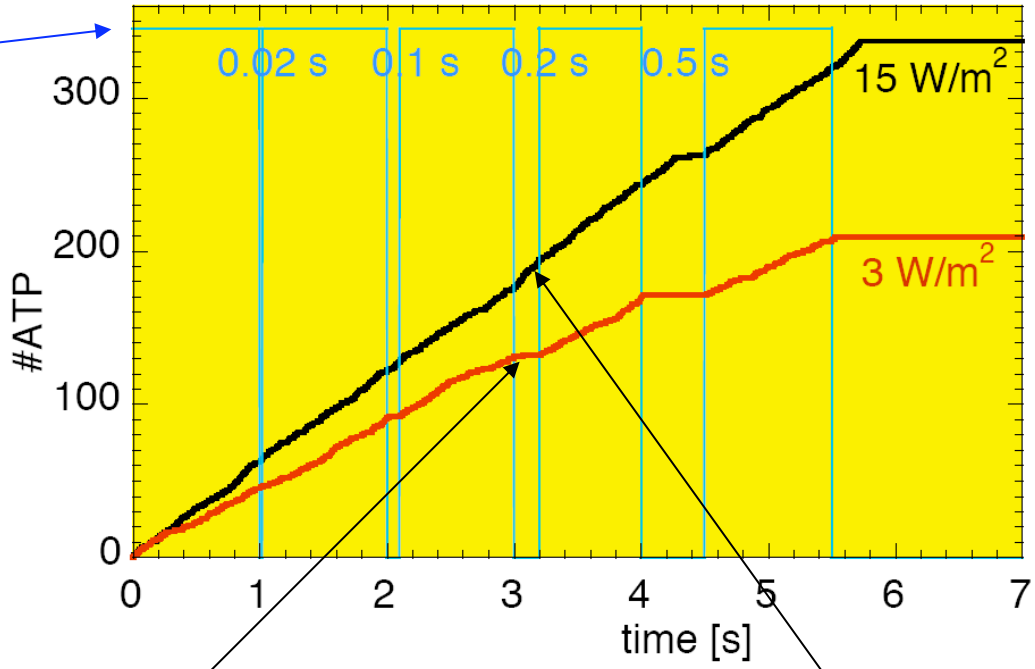
oxidation state of cytochrome c_2 pool



more bc_1 complexes
can load more
cytochrome c_2s

total number of produced ATP

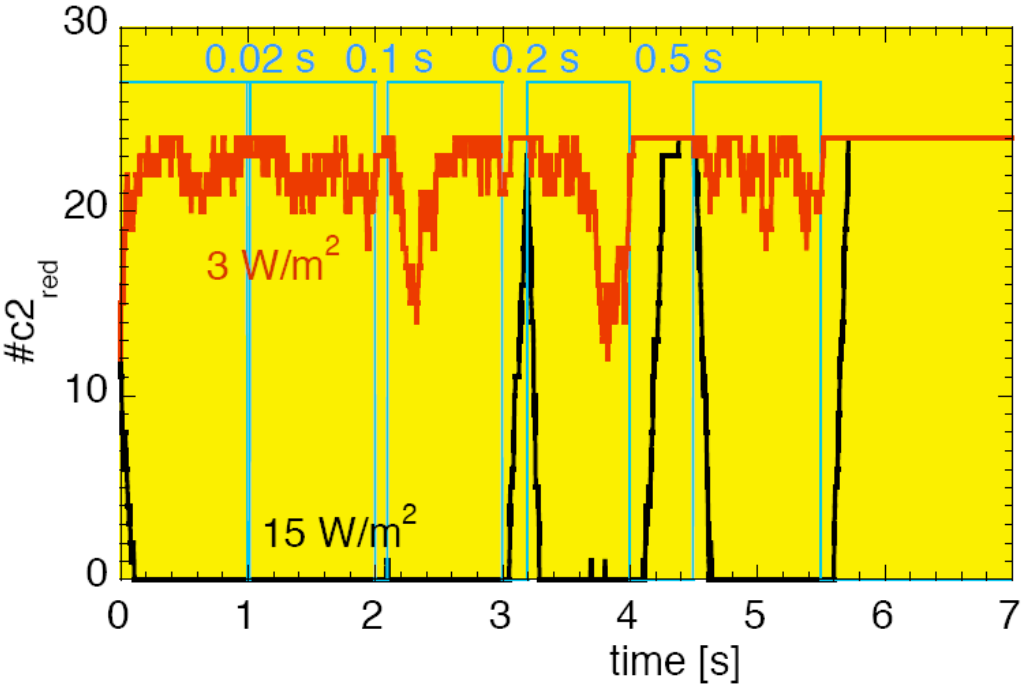
blue line:
illumination



low light intensity: any interruption stops ATP production

high light intensity: interruptions are buffered up to 0.3 s duration

c_2 pool acts as buffer



At high light intensity, c2 pool is mainly oxidized.

If light is turned off, bc1 can continue to work (load c2s, pump protons, let ATPase produce ATP) until c2 pool is fully reduced.

What if parameters are/were unknown ?

Bridging the Gap: Linking Molecular Simulations and Systemic Descriptions of Cellular Compartments

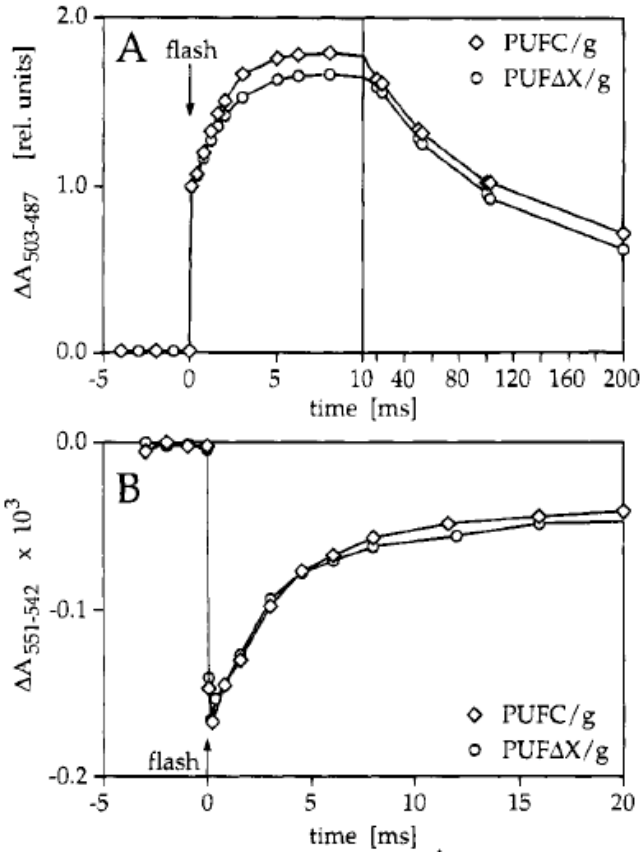
Tihamér Geyer*, Xavier Mol, Sarah Blaß, Volkhard Helms

Center for Bioinformatics, Saarland University, Saarbrücken, Germany

PLoS ONE (2010)

choose 25 out of 45 system parameters for optimization.

take 7 different non-equilibrium time-resolved experiments from Dieter Oesterhelt lab (MPI Martinsried).



Biochemistry **1995,** *34,* 15235−15247

15235

Role of PufX Protein in Photosynthetic Growth of *Rhodobacter sphaeroides*. 1. PufX Is Required for Efficient Light-Driven Electron Transfer and Photophosphorylation under Anaerobic Conditions†

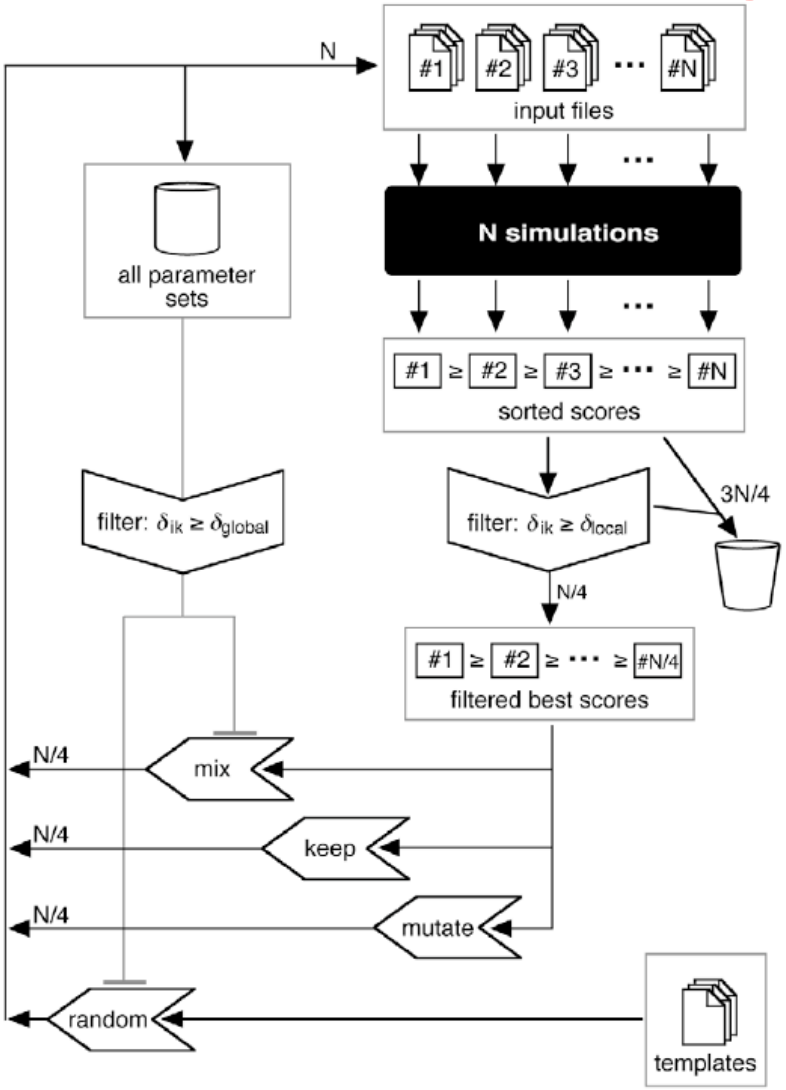
Wolfgang P. Barz,‡,§ Francesco Francia,|| Giovanni Venturoli,|| B. Andrea Melandri,|| André Verméglio,⊥ and Dieter Oesterhelt*,‡

Parameters not optimized

Parameter	Value	Description
$bc_1::k_{on}(H^+_{out})$	10^{10} nm^3 s^{-1}	rate for proton uptake from the cytoplasm by bc_1
$bc_1::k_{tr}(e:Q_o\Rightarrow FeS)$	$2.3 * 10^3$ s^{-1}	rate for electron transfer from Q_o to FeS
$bc_1::k_{tr}(e:c_1\Rightarrow c_2)$	10^5 s^{-1}	electron transfer rate from c_1 to bound cytochrome c_2
$bc_1::k_{tr}(e:Q_o\Rightarrow b_L)$	10^4 s^{-1}	electron transfer from Q_0 to b_L heme
$bc_1::k_{tr}(e:b_L\Rightarrow b_H)$	10^4 s^{-1}	electron transfer from b_L to b_H heme
$\Delta\Phi::V$	$2.65 * 10^4$ nm^3	inner volume of the vesicle
$\Delta\Phi::A$	$5.28 * 10^3$ nm^2	membrane area (Q pool „volume“)
$\Delta\Phi::C_{Hin}$	1.0 e	effective charge of a free proton in the vesicle
$\Delta\Phi::C_{Hin}$	1.0 e	effective charge of a proton on the titratable groups
$\Delta\Phi::C_{prot}$	–1.0 e	effective charge of an e^- translocated through an RC
$\Delta\Phi::C_{cred}$	–0.5 e	effective charge of a reduced cytochrome c_2
$\Delta\Phi::C_{cox}$	0.5 e	effective charge of an oxidized cytochrome c_2
PR::N_p	80	number of titratable groups in the vesicle
PR::pK	5.0	pK of the titratable groups
N_{core}	10	number of dimeric core complexes (2 RC + 1 LHC)
N_{bc1}	10	number of cytochrome bc_1 complexes
N_{ATPase}	1	number of ATPases
N_{c2}	20	total number of cytochrome c_2
N_Q	200	total number of quinones

Table S1: Model Parameters Not Included in the Optimization Process

Parameter optimization through evolutionary algorithm



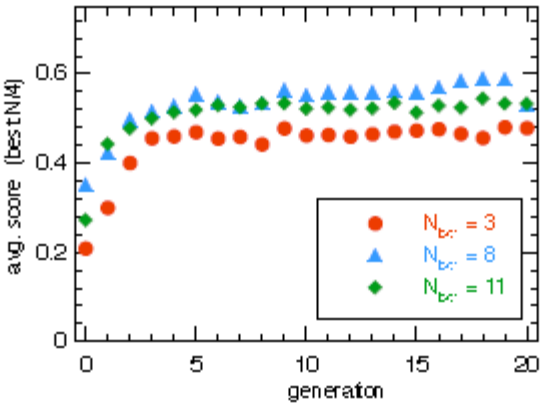


Figure S1: Determining the Number of bc1 Complexes: Evolution of the Master Score

25 optimization parameters

Analyze 1000 best parameter sets among 32.800 simulations:

$$\langle P \rangle = \exp[\langle \log P \rangle]$$

$$\sigma^2 = \langle (\log P - \langle \log P \rangle)^2 \rangle$$

$$P_{\min} = \exp[\langle \log P \rangle - \sigma]$$

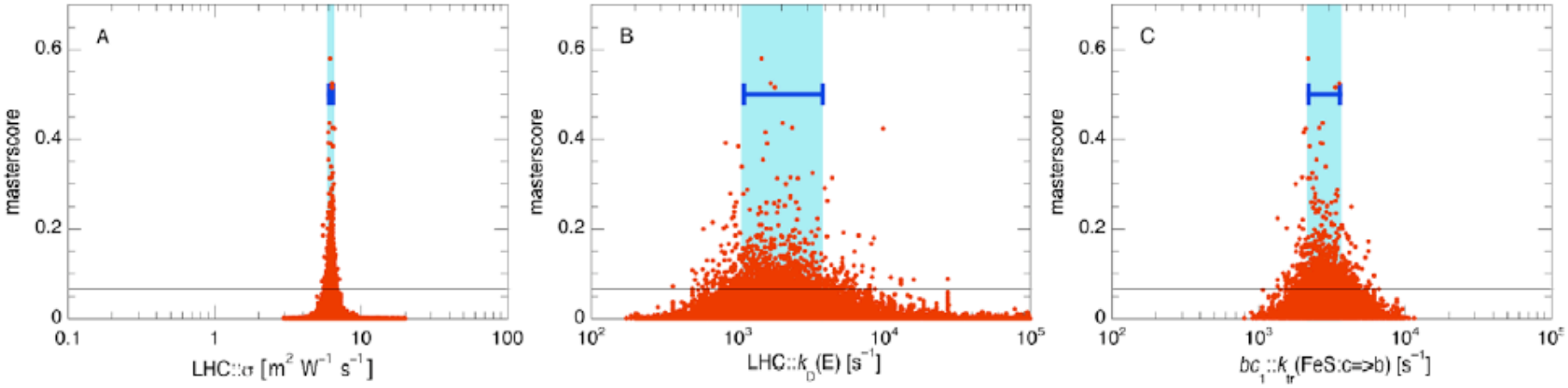
$$P_{\max} = \exp[\langle \log P \rangle + \sigma]$$

parameter	units	<P>	P_{min}...P_{max}	P_{min}/P_{max}
LHC::σ	m^2 W $^{-1}$ s^{-1}	6.22	6.02...6.42	0.94
LHC::N_0	1	1.31	0.81 ... 2.13	0.38
LHC::k_D(E)	s^{-1}	$1.9 * 10^3$	$(1.1...3.8) * 10^3$	0.29
RC::k_{on}(E)	s^{-1}	$2.4 * 10^6$	$(1.2...4.5) * 10^6$	0.27
RC::k_{on}(H^+)	nm^3 s^{-1}	$1.4 * 10^8$	$(1.3...1.6) * 10^8$	0.81
RC::k_{on}(Q)	nm^2 s^{-1}	$6.0 * 10^4$	$(4.4...8.1) * 10^4$	0.54
RC::k_{off}(QH2)	s^{-1}	87	70...108	0.65
RC::k_{on}(c2red)	nm^3 s^{-1}	$9.2 * 10^5$	$(7.3...11.5) * 10^5$	0.63
RC::k_{off}(c2ox)	s^{-1}	$2.2 * 10^3$	$(1.6...3.0) * 10^3$	0.53
bc1::kon(QH2@Qo)	nm^2 s^{-1}	$1.2 * 10^4$	$(0.79...1.7) * 10^4$	0.46
bc1::koff(Q@Qo)	s^{-1}	28.3	26.3...30.4	0.86
bc_1::k_{τ}(Q:Q_o -> Q_i)	s^{-1}	$4.9 * 10^3$	$(3.6...6.7) * 10^3$	0.54
bc1::kon(Q@Qi)	mm^2 s^{-1}	$6.7 * 10^5$	$(4.5...10) * 10^5$	0.45
bc1::koff(QH2@Qi)	s^{-1}	86	68...110	0.62
bc_1::k_{τ}(QH2:Q_i -> Q_o)	s^{-1}	$3.8 * 10^3$	$(2.6...5.5) * 10^3$	0.47
bc_1::k_{on}(c2ox)	nm^3 s^{-1}	$9.4 * 10^6$	$(6.3...14) * 10^6$	0.47
bc_1::k_{off}(c2red)	s^{-1}	$6.0 * 10^3$	$(3.3...11) * 10^3$	0.30
bc1::koff(H+@Qo)	s^{-1}	$2.4 * 10^4$	$(1.3...4.3) * 10^4$	0.30
bc_1::k_{τ}(FeS:b -> c)	s^{-1}	$3.9 * 10^3$	$(3.1...5.1) * 10^3$	0.61
bc_1::k_{τ}(FeS:c -> b)	s^{-1}	$2.8 * 10^3$	$(2.2...3.6) * 10^3$	0.61
bc_1::k_{τ}(e:b_H -> Q_i)	s^{-1}	$7.7 * 10^3$	$(5.0...12) * 10^3$	0.42
bc_1::Φ_0	mV	102	83...114	0.73
$\Delta\Phi$::U_0	mV/e	10.3	9.5...11	0.85
$\Delta\Phi$::$\Delta\Phi_0$	mV/pH	10	7.6...13.7	0.55
PR::pK	1	4.84	3.9...5.9	0.66

Sensitivity of master score

Decay rate of excitons

in LHC



Absorption cross section

light harvesting complex

Kinetic rate for hinge motion of FeS domain in bc1 complex

Some parameters are very sensitive, others not.

Three best-scored parameter sets

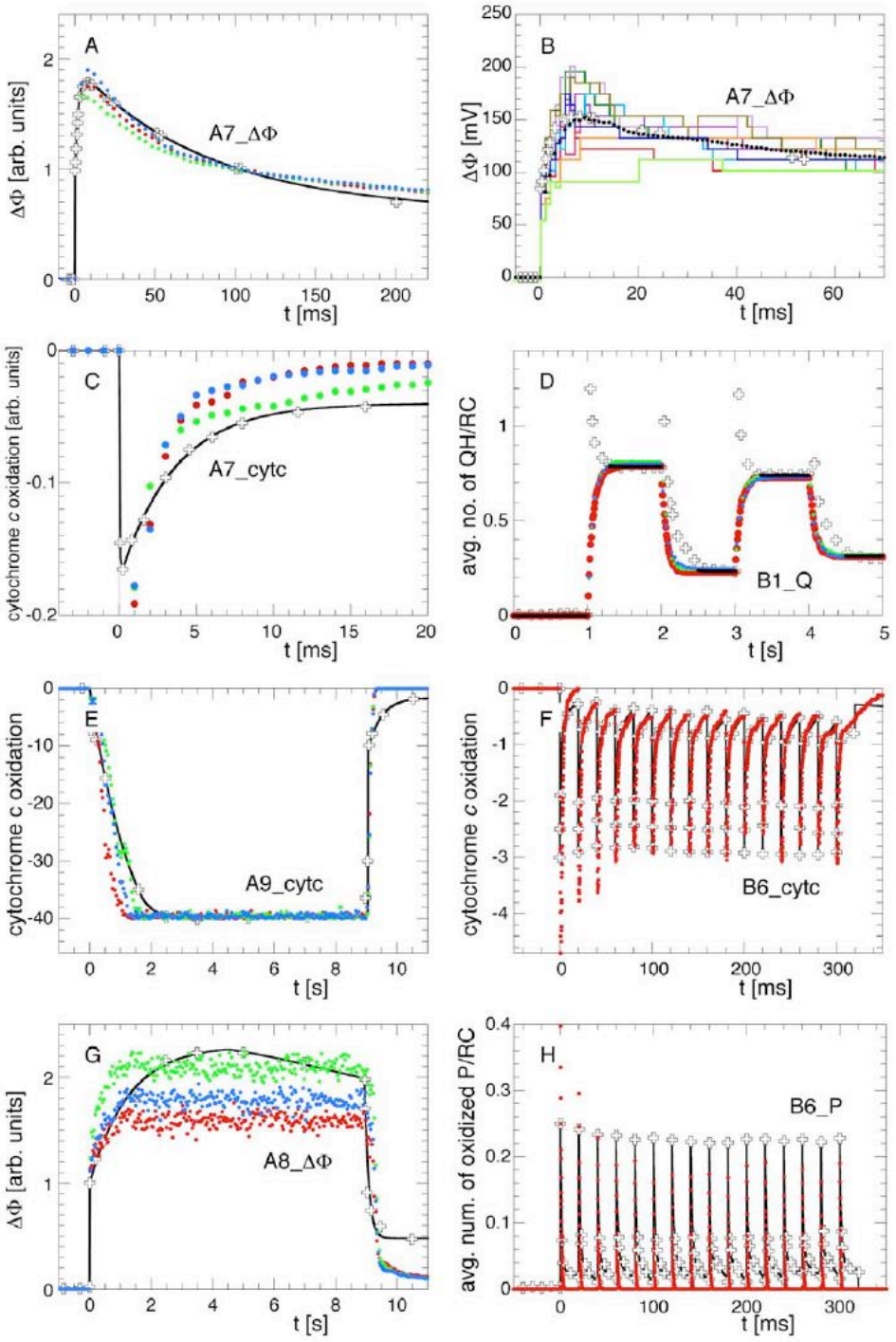
Score of individual parameter set *i* for matching one experiment:

$$s_i = \frac{C_i}{\sum (x(t_i) - f(t_i))^2}$$

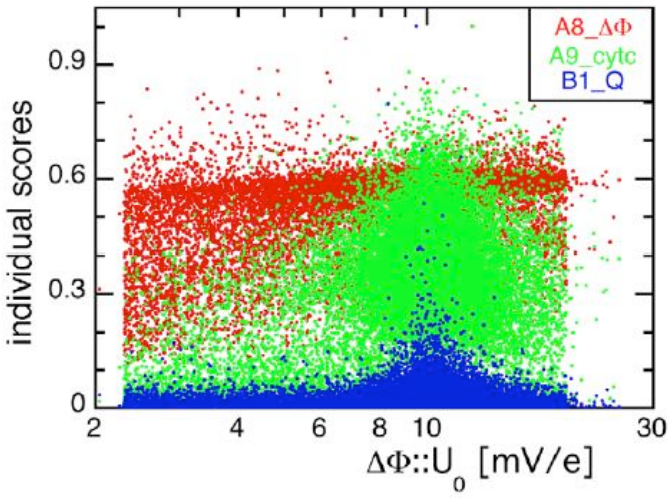
$x(t_i)$: simulation result

$f(t_i)$: smooth fit of exp. data

Master score for one parameter set: defined as product of the individual scores s_i



Different experiments yield different sensitivity



"importance score":
Sum of the sensitivities P_{min}/P_{max} of all relevant parameters

Table 2. Importance scores and correlation coefficients between the master score and the respective individual scores of the experimental scenarios denoting the relative importance of each of the experiments for the parameter value optimization.

experiment	A7 cytc	A7 ΔΦ	A8 ΔΦ	A9 cytc	B1 Q	B6 P	B6 cytc	BC1
importance score	4.4	7.7	5.8	9.7	3.8	5.2	8.9	5.5
correlation	0.09	0.44	0.22	0.38	0.83	0.17	0.31	0.41

The importance scores are determined as the sums of the sensitivities of all relevant parameters against the individual scores (see table S2 for all the individual values). The correlation coefficients are obtained from a linear fit of the master score against the respective individual score.

Analysis could suggest new experiments that would be most informative!

Summary

Only 1/3 of the kinetic parameters previously known.

Stochastic parameter optimization converges robustly into the same parameter basin as known from experiment.

Two large-scale runs (15 + 17 parameters) yielded practically the same results.

If implemented as grid search, less than 2 points per dimension.

It appears enough to know 1/3 – 1/2 of kinetic rates about a system to be able to describe it quantitatively (IF connectivities are known).