V23 Integrated Metabolic and Transcriptional Networks

Two methods:

- Probabilistic Regulation of Metabolism (PROM)

and

- Integrated Deduced REgulation And Metabolism (IDREAM),

by group of

Nathan Price @ Institute of Systems Biology / Seattle

PROM (2010)

The construction of an integrated metabolic-regulatory network using PROM requires the following:

(i) a reconstructed genome-scale **metabolic network**;

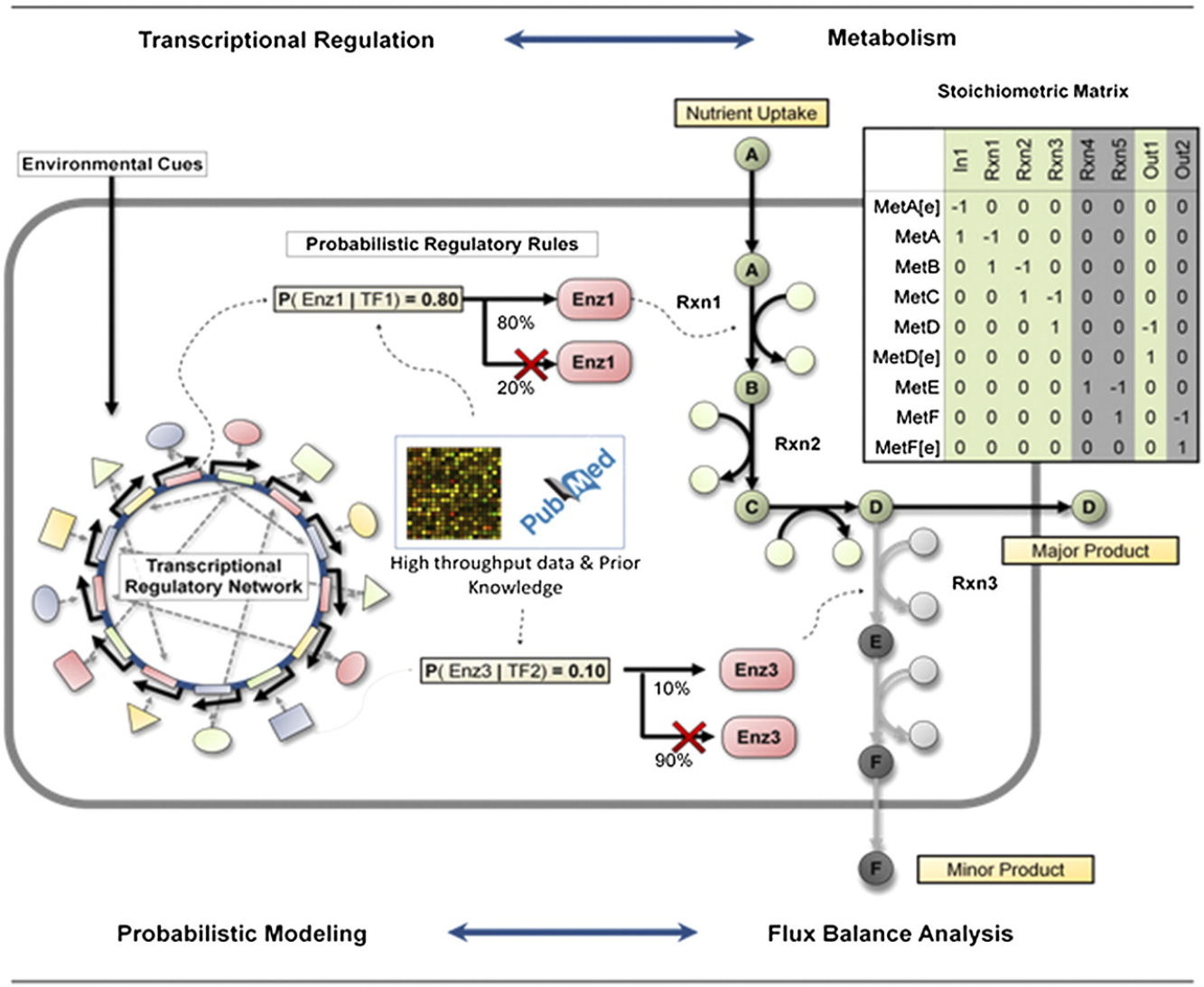
(ii) a **regulatory network** structure, consisting of transcription factors (TFs) and their targets;

(iii) abundant **gene expression** data, in which the transcriptome has been measured under various environmental and genetic perturbations; and

(iv) additional interactions involving enzyme regulation by metabolites and proteins.

Chandrasekaran & Price,
PNAS 107:17845 (2010)

PROM - overview



	In1	Rxn1	Rxn2	Rxn3	Rxn4	Rxn5	Out1	Out2
MetA[e]	-1	0	0	0	0	0	0	0
MetA	1	-1	0	0	0	0	0	0
MetB	0	1	-1	0	0	0	0	0
MetC	0	0	1	-1	0	0	0	0
MetD	0	0	0	1	0	0	-1	0
MetD[e]	0	0	0	0	0	0	1	0
MetE	0	0	0	0	1	-1	0	0
MetF	0	0	0	0	0	1	0	-1
MetF[e]	0	0	0	0	0	0	0	1

Chandrasekaran & Price,
PNAS 107:17845 (2010)

Metabolic network is represented using a **stoichiometric matrix. Regulatory interactions** are represented as **probabilities**.
The TF states are determined based on environmental conditions; the state of TF is then used to determine the on/off state of the target genes based on probabilities estimated from microarray data. The probabilities are then used to constrain the fluxes through the metabolic network.

PROM – regulatory interactions

PROM uses **probabilities** to represent gene states and gene–TF interactions.

The probability of gene A being **on** when the regulating TF B is **off** is given by $P(A = 1|B = 0)$.

Similarly, $P(A = 1|B = 1)$ gives the probability of A being **on** when B is **on**.

The transcriptomic data were **binarized** with respect to a fixed low value threshold for all genes.

Gene expression values less than a threshold were considered to be **off** and the remaining values were set to **on**.

Chandrasekaran & Price,
PNAS 107:17845 (2010)

PROM – use of transcriptomic data

The relationship between a TF and a target gene is then quantified by using transcriptomics data.

By using this interaction data, one models the effect of perturbations to the regulatory network using PROM.

To predict the effect of a TF KO on a gene A, which is the probability P(A = 1|B = 0), we count or estimate the number of microarray samples wherein the target gene A is on when the TF B is off.

$$\frac{N(Target = 1TF = 0)}{N(TF = 0)}$$

If the data set is large enough, we can get a robust estimate of the probability for this interaction.

So, if the probability associated with a gene being on is 0.8,
then we estimate that in 80% of the samples we find the gene to be on,
and 20% of the samples it is off or not expressed.

Chandrasekaran & Price,
PNAS 107:17845 (2010)

PROM – effect of regulatory links

To model the effect of the KO at the genome scale, the states of all its target genes are determined.
These probabilities are then used to constrain the fluxes through the reactions controlled by the target genes.

For the example just discussed, the flux through the reaction regulated by gene A cannot exceed the maximum flux possible, V_{max}, through the reaction if it is on, and would be zero when it is off.

Hence, on average, the maximum flux through the reaction in the population would be $0.8 \times V_{max}$ or, in general, the upper bound for the flux is $p \times V_{max}$, where p is the probability of the gene being on.

The systemic reaction V_{max} is estimated by flux variability analysis (FVA) on the unregulated metabolic model.

Chandrasekaran & Price,
PNAS 107:17845 (2010)

PROM – optimization function

When the constraints have been set, the optimal growth of the regulated network is determined by solving a linear optimization problem as in FBA.

PROM finds a flux distribution that satisfies the same constraints as FBA **plus additional constraints** resulting from the transcriptional regulation:

$$\min(\kappa.\alpha + \kappa.\beta),$$

subject to constraints $lb' - \alpha \leq v \leq ub' + \beta$ and $\alpha,\beta \geq 0$,

where lb' and ub' are constraints based on transcriptional regulation,
(lb und ub stand for lower bound and upper bound),
α and β are positive constants that represent deviation from those constraints, and
κ represents the **penalty** for such deviations.

Chandrasekaran & Price,
PNAS 107:17845 (2010)

PROM

The higher the value of κ, the greater is the constraint on the system based on transcriptional regulation.

For values of κ significantly greater than 1, the regulatory constraints become “hard”.
For values less than 0.1 they become less pronounced.

Chandrasekaran & Price used a κ value of 1 for all their simulations as it represents a tradeoff between the two extremes.

Chandrasekaran & Price,
PNAS 107:17845 (2010)

EGRIN: construct transcriptional regulatory network

Approach:

- perturb the cells (genetically or environmentally) – *Halobacterium salinarum,*
- characterize their growth and/or survival phenotype,
- quantitatively measure steady-state and dynamic changes in mRNAs,
- assimilate these changes into a network model that recapitulates all observations, and,
- finally, experimentally validate hypotheses formulated from the model.

Realization:

This approach required the integrated development and implementation of computational and experimental technologies and consisted of the following steps:

Integrated approach

1 Sequence the genome and **assign functions** to genes using protein sequence and structural similarities.

2 Perturb cells by changing relative concentrations of **environmental factors** (EFs - light, oxygen, UV radiation, gamma radiation, Mn, Fe, Co, Ni, Cu, and Zn) and/or **gene knockouts**.

3 Measure the resulting **dynamic** and/or **steady-state transcriptional changes** in all genes using microarrays.

4 Integrate diverse data (mRNA levels, evolutionarily conserved associations among proteins, metabolic pathways, *cis*-regulatory motifs, etc.) with the `cMonkey` algorithm to reduce data complexity and identify **subsets** of genes that are **coregulated** in certain environments (biclusters).

5 Using the machine learning algorithm `Inferelator` construct a dynamic network model for the influence of changes in EFs and TFs on the expression of coregulated genes.

6 Explore the network to **formulate and then experimentally test hypotheses** to drive additional iterations of steps 2–6.

Examples of biclusters

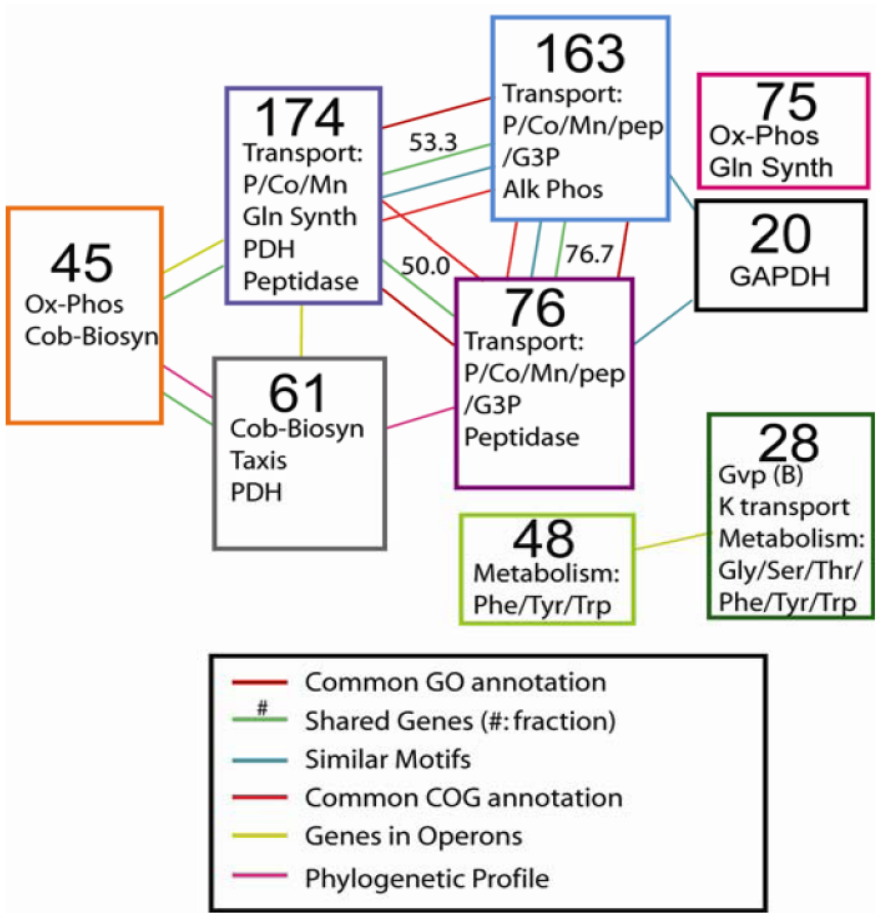


Figure S5. Nine biclusters representing components of C-, N- and energy metabolism. The biclusters (boxes) are all numbered and color coded to match the scheme described in Fig 3 in the main text of this report. The relationships among biclusters on basis of their gene memberships are indicated with different colored lines that connect the biclusters. The overrepresented GO annotations for each bicluster are also indicated (for statistical significance and other details see ST7).

Cell 131, 1354 (2007)

Inferelator algorithm for biclustering

5. Use the machine learning algorithm Inferelator to discover the dynamic influences of EFs and TFs on the expression of co-regulated genes within biclusters.

Briefly, the *Inferelator* (a) selects parsimonious models (i.e. **minimum number of regulatory influences** for each bicluster) that are predictive;

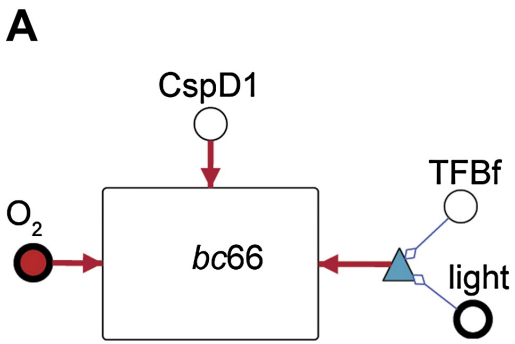
(b) explicitly models temporal behavior (ODEs) to discover causal influences; and

(c) models combinatorial logic i.e. interactions between EFs and TFs and between pairs of TFs.

The resulting model is a set of differential equations that can take as input measured changes in a few TFs and/or EFs to predict kinetic and steady-state transcriptional changes in 80% of genes of *Halobacterium salinarum* with an average (Pearson) correlation of 0.8 to their actual measured changes.

Cell 131, 1354 (2007)

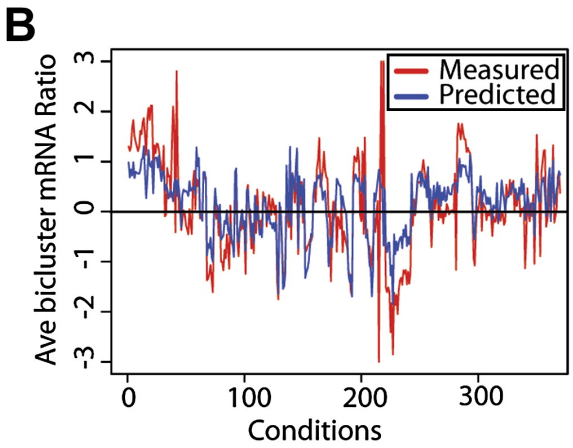
EGRIN predicts novel regulatory influences for known biological processes



Bicluster bc66 contains 34 genes including cytochrome oxidase, ribosomal proteins, and RNA polymerase.

Their transcriptional behavior is nearly perfectly modelled by corresponding changes of 2 EFs (oxygen and light) and 2 TFs (Cspd1 and TFBf).

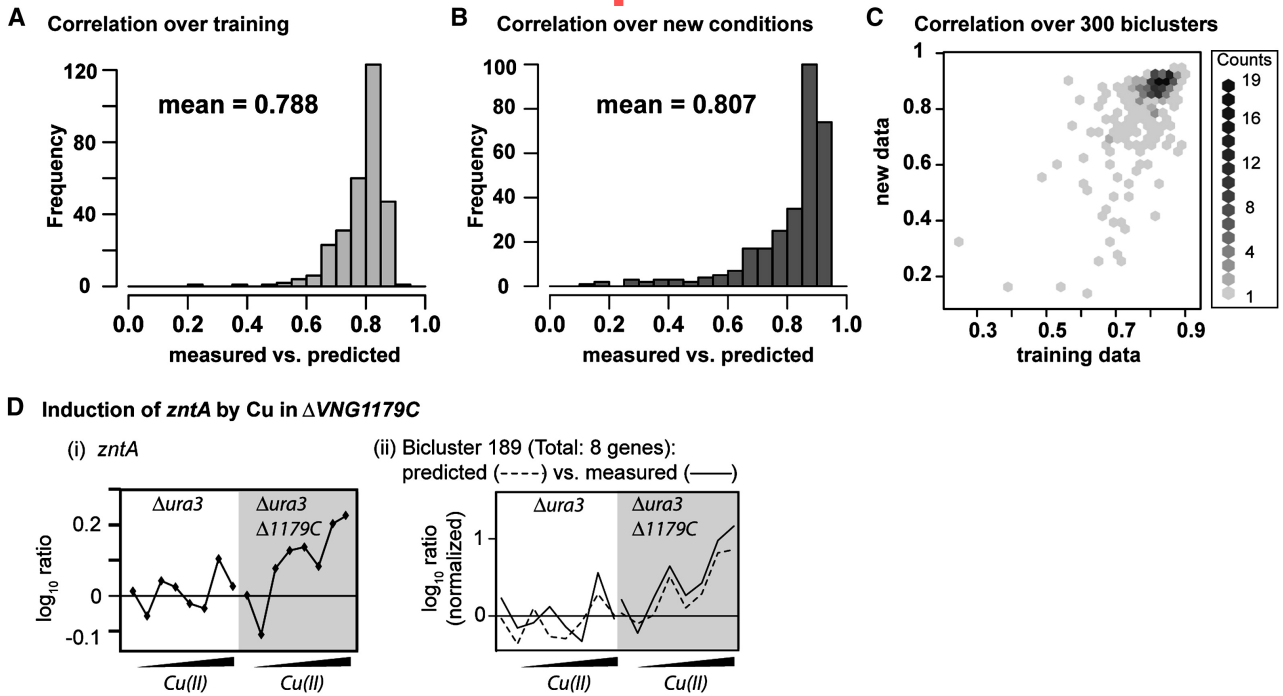
The influences from TFBf and light act through an AND logic gate (triangle).



(B) The mRNA profile of *bc*66 recreated by the combined TFs and environmental influences is nearly identical to the actual (averaged) mRNA levels over 398 experiments.

Cell 131, 1354 (2007)

Correlation of predicted and measured mRNA levels

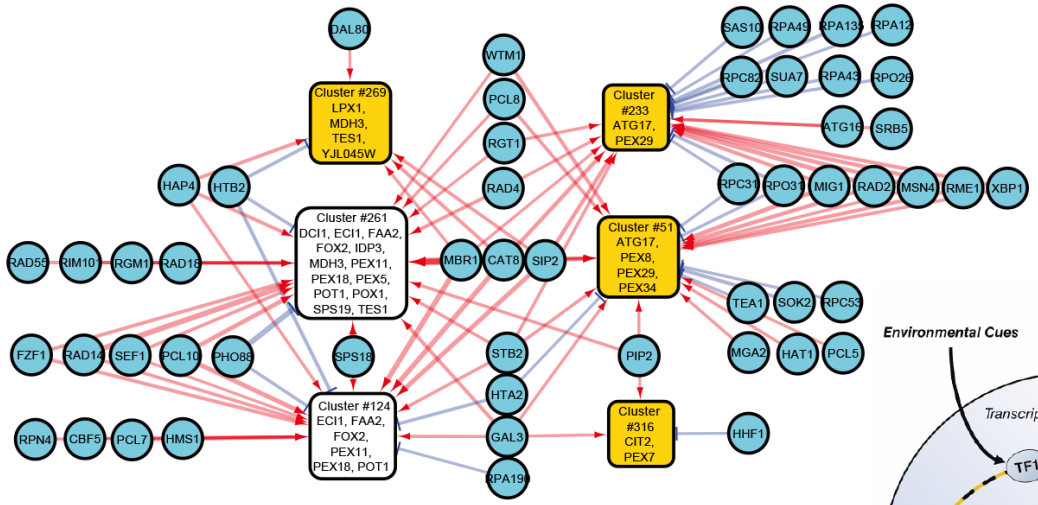


147 new experiments:
(1) New combinatorial perturbations of EFs already in training set
(2) New EF perturbations: oxidative stress agent hydrogen peroxide, chemical mutagen ethyl methyl sulfonate
(3) New combinations of TF and EF perturbations.

Histogram of Pearson correlations of predicted and measured mRNA levels of individual biclusters over the 266 experiments in the training set (A) and the 131 newly collected experiments (B). (C) shows a comparison of correlations between predicted and measured mRNA levels for all 300 biclusters in training set and new data. (D) Transcription of the broad specificity metal ion efflux pump ZntA is upregulated under Cu stress in the *ΔVNG1179C* strain background in which the primary efflux pump is transcriptionally inactivated (*Δura3* is the parent strain in which knockouts are constructed). This altered transcriptional response of ZntA to Cu was accurately modeled by the regulatory influences on *bc*189, which contains this gene along with 7 other genes.

Integrated Deduced REgulation And Metabolism (IDREAM)

IDREAM uses bootstrapping-EGRIN inferred TF regulation of enzyme-encoding genes, then applies a PROM-like approach to apply metabolic network constraints in an effort to improve phenotype prediction.



EGRIN

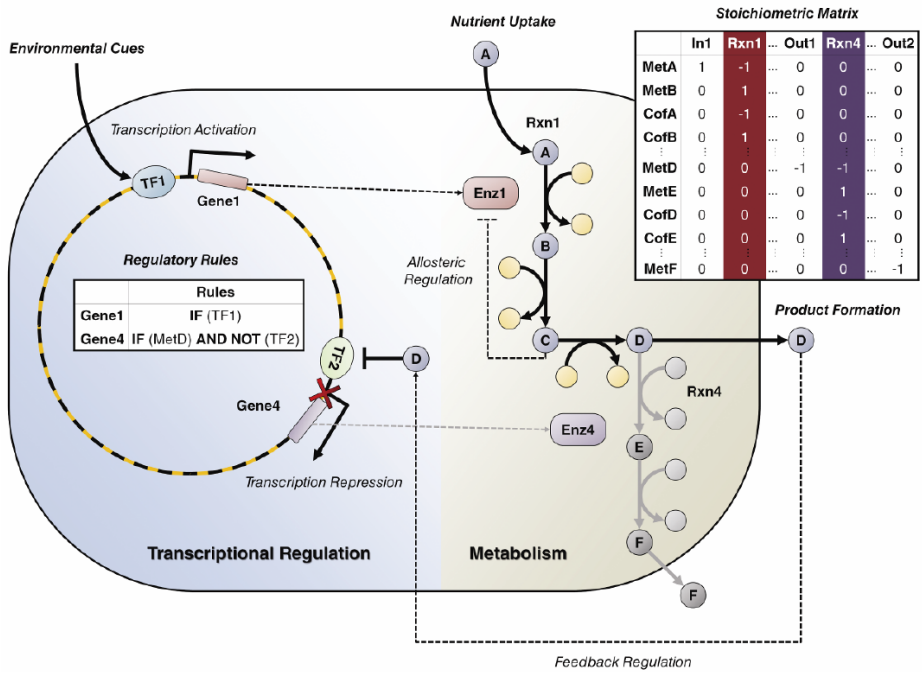
(**E**nvironment & **G**ene **R**egulatory **I**nfluence **N**etwork)

Identify Condition Specific Regulatory Network

PROM

(**P**robabilistic **R**egulation **o**f **M**etabolism)

Predict Condition Specific Growth Rates



Wang … Price (2017) PLoS Comput Biol 13: e1005489.

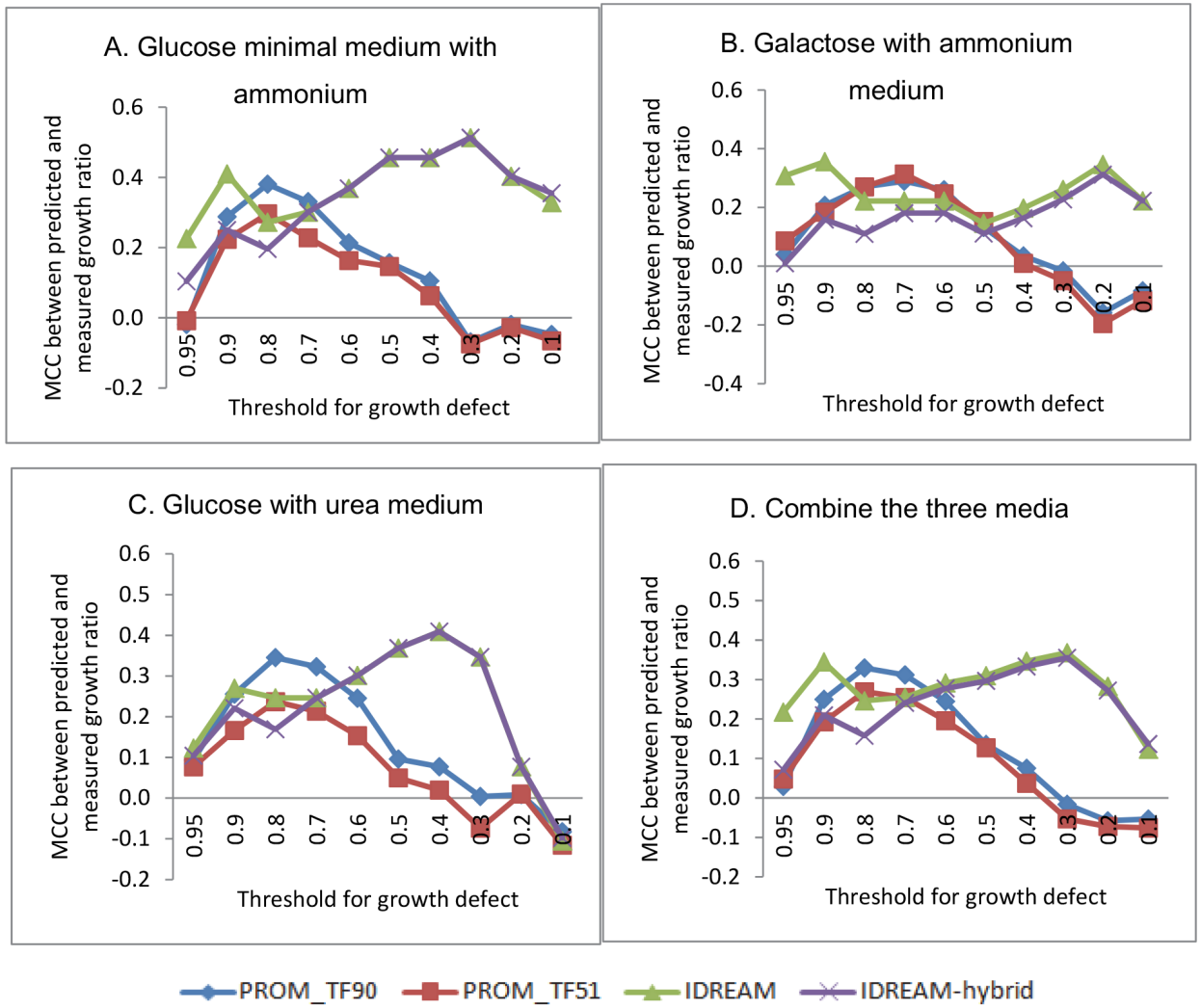
Predicted (IDREAM) vs. exp. growth

4 conditions are presented in the four panels (A, B, C and D).

Under each condition, we calculated the ratio of growth rates between TF knockout and wild-type.

When the ratio was lower than some particular **threshold**, the corresponding TF is considered growth defective.

By adjusting the threshold of growth ratio from 0.1 to 0.95, the MCCs between prediction and measurement were derived.



Wang … Price (2017) PLoS Comput Biol 13: e1005489.

OptRAM algorithm

AIM: OptRAM (Optimization of Regulatory And Metabolic Network) identifies combinatorial optimization strategies including overexpression, knockdown or knockout of both TFs and metabolic genes, based on the IDREAM integrated network framework.

Considered in silico mutations:

Table 1. Mutations of the gene expression fold change (FC) over wildtype.

Overexpression FC	2	4	8	16	32
Knockout FC	0.001				
Knockdown FC	1/2	1/4	1/8	1/16	1/32

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

OptRAM algorithm

The expression level of TFs and metabolic genes will be translated to corresponding metabolic reactions by the integrative network.

First, expression levels of metabolic genes are calculated according to the expression of corresponding TFs.

The EGRIN algorithm generates a linear equation of the target gene and the TFs:

$$target = coeff_1 TF_1 + coeff_2 TF_2 + \cdots + coeff_n TF_n$$

where *target* : expression level of a target gene regulated by *n* TFs,
TF_i : expression level of these TFs, and
$coeff_i$: corresponding coefficients of each TF.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

OptRAM – links TF -> target gene expression

In OptRAM, for a target gene regulated by one TF, *tfExpr* is the relative expression level of the mutated TF.

Then the relative expression level of the target gene is calculated as:

$$targExpr = 2^{coeff \times \log_2 tfExpr}$$

When a target gene is affected by more than one TF, the expression level of the target gene is calculated as:

$$targExpr = 2^{\sum_i^n coeff_i \times \log_2 tfExpr_i}$$

Having the relative expression level of all metabolic genes, the change of relevant reactions, represented as FC(R), is calculated according to the gene-reaction rules in the metabolic model.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

References fluxes from pFBA

In order to simulate the flux change of reactions induced by the gene expression mutation, we first need a **reference flux** value for each reaction, which is obtained by the pFBA (parsimonious enzyme usage FBA) method.

pFBA is an algorithm based on FBA. For a metabolic network with *M* metabolites and *N* reactions, the FBA formulation is as follows:

$$Maximize\ v_{objective}$$

$$Subject\ to\ \sum_{j=1}^{N} S_{ij} v_j = 0, i = 1, \ldots, M$$

$$lb_j \leq v_j \leq ub_j, j = 1, \ldots, N$$

where S_{ij} : stoichiometric coefficient of metabolite *i* in reaction *j*,
v_j : flux of reaction *j*,
lb_j (lower bound) and ub_j (upper bound) : constraints for reaction *j*.

The most commonly used **objective function** ($v_{objective}$) is biomass synthesis.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

pFBA: most efficient solution

The pFBA algorithm is divided into three steps.

(1) The **max biomass rate** is obtained by FBA with the original model.

(2) The constraint of biomass is set equal to the max biomass value.

(3) A new objective function is set as the **minimization of total flux values** carried by all reactions. Then, an optimal flux distribution is computed that maintains optimal growth.

This proxy computes the pFBA optima, representing the set of genes associated with all maximum-growth, minimum-flux solutions, thereby predicting the most **stoichiometrically efficient pathways**.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

Optimization criterion BPCY

In previous meta-heuristic strain optimization methods, such as OptGene, BPCY (**biomass-product coupled yield**) is used as the **objective function**

$$BPCY = \frac{Product \times Growth}{Substrate}$$

Product : flux of the reaction synthesizing the desired product,
Growth : flux of biomass, and
Substrate : uptake rate of the nutrient substrate.

The ultimate goal of the optimization algorithm is to identify the mutated solution with the largest BPCY value, which ensures a considerable growth when improving the target product.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

Limitations of BPCY

A limitation of the simulation using pFBA is that this framework does not guarantee that the target reaction flux will be coupled to biomass.

That is, even if the BPCY score of a mutated solution is high, the flux value of the target reaction is unstable with the max biomass.

Because the flux variability of target reaction is a wide range and the minimum flux may even be zero, there is no guarantee that the target product can have a certain output under natural growth.

Moreover, since the objective function of pFBA is biomass, there is often no flux through the desired target reaction, although the flux range of that reaction may be 0 to a positive value.

In this situation, BPCY remains 0 and the algorithm reports no feasible solution.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

Optimization criterion of OptRAM

Shen *et al*. defined a new objective function in OptRAM to couple maximizing biomass production and the target reaction of interest.

$$Obj = \frac{Target \times Growth}{Substrate} \times \left(1 - \log \frac{Range}{Target}\right)$$

where $Target = \frac{V_{max}+V_{min}}{2}$, $Range = \frac{V_{max}-V_{min}}{2}$

V_{max} : maximum flux value of target reaction

V_{min} : minimum flux value of target reaction by FVA (flux variability analysis).

Target : average flux value of target product.

Range : half of the interval between min and max target flux value.

When V_{min} is 0, $\frac{Range}{Target} = 1$, the coefficient $\left(1 - \log \frac{Range}{Target}\right)$ is 1.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

Optimization criterion

And when V_{min} is greater than 0, $\frac{Range}{Target} < 1$

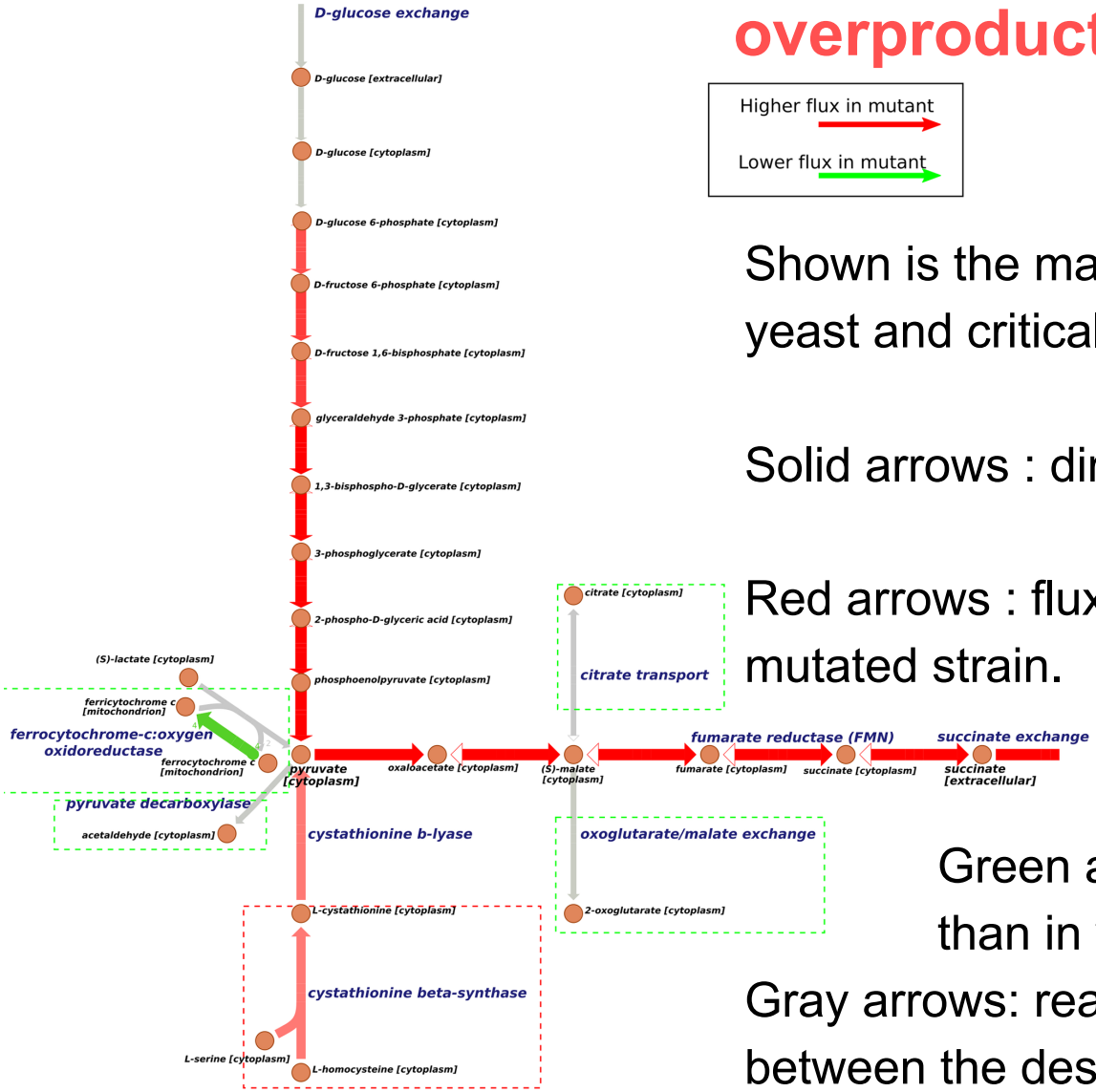
the coefficient in the () bracket will be greater than 1,
which is essentially a reward coefficient for BPCY.

Compared to BPCY, this objective function will induce solutions to have a higher and narrower flux range of target product, which reduces the uncertainty caused by alternative solutions in constraint-based modeling.

Hence, by using the refined objective function, OptRAM can provide solutions with better biomass-product coupled.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

Flux comparison of mutated model and wt for succinate overproduction



Higher flux in mutant

Lower flux in mutant

Shown is the main path of **succinate production** in yeast and critical reactions identified by OptRAM.

Solid arrows : direction of metabolic reactions.

Red arrows : fluxes are predicted to be higher in the mutated strain.

Green arrows: flux is predicted to be lower than in wildtype.

Gray arrows: reactions are not significantly different between the designed strain and the wildtype.

Red dotted boxes: critical up-regulated reactions.

Green dotted boxes : down-regulated reactions.

Shen et al. PLoS Comp Biol 15: e1006835 (2019)

Stochastic Dynamics simulations of a photosynthetic vesicle

where bioinformatics meets biophysics

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

Photons

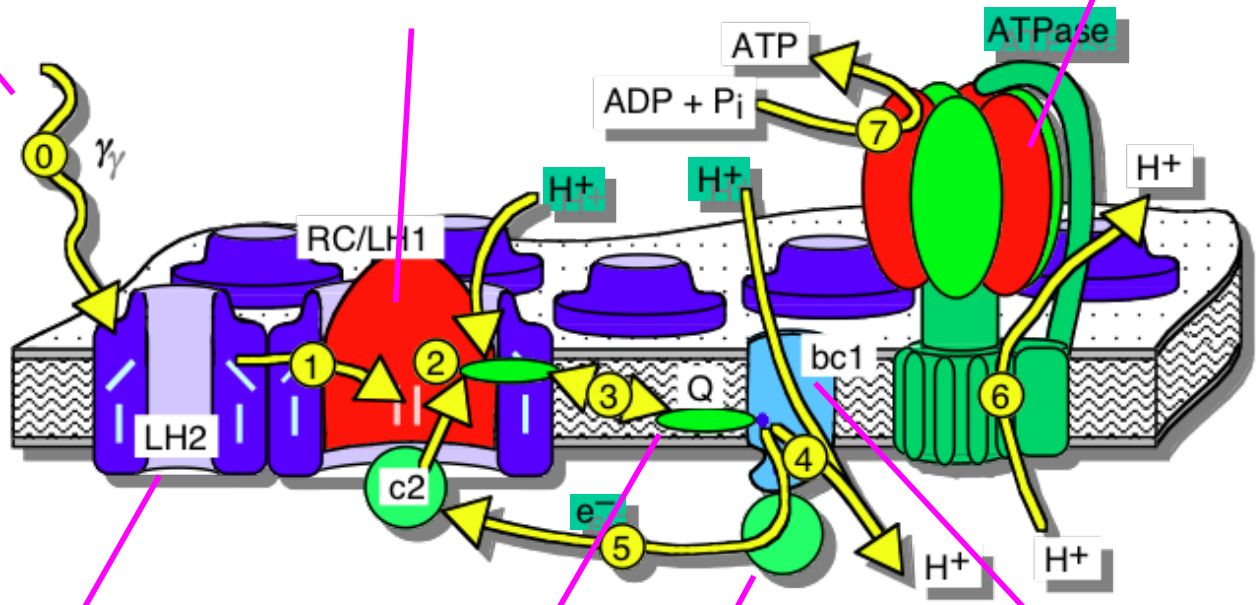
light energy

Reaction Center

$e^- - H^+$ –pairs

ATPase

chemical energy



outside

inside

Light Harvesting Complexes

electronic excitation

ubiquinon

cytochrome c_2

electron carriers

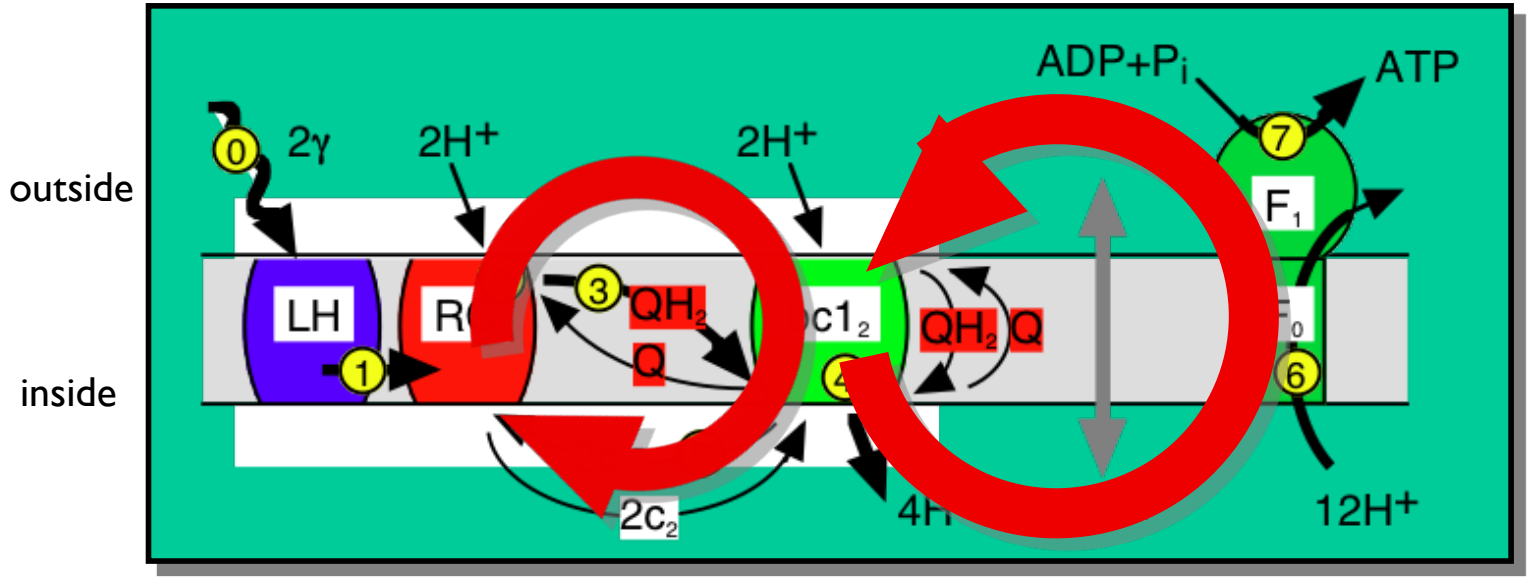
cytochrome bc_1 complex

H^+ gradient; transmembrane potential

Photosynthesis – cycle view

The conversion chain: stoichiometries must match turnovers!

light energy → electronic excitation → $e^- - H^+$ –pairs → H^+ gradient, transmembrane voltage → chemical energy



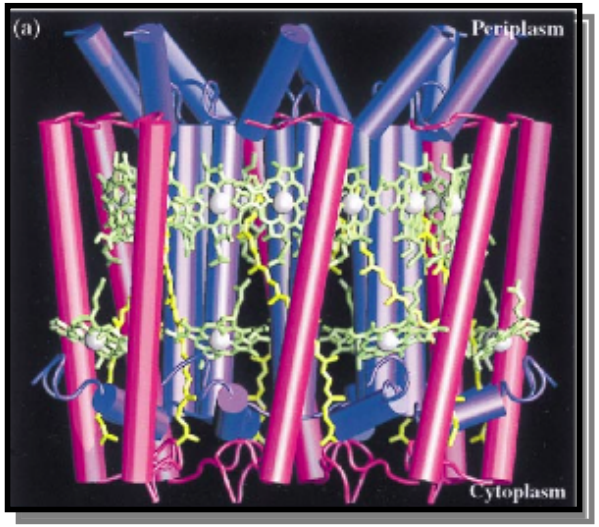
2 cycles:

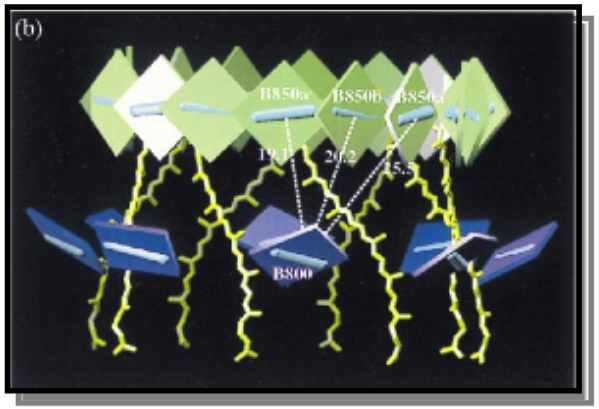
electrons protons

LH1 / LH2 / RC — a la textbook

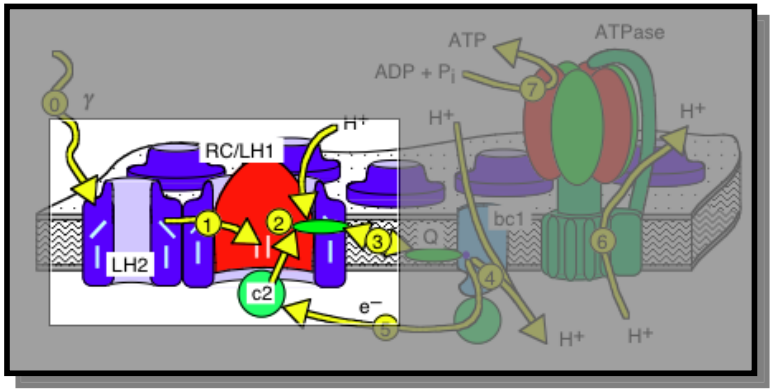
Collecting photons

LH2: 8 αβ dimers

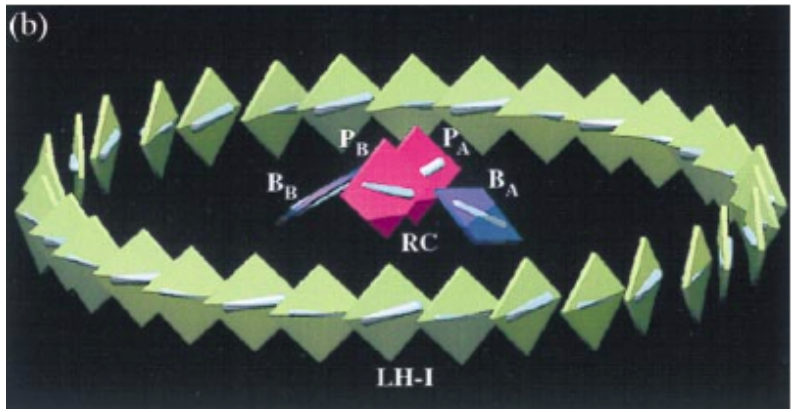


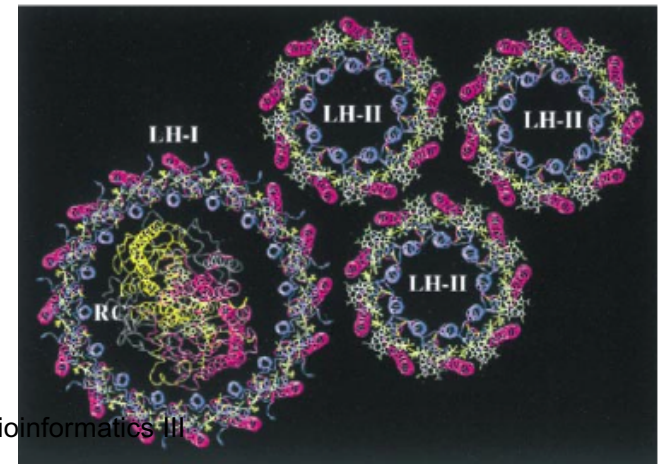


B800, B850, Car.



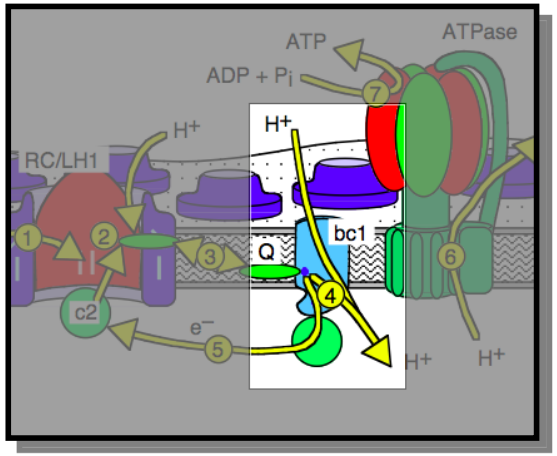
LH1: 16 αβ dimers



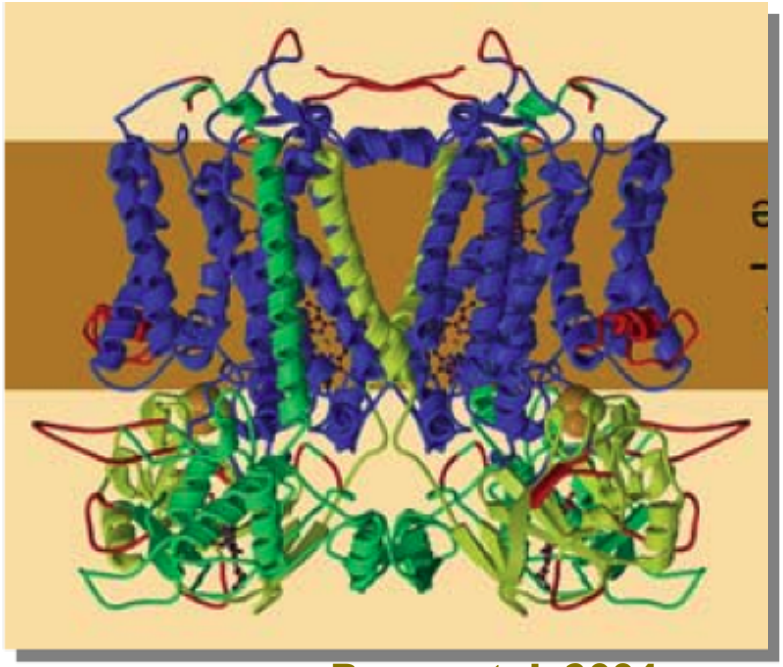


Hu *et al*, 1998

The Cytochrome bc_1 complex

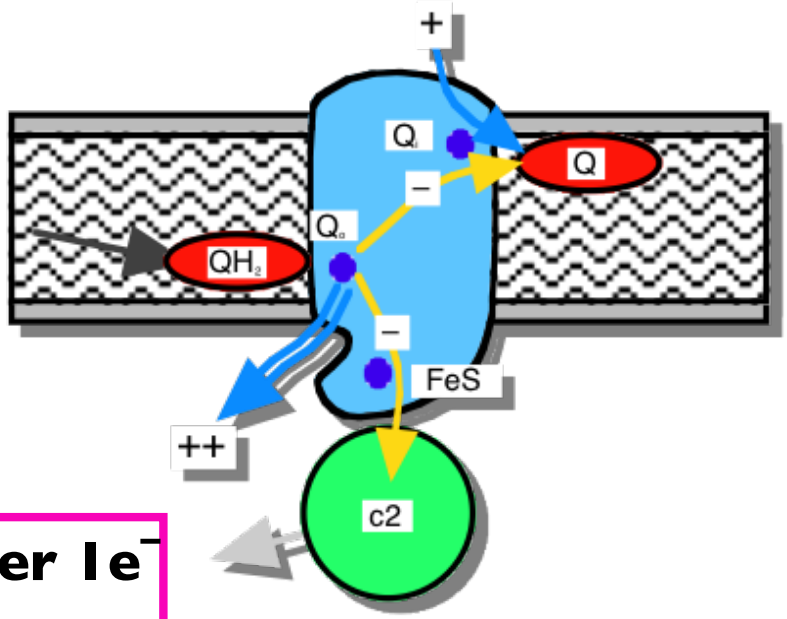


the "proton pump"



Berry, etal, 2004

Q-cycle:



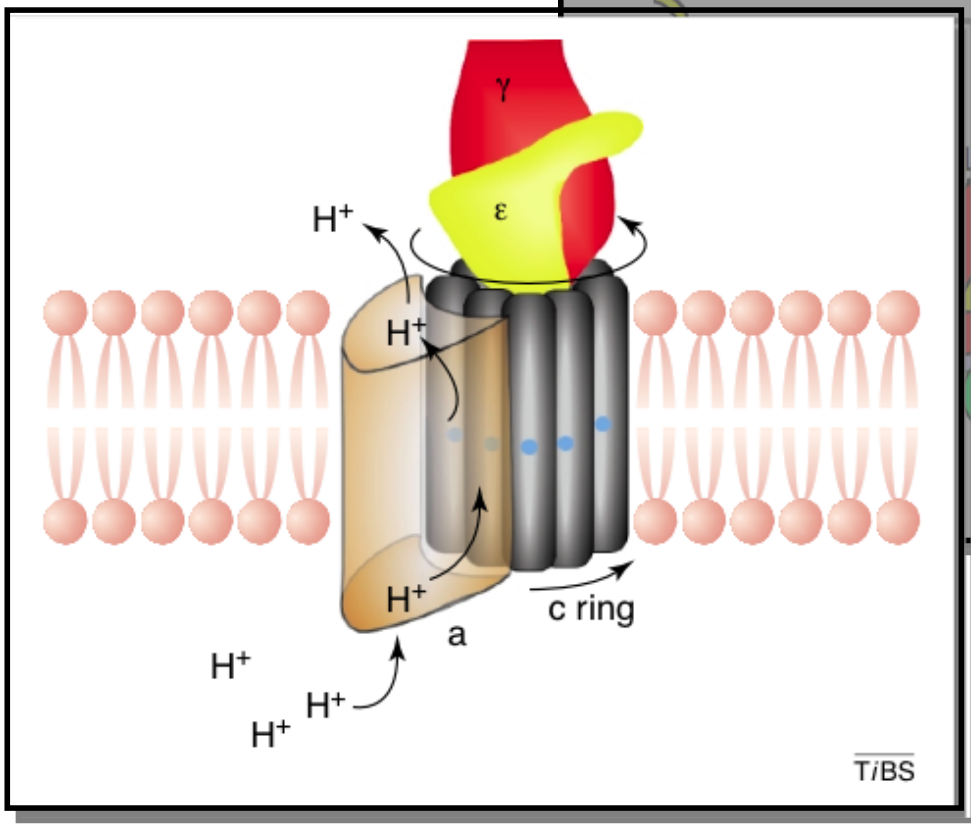
$2H^+$ per $1e^-$

X-ray structures known

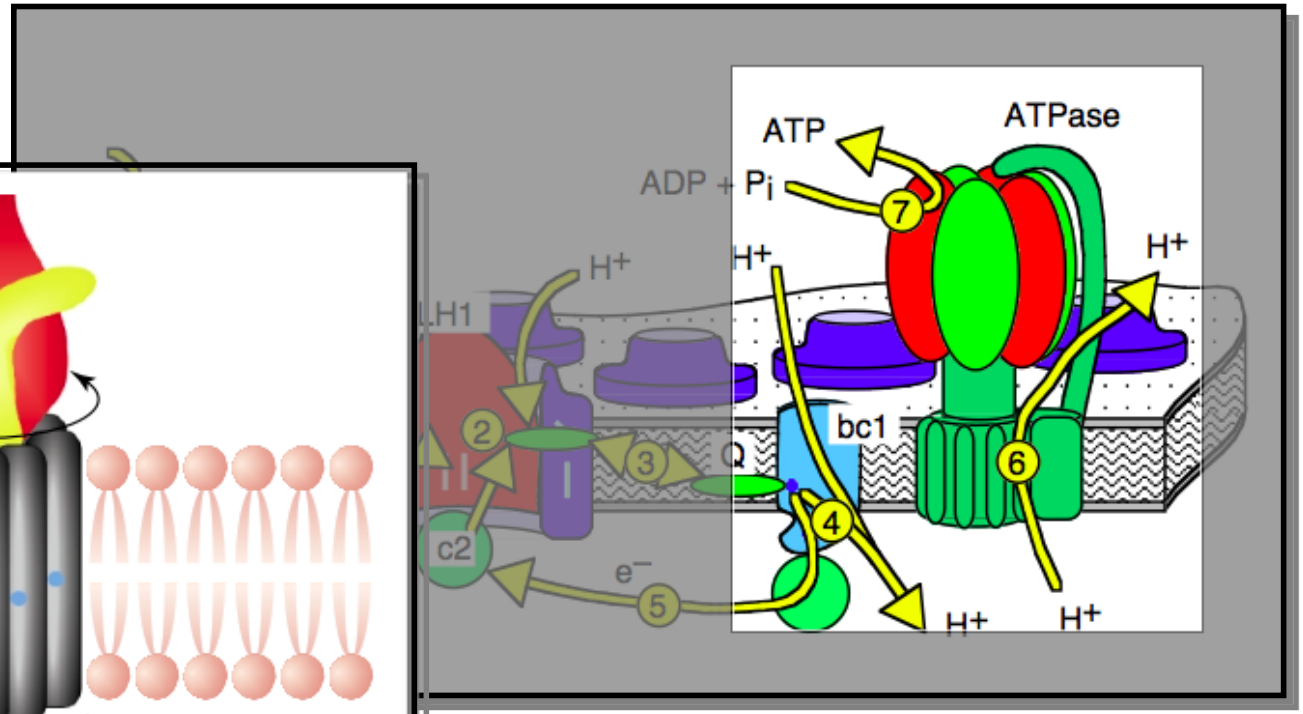
always forms a dimer

The F_oF_1-ATP synthase I

at the end of the chain: producing ATP from the H+ gradient



Capaldi, Aggeler, 2002



per turn:

10–14 H^+ $\longrightarrow$ 3 ATP

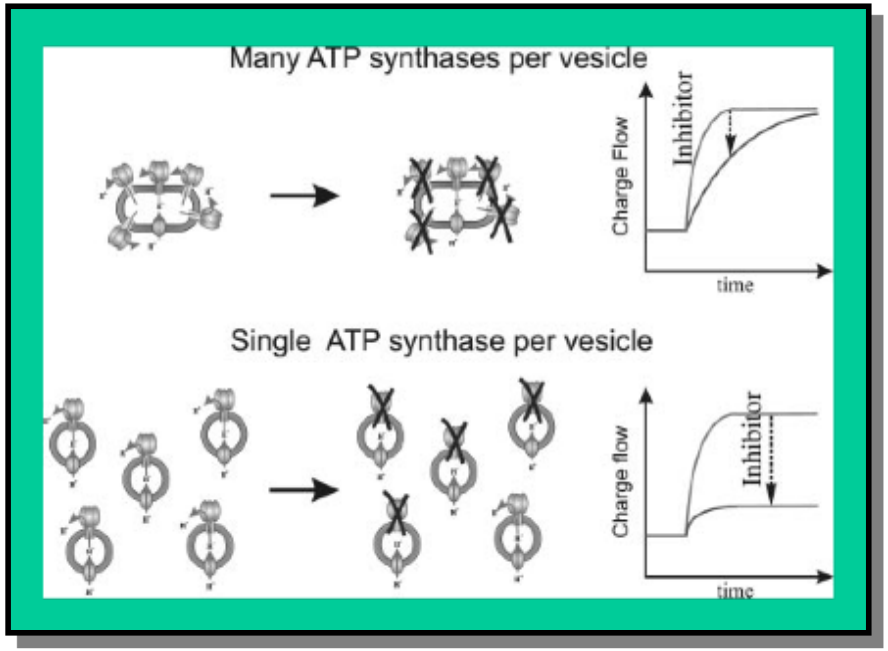
$\boxed{1\ \text{ATP} \triangleq 4\ H^+}$

The F_1F_0-ATP synthase

"…mushroom like structures observed in AFM images…"

ATPase is "visible"

1 ATPase per vesicle

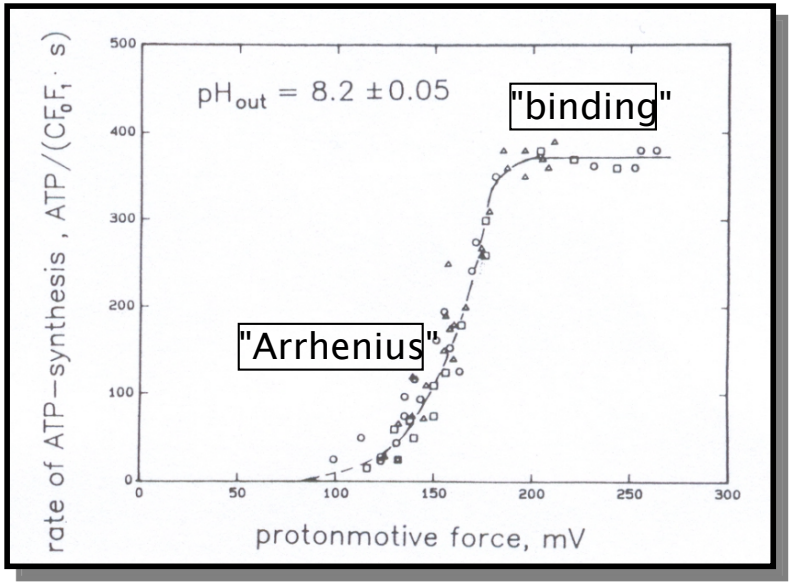


Feniouk et al, 2002

per turn: 10–14 H^+ per 3 ATP

1 ATP ≙ 4 H^+

limited throughput of the ATPase



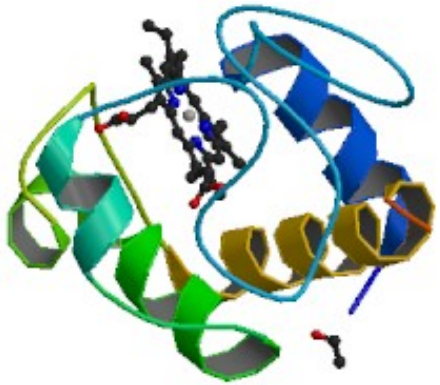
ATPase from	ATP/s	H^+/s
chloroblasts	<400	1600
E. coli	<100	400

Gräber et al, 1991, 1999

The electron carriers

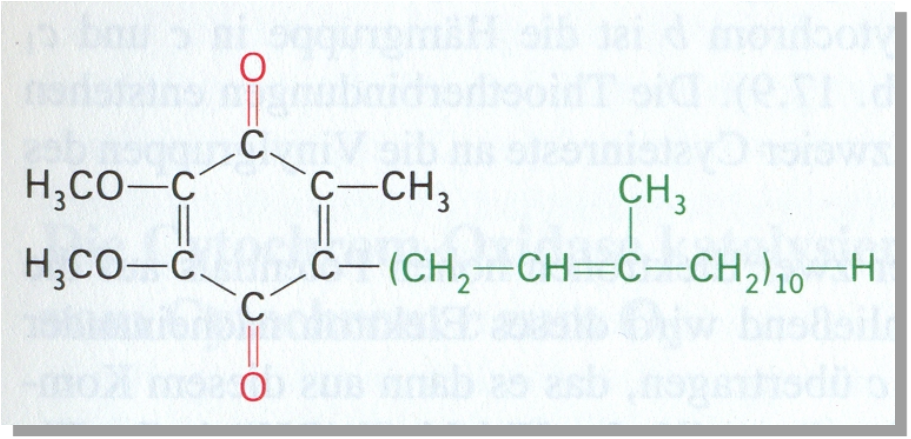
Cytochrome *c*: carries electrons from bc_1 to RC

- heme in a hydrophilic protein shell
- 3.3 nm diameter, water-soluble



Ubiquinone UQ10:
carries electron-proton pairs
from RC to bc_1

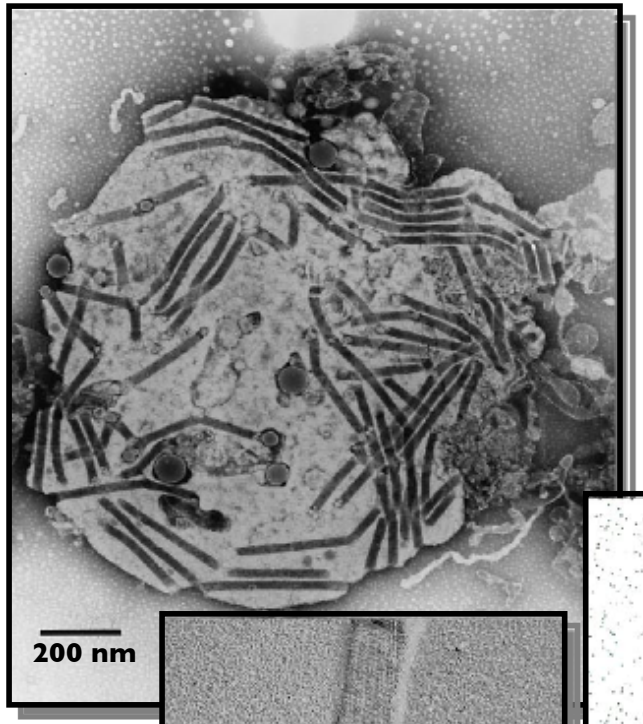
- long (2.4 nm) hydrophobic isoprenoid tail, membrane-soluble

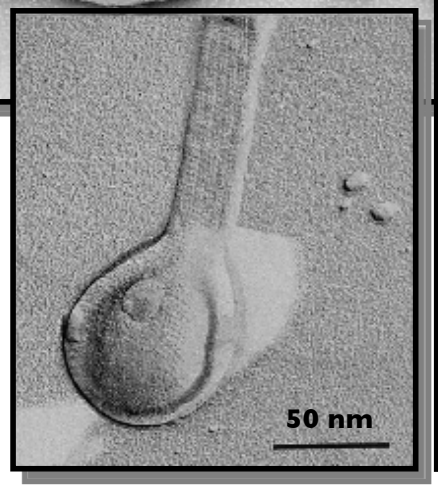


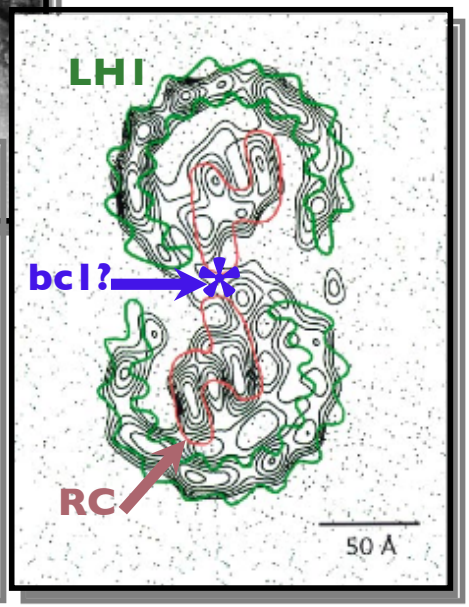
taken from Stryer

Tubular membranes – photosynthetic vesicles where are the bc_1 complexes and the ATPase?

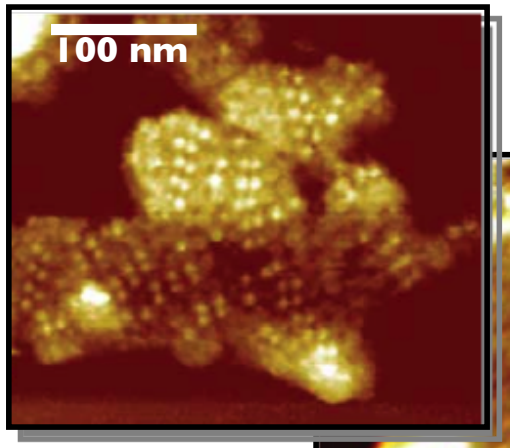
Jungas *et al.*, 1999



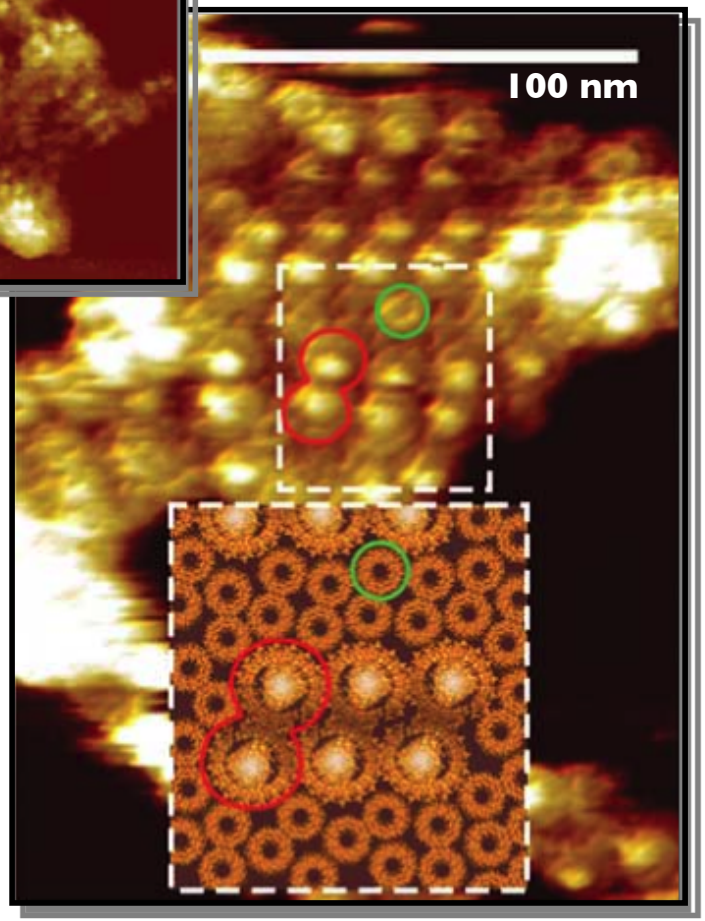




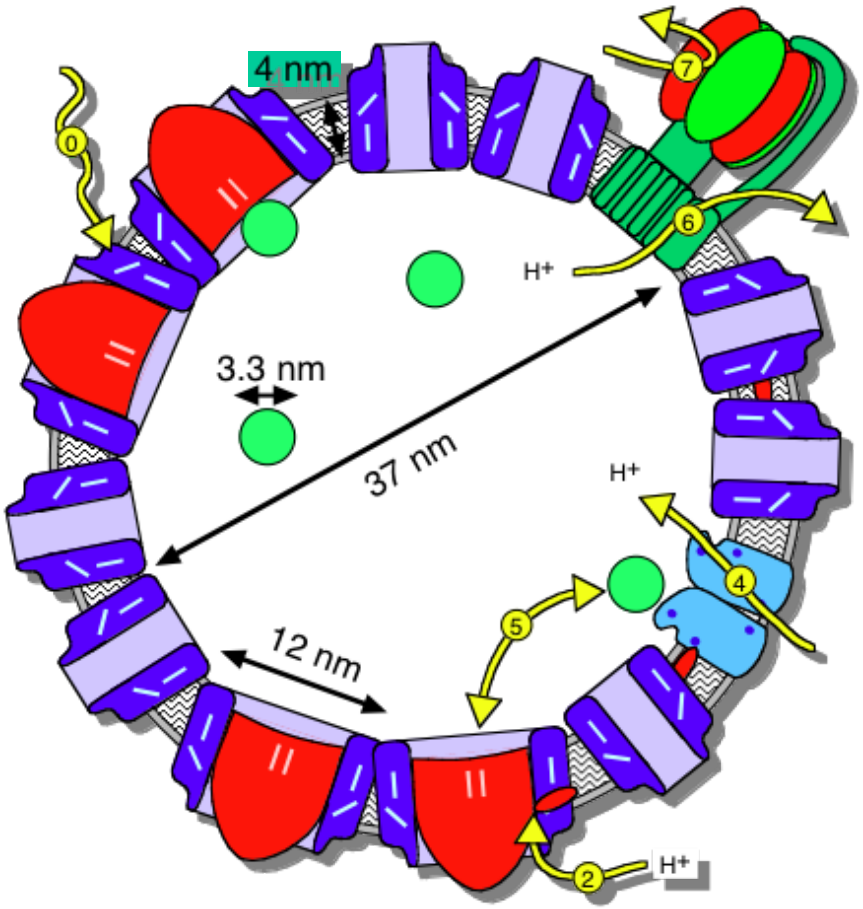
Bahatyrova *et al.*, 2004



no bc_1 found!



Chromatophore vesicle: typical form in *Rh. sphaeroides*



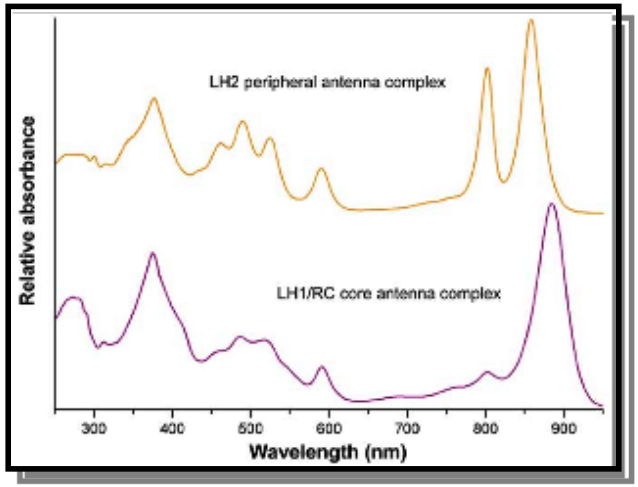
Lipid vesicles
30–60 nm diameter
H^+ and cyt *c* inside

average chromatophore vesicle, 45 nm Ø: surface 6300 nm²

Vesicles are really small!

Photon capture rate of LHC's

relative **absorption spectrum** of LH1/RC and LH2

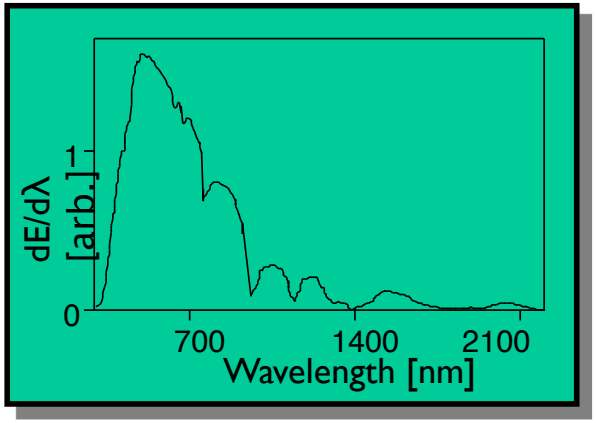


Cogdell etal, 2003

+ Bchl extinction coeff. normalization (σ_{Bchl} = 2.3 Å^2)

Franke, Amesz, 1995

sun's spectrum at ground (total: 1 kW/m^2)



Gerthsen, 1985

multiply

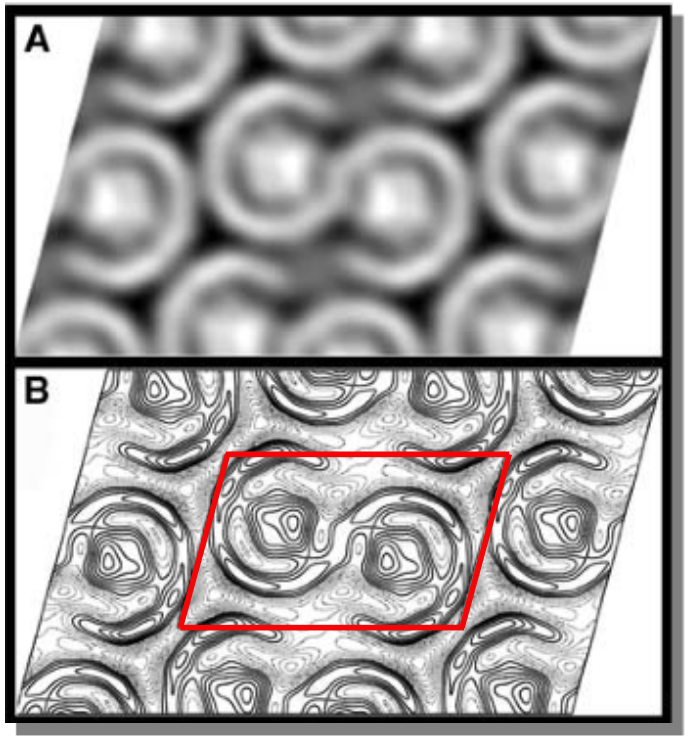
capture rate: $0.1 \frac{\gamma}{s\ kW\ Bchl}$

typical growth condition: 18 W/m^2 Feniouk et al, 2002

LH1:	16 * 3 Bchl	14 γ/s
LH2:	10 * 3 Bchl	10 γ/s

LH1 / LH2 / RC — native

electron micrograph
and density map



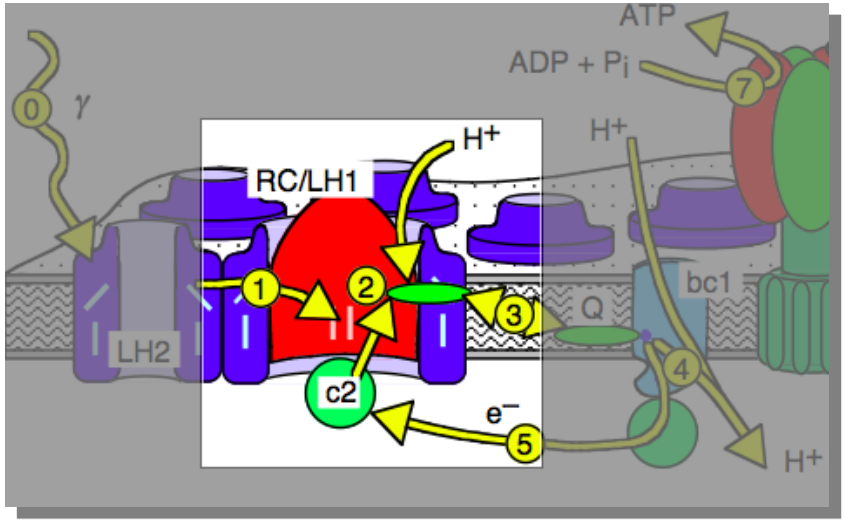
Siebert et al, 2004
125 * 195 Å², γ = 106°

	Area per:	per vesicle (45 nm)
LH1 monomer (hexagonal)	146 nm²	
LH1 dimer	234 nm²	
LH2 monomer	37 nm²	
$LH1_2$ + 6 LH2	456 nm²	11

Chromatophore vesicle, 45 nm Ø: surface 6300 nm²

Photon processing rate at the RC

Which process limits the RCs turnover?



Unbinding of the quinol

25 ms Milano *et al.* 2003

+ binding, charge transfer

≈ 50 ms per quinol (estimate)

with $2e^-$ H^+ pairs per quinol

40–50 γ/s per RC

≈ 22 QH_2/s

1 RC can serve 1 LH1

+ 3 LH2

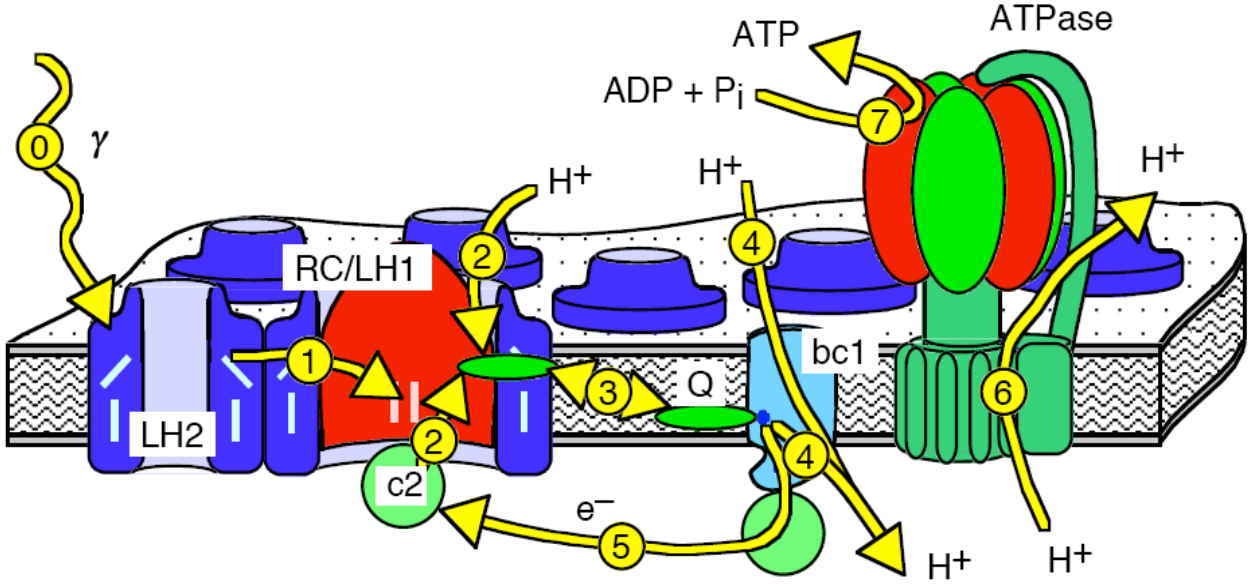
= 44 γ/s

$LH1_2$ + 6 LH2 ≙ 456 nm² → 11 LH1 dimers including 22 RCs on one vesicle

480 Q/s can be loaded @ 18 W/m² per vesicle

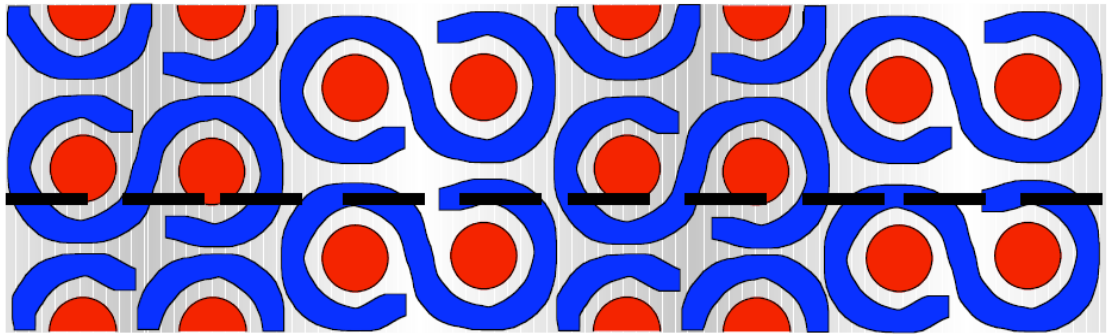
Parameters

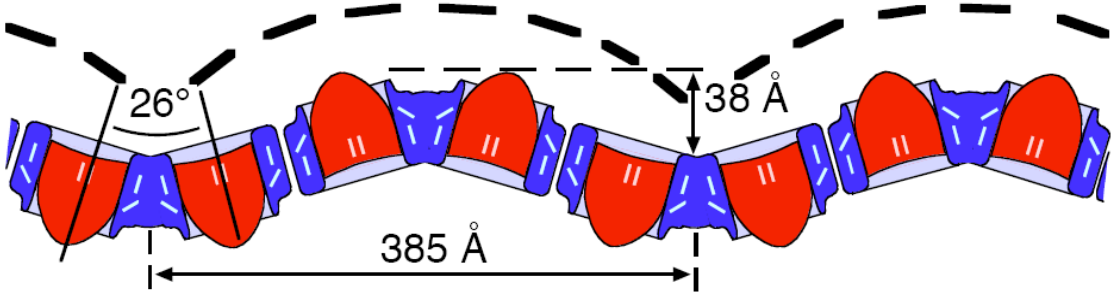
protein	throughput per protein (natural units)	H^+ equivalents per protein $[1/s]$	total number per avg. vesicle of 45 nm diameter	rate determined from	explained in section
LH2	$10\,\gamma/s$	20	60	absorption spectra +	III A
LH1 dimer	$2 \times 14\gamma/s$	56	10	+ light intensity of 18 W/m^2	III A
RC	22 QH2/s	88	20	QH2 (un)binding	III B
bc1 dimer	$\leq 2 \times 42$ c2/s	168	3 (... 10)	measured activity at ΔpH = 0	III C
ATPase	≤ 100 ATP/s	400	1	measured throughput	III D
cytochrome c_2	80 e^-/s	160	20	(un)binding at the bc1	V A (III B, III C)
ubiquinone	$10 \times 2(e^- H^+)$/s	40	100	(un)binding at the RC and the bc1	V A (III B, III C)

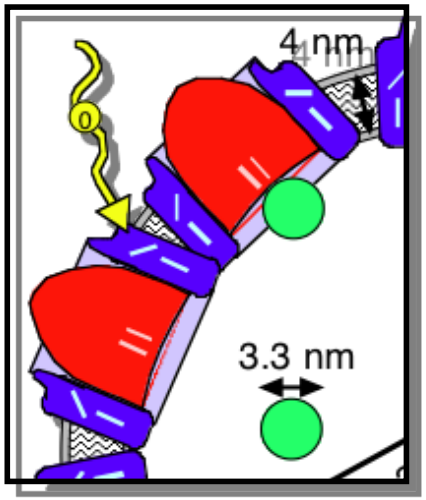


reconstituted LH1 dimers in planar lipid membranes explain intrinsic curvature of vesicles

Drawn after AFM images of Scheuring *et al* of LH1 dimers reconstituted into planar lipid membranes.







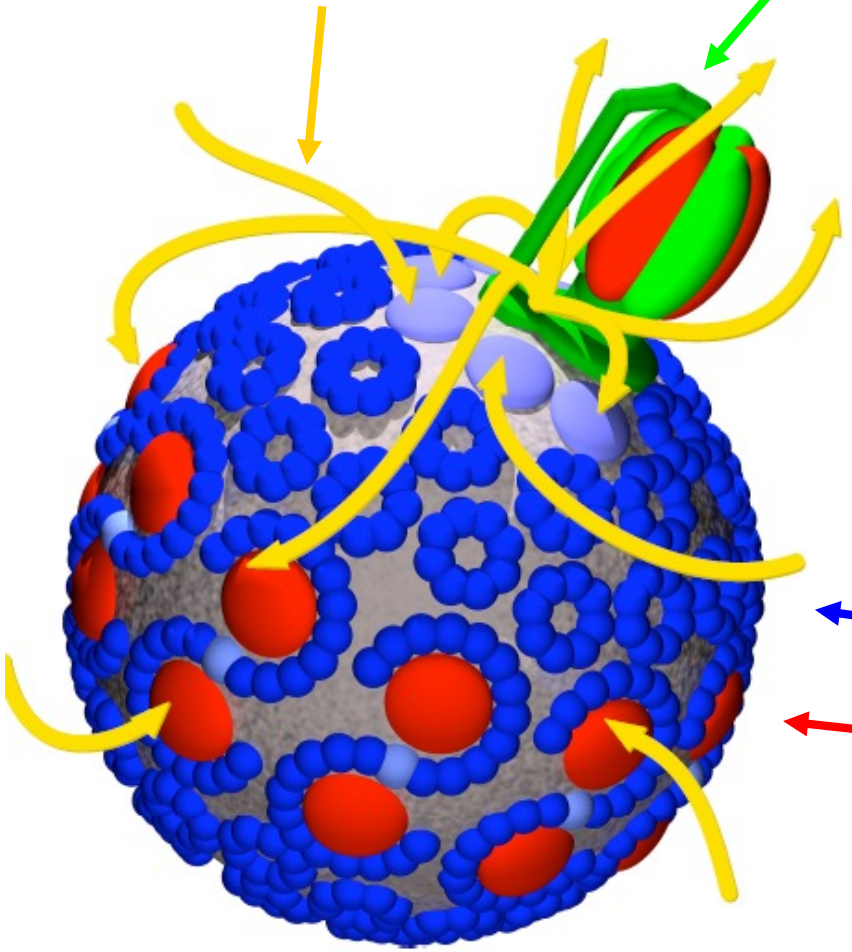
Values fit nicely to the proposed arrangement of LH1 dimers, when one assumes that they are stiff enough to retain the bending angle of 26˚ that they would have on a spherical vesicle of 45 nm diameter and taking into account the length of a single LH1 dimer of about 19.5 nm.

Proposed setup of a chromatophore vesicle

yellow arrows: diffusion of the protons out of the vesicle via the ATPase and to the RCs and bc1s.

At the „poles“
green/red: the ATPase
light blue: the bc1 complexes

Increased proton density close to the ATPase suggests close proximity of ATPase and bc_1 complexes.



blue: small LH2 rings (blue)

blue/red: Z-shaped LH1/RC dimers form a linear array around the “equator” of the vesicle, determining the vesicle’s diameter by their intrinsic curvature.

Geyer & Helms, Biophys J. (2006)

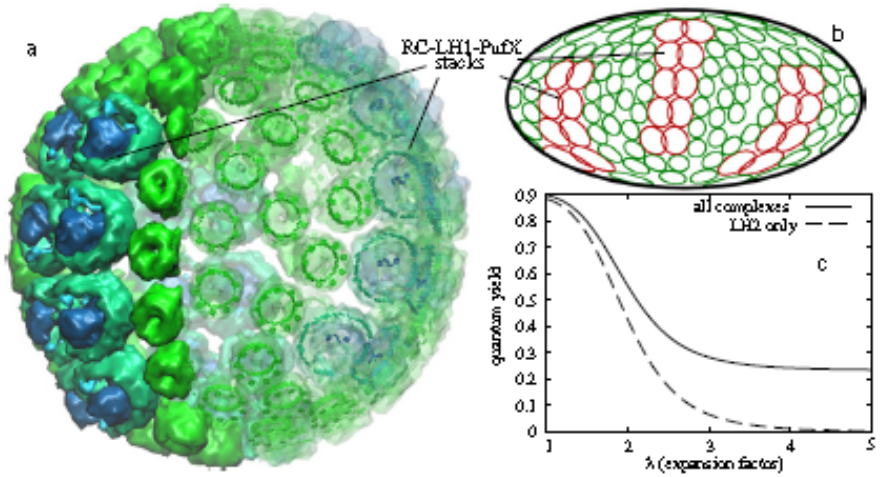
Summary

Integrated model of binding + photophysical + redox processes inside of chromatophore vesicles

Various experimental data fit well together

Equilibrium state.

How to model non-equilibrium processes?

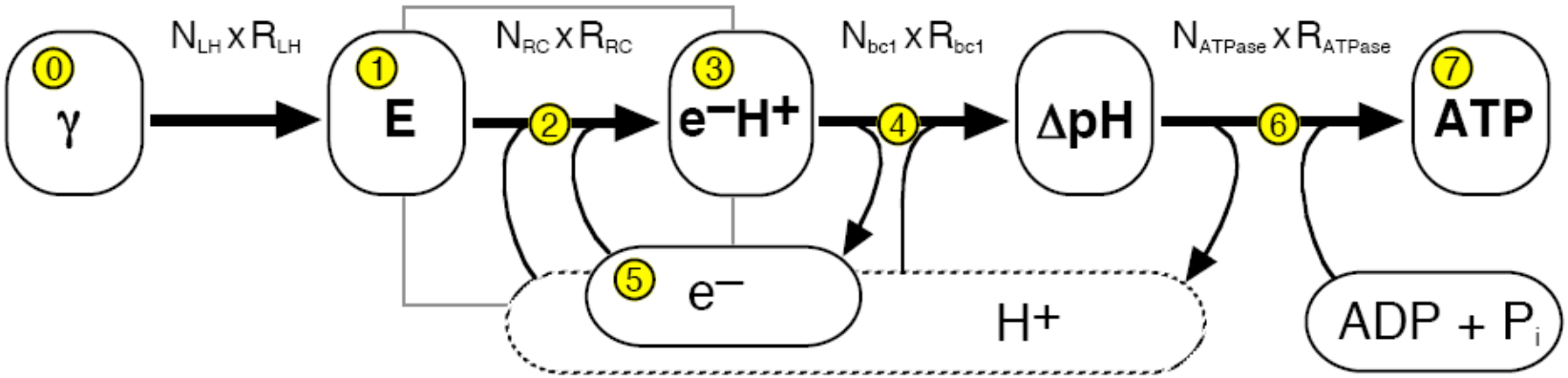


Biophysical Journal Volume 99 July 2010 67–75

Photosynthetic Vesicle Architecture and Constraints on Efficient Energy Harvesting

Melih Şener,[†‡] Johan Strümpfer,[†§] John A. Timney,[¶] Arvi Freiberg,[‖**] C. Neil Hunter,[¶] and Klaus Schulten[†‡§*]
[†]Beckman Institute for Advanced Science and Technology, [‡]Department of Physics, and [§]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois; [¶]Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, United Kingdom; and [‖]Institute of Physics and [**]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia

Viewing the photosynthetic apparatus as a conversion chain

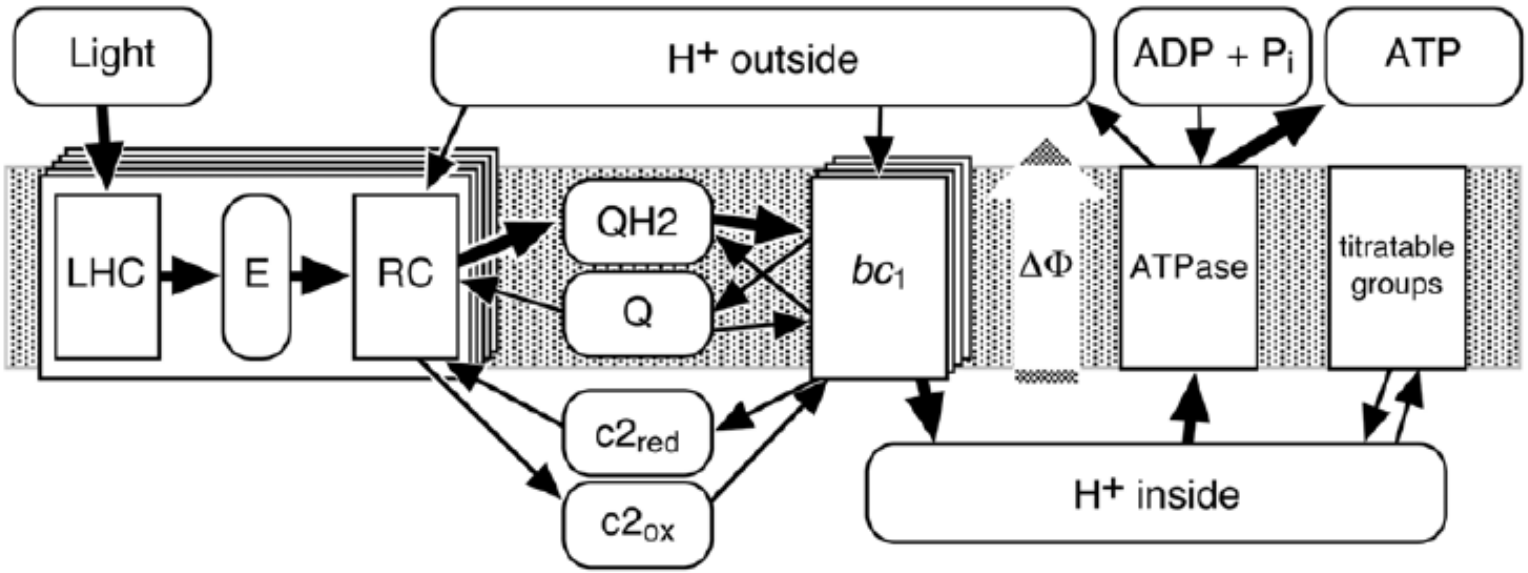


Thick arrows : path through which the photon energy is converted into chemical energy stored in ATP via the intermediate stages (rounded rectangles).

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



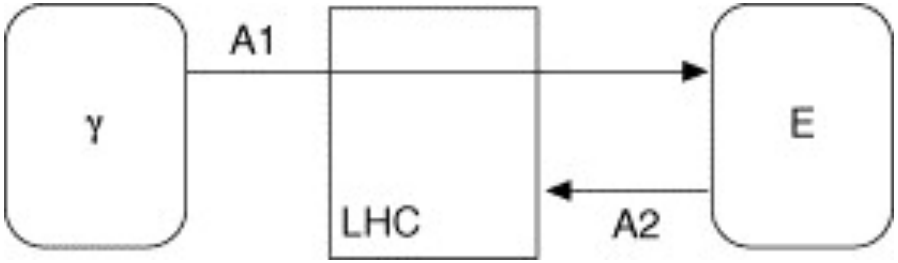
Round edges: pools for metabolite molecules

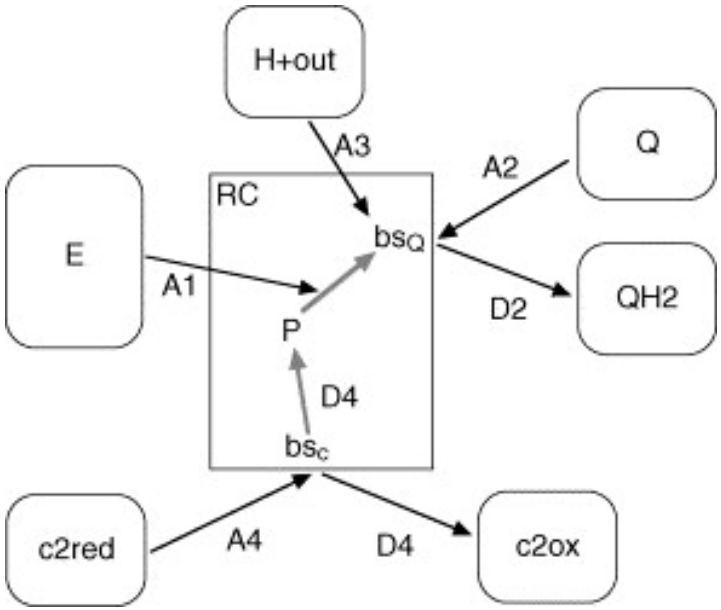
Rectangles: protein machines are modeled explicitly as multiple copies

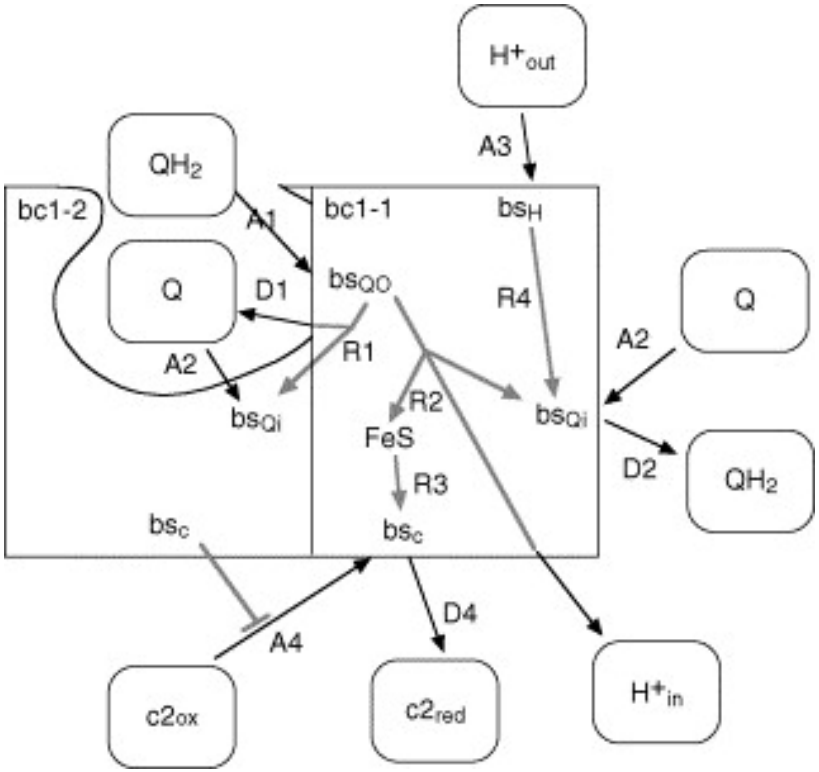
fixed set of parameters

integrate rate equations with stochastic algorithm

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)

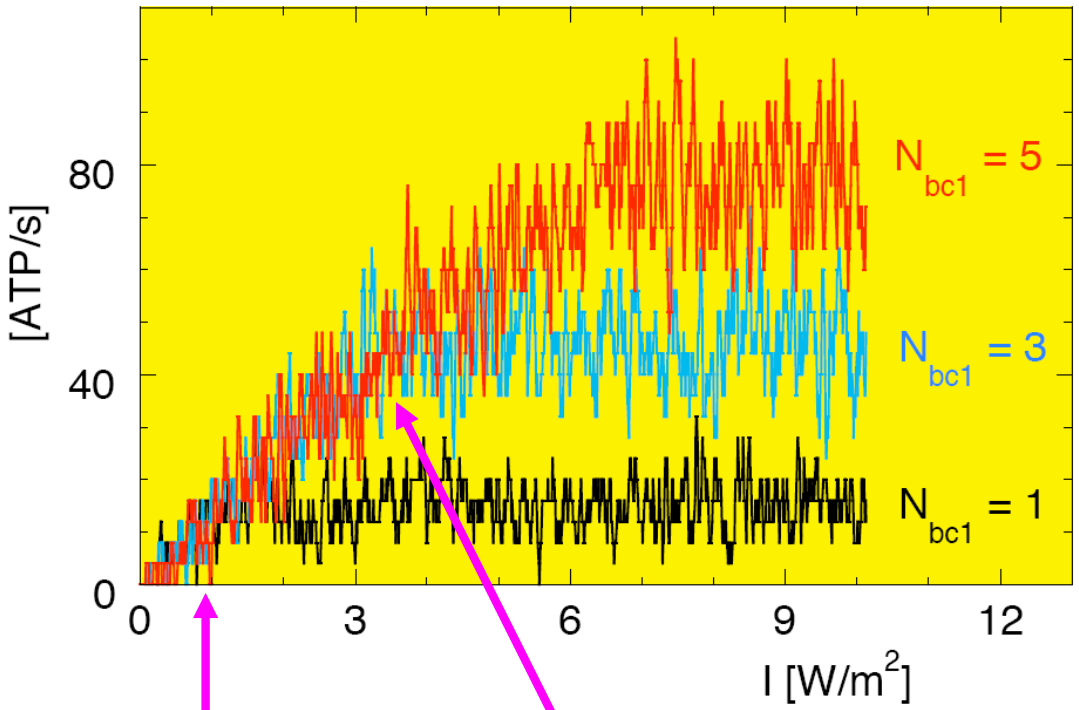
- Timer algorithm (reactions); 1 random number determines when reaction occurs

simulating 1 minute real time required 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)

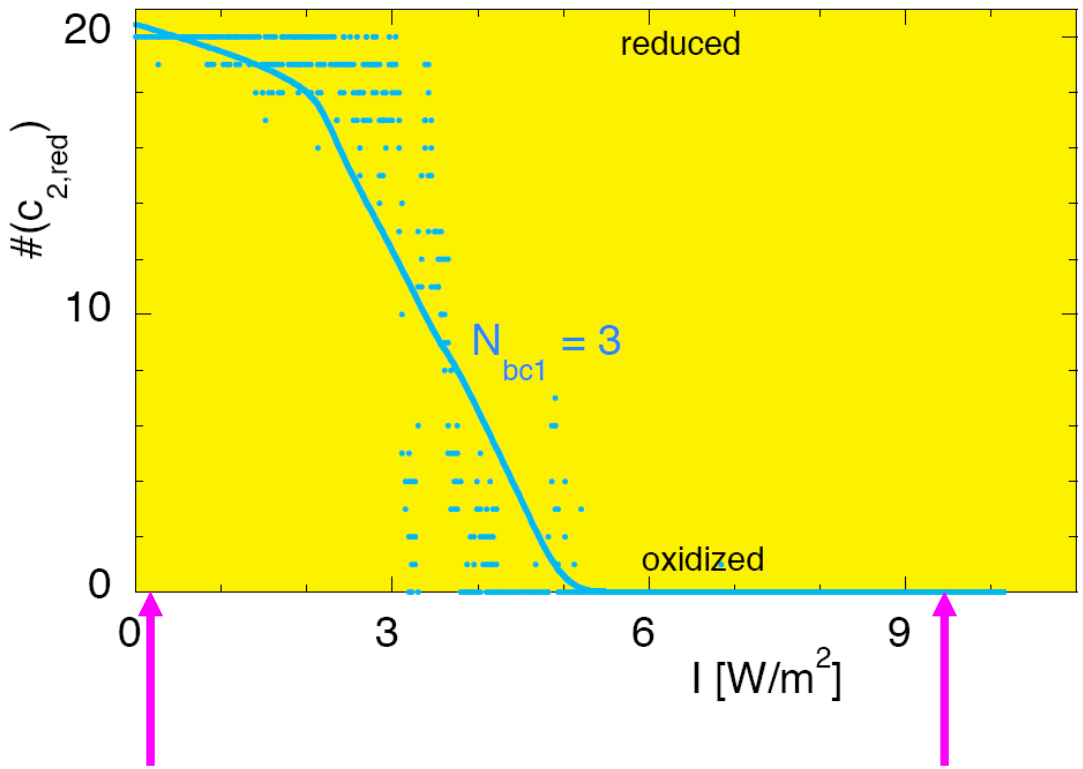
→ there are two regimes
- one limited by available light
- one limited by bc$_1$ throughput



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

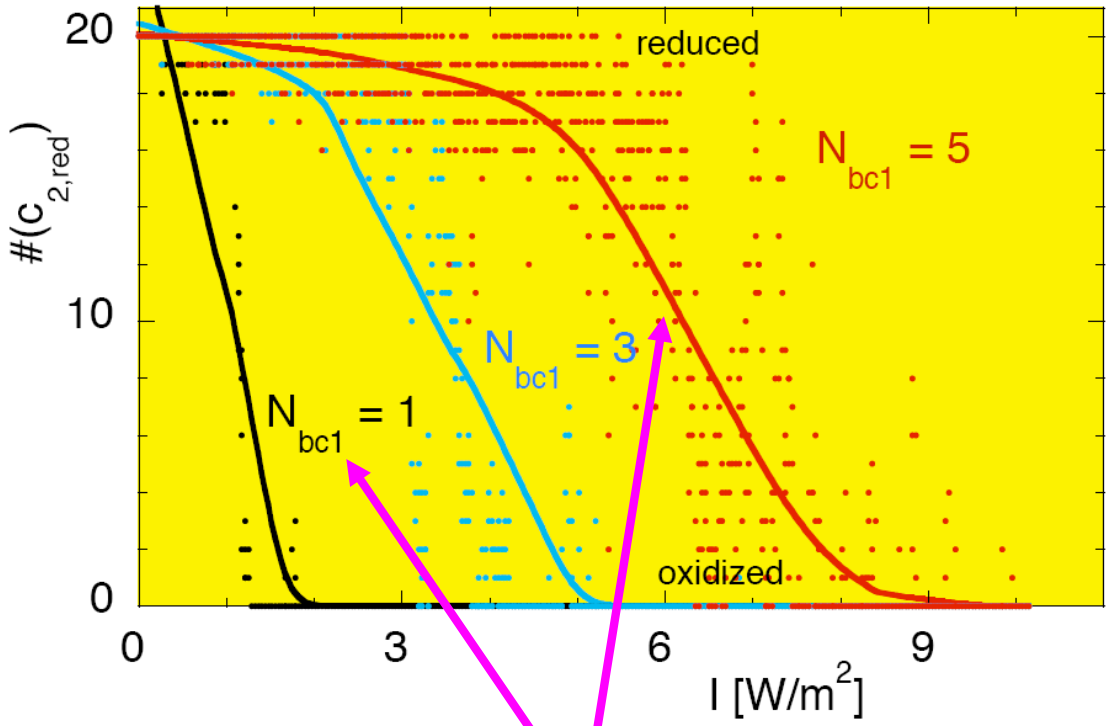
oxidation state of cytochrome c_2 pool



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

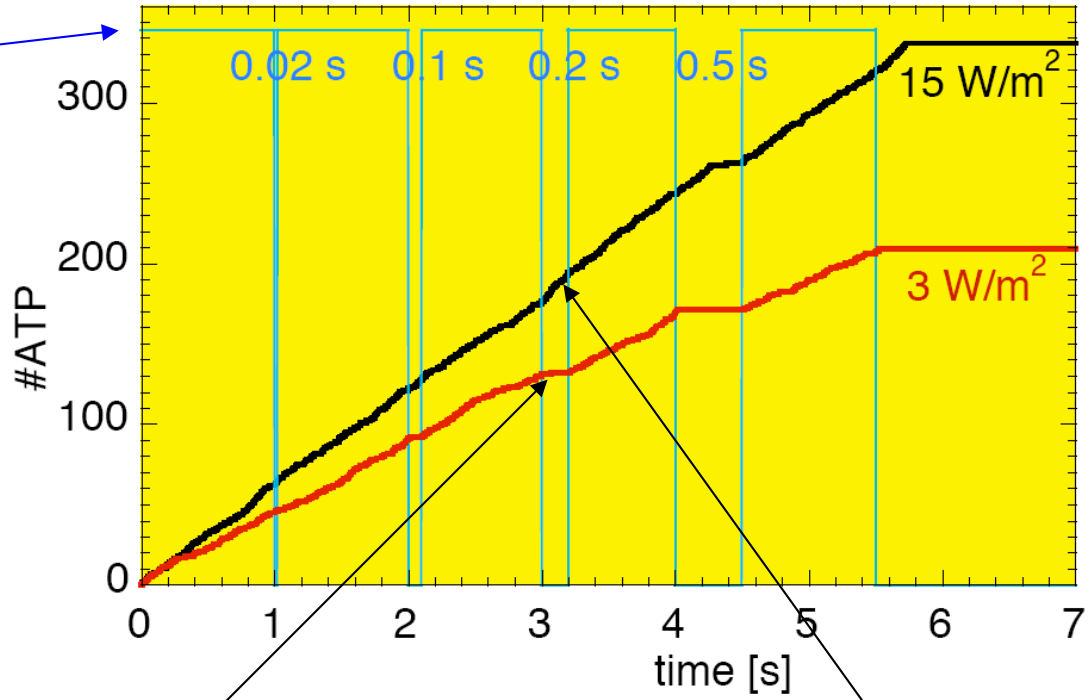
oxidation state of cytochrome c_2 pool



more bc_1 complexes
can load more
cytochrome c_2s

total number of produced ATP

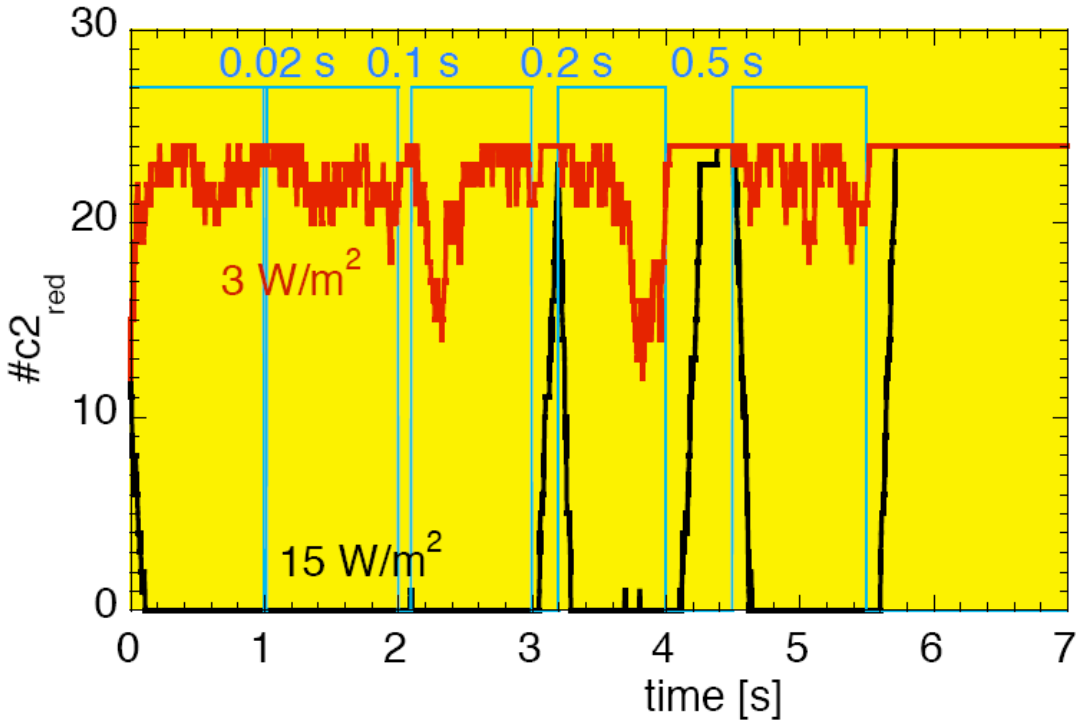
blue line:
illumination



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

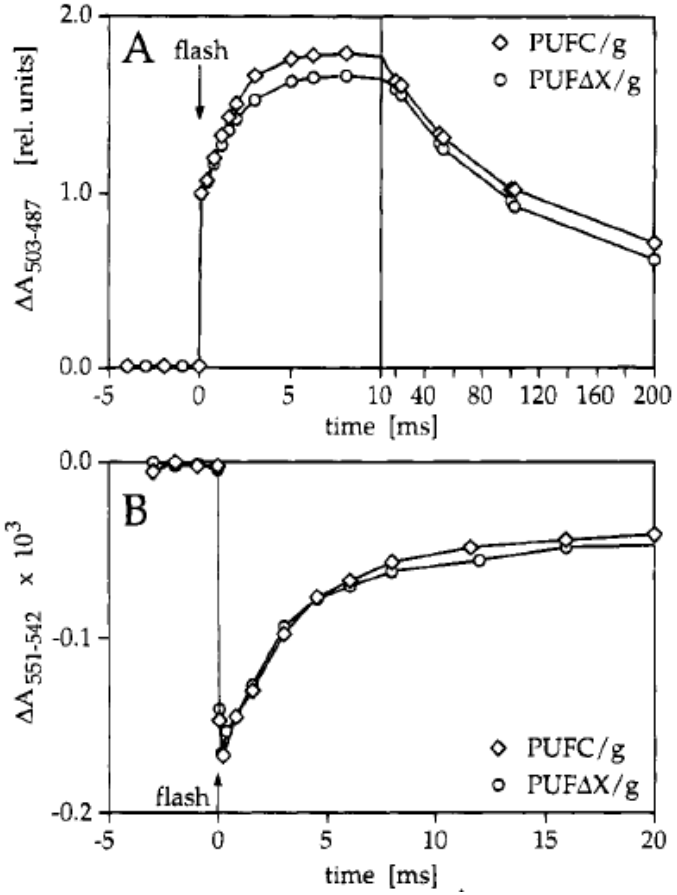
Tihamér Geyer*, Xavier Mol, Sarah Blaß, Volkhard Helms

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†

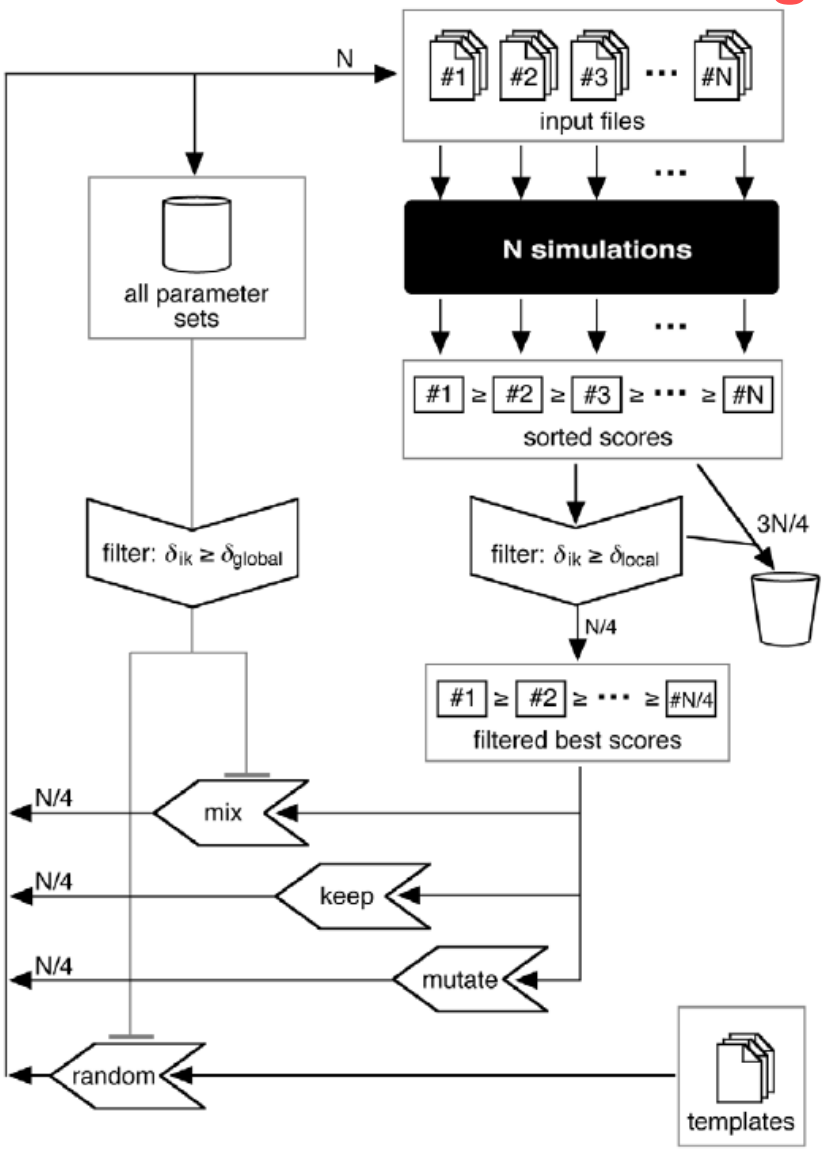
Wolfgang P. Barz,‡,§ Francesco Francia,‖ Giovanni Venturoli,‖ B. Andrea Melandri,‖ André Verméglio,⊥ and Dieter Oesterhelt*,‡

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
$bc_1::k_{tr}(e:Q_o{=>}FeS)$	$2.3 * 10^3$ s^{-1}	rate for electron transfer from Q_o to FeS
$bc_1::k_{tr}(e:c_1{=>}c_2)$	10^5 s^{-1}	electron transfer rate from c_1 to bound cytochrome c_2
$bc_1::k_{tr}(e:Q_o{=>}b_L)$	10^4 s^{-1}	electron transfer from Q_0 to b_L heme
$bc_1::k_{tr}(e:b_L{=>}b_H)$	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
$\Delta\Phi::C_{red}$	–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



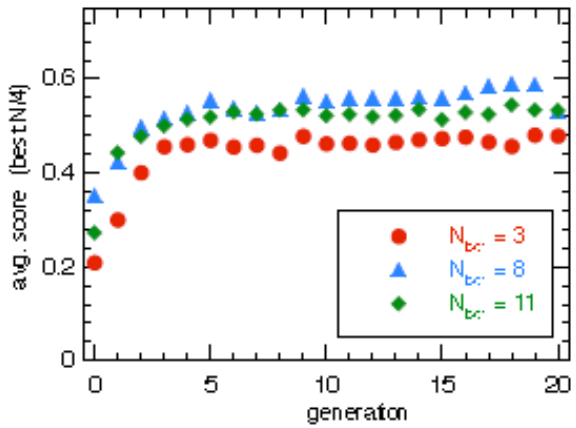


Figure S1: Determining the Number of bcl Complexes: Evolution of the Master Score

25 optimization parameters

Analyze 1000 best parameter sets among 32.800 simulations:

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

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

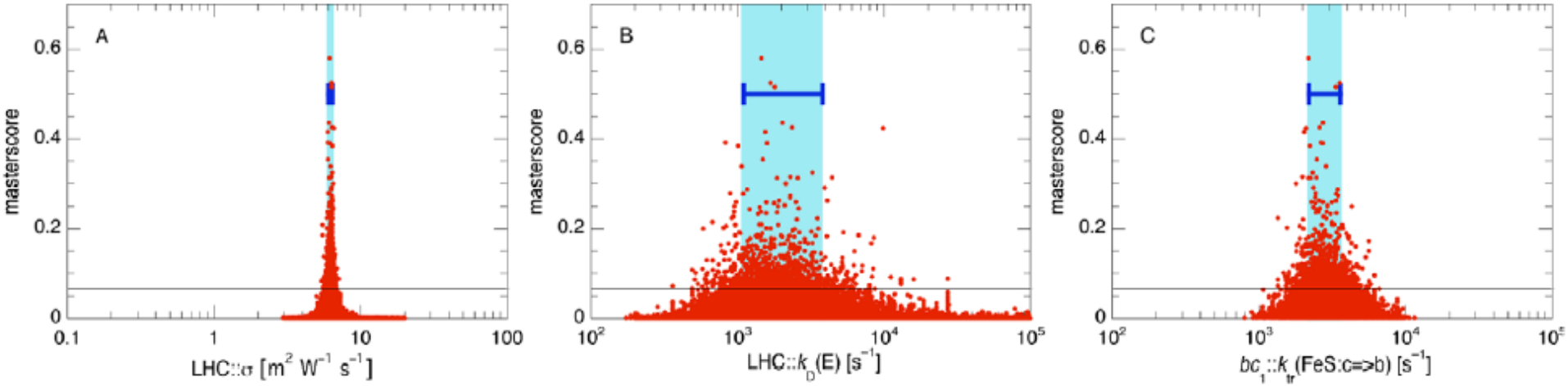
$$P_{\min} = \exp[\langle \log P \rangle - \sigma]$$

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

parameter	units	$\langle P \rangle$	$P_{min} \ldots P_{max}$	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}	$1.9 * 10^3$	$(1.1\ldots3.8) * 10^3$	0.29
RC::k_{on}(E)	s^{-1}	$2.4 * 10^6$	$(1.2\ldots4.5) * 10^6$	0.27
RC::k_{on}(H^+)	$nm^3\ s^{-1}$	$1.4 * 10^8$	$(1.3\ldots1.6) * 10^8$	0.81
RC::k_{on}(Q)	$nm^2\ s^{-1}$	$6.0 * 10^4$	$(4.4\ldots8.1) * 10^4$	0.54
RC::k_{off}(QH2)	s^{-1}	87	70…108	0.65
RC::k_{on}(c2red)	$nm^3\ s^{-1}$	$9.2 * 10^5$	$(7.3\ldots11.5) * 10^5$	0.63
RC::k_{off}(c2ox)	s^{-1}	$2.2 * 10^3$	$(1.6\ldots3.0) * 10^3$	0.53
bc1::kon(QH2@Qo)	$nm^2\ s^{-1}$	$1.2 * 10^4$	$(0.79\ldots1.7) * 10^4$	0.46
bc1::koff(Q@Qo)	s^{-1}	28.3	26.3…30.4	0.86
bc_1::k_{tr}(Q:Q_o –>Q_i)	s^{-1}	$4.9 * 10^3$	$(3.6\ldots6.7) * 10^3$	0.54
bc1::kon(Q@Qi)	$mm^2\ s^{-1}$	$6.7 * 10^5$	$(4.5\ldots10) * 10^5$	0.45
bc1::koff(QH2@Qi)	s^{-1}	86	68…110	0.62
bc_1::k_{tr}(QH2:Q_i –>Q_o)	s^{-1}	$3.8 * 10^3$	$(2.6\ldots5.5) * 10^3$	0.47
bc_1::k_{on}(c2ox)	$nm^3\ s^{-1}$	$9.4 * 10^6$	$(6.3\ldots14) * 10^6$	0.47
bc_1::k_{off}(c2red)	s^{-1}	$6.0 * 10^3$	$(3.3\ldots11) * 10^3$	0.30
bc1::koff(H+@Qo)	s^{-1}	$2.4 * 10^4$	$(1.3\ldots4.3) * 10^4$	0.30
bc_1::k_{tr}(FeS:b –>c)	s^{-1}	$3.9 * 10^3$	$(3.1\ldots5.1) * 10^3$	0.61
bc_1::k_{tr}(FeS:c –>b)	s^{-1}	$2.8 * 10^3$	$(2.2\ldots3.6) * 10^3$	0.61
bc_1::k_{tr}(e:b_H –>Q_i)	s^{-1}	$7.7 * 10^3$	$(5.0\ldots12) * 10^3$	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

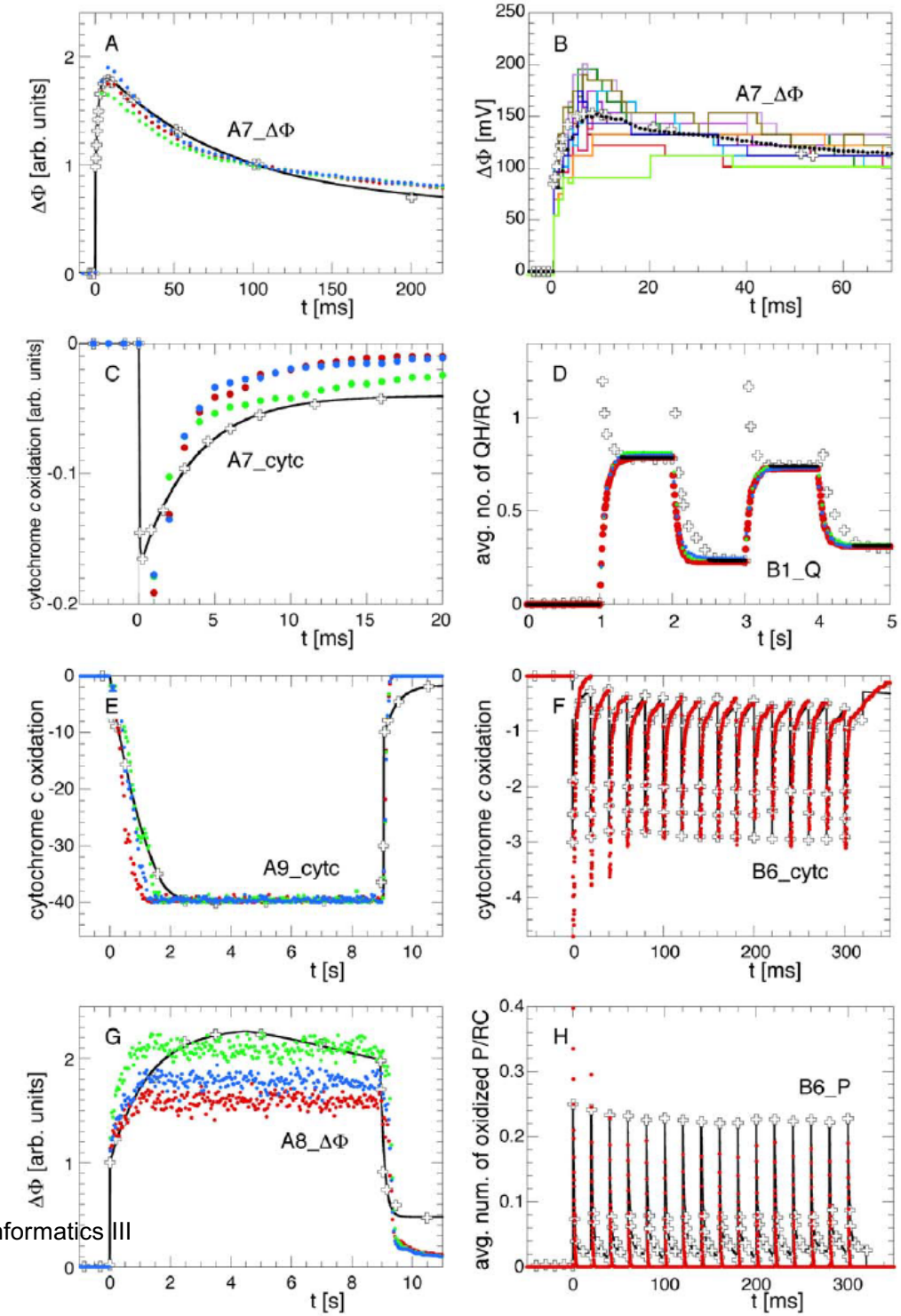
Score of individual parameter set *i* for matching one experiment:

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

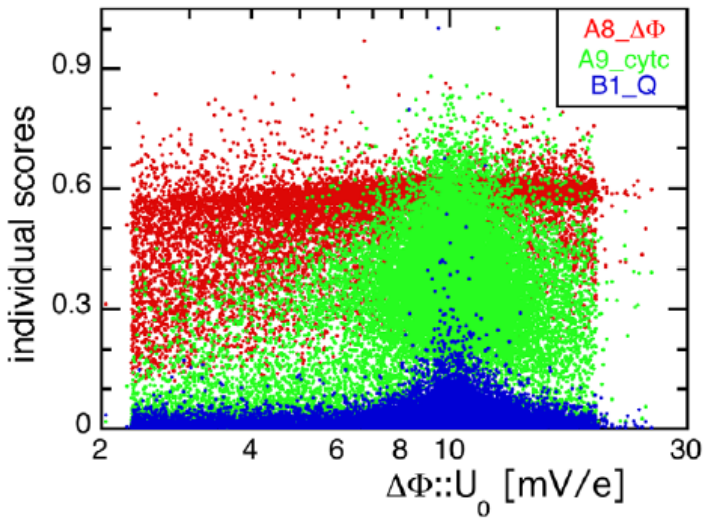
$x(t_i)$: simulation result

$f(t_i)$: 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).