Monte Carlo simulation of electron transport in a photosynthetic enzyme: comparison with experiment using Latin hypercube sampling

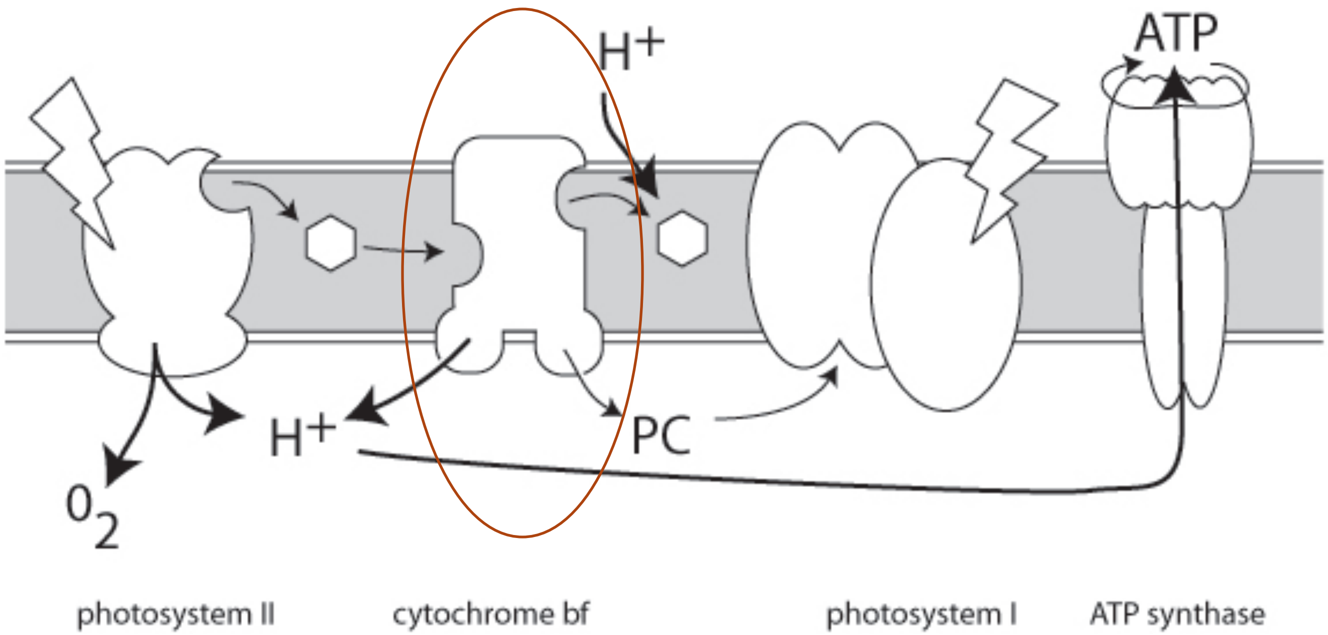
Mark Schumaker
Department of Mathematics
Washington State University

David Kramer*
Institute for Biological Chemistry
Washington State University

Present address:
DOE-Plant research laboratory
Michigan State University

Bulletin of Mathematical Biology, in press, online version available

Light Reactions of Photosynthesis

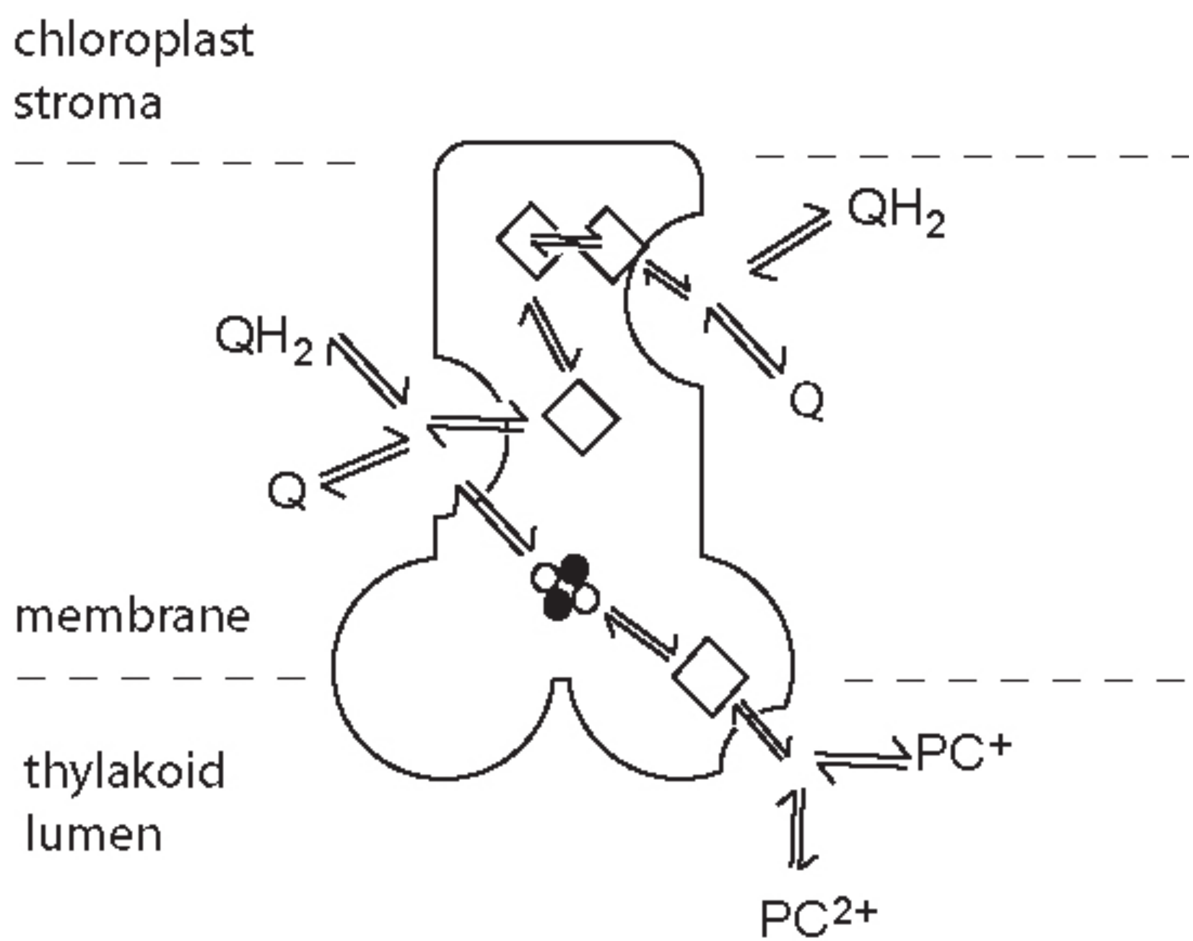


The energy of sunlight is used to separate water into oxygen and hydrogen.

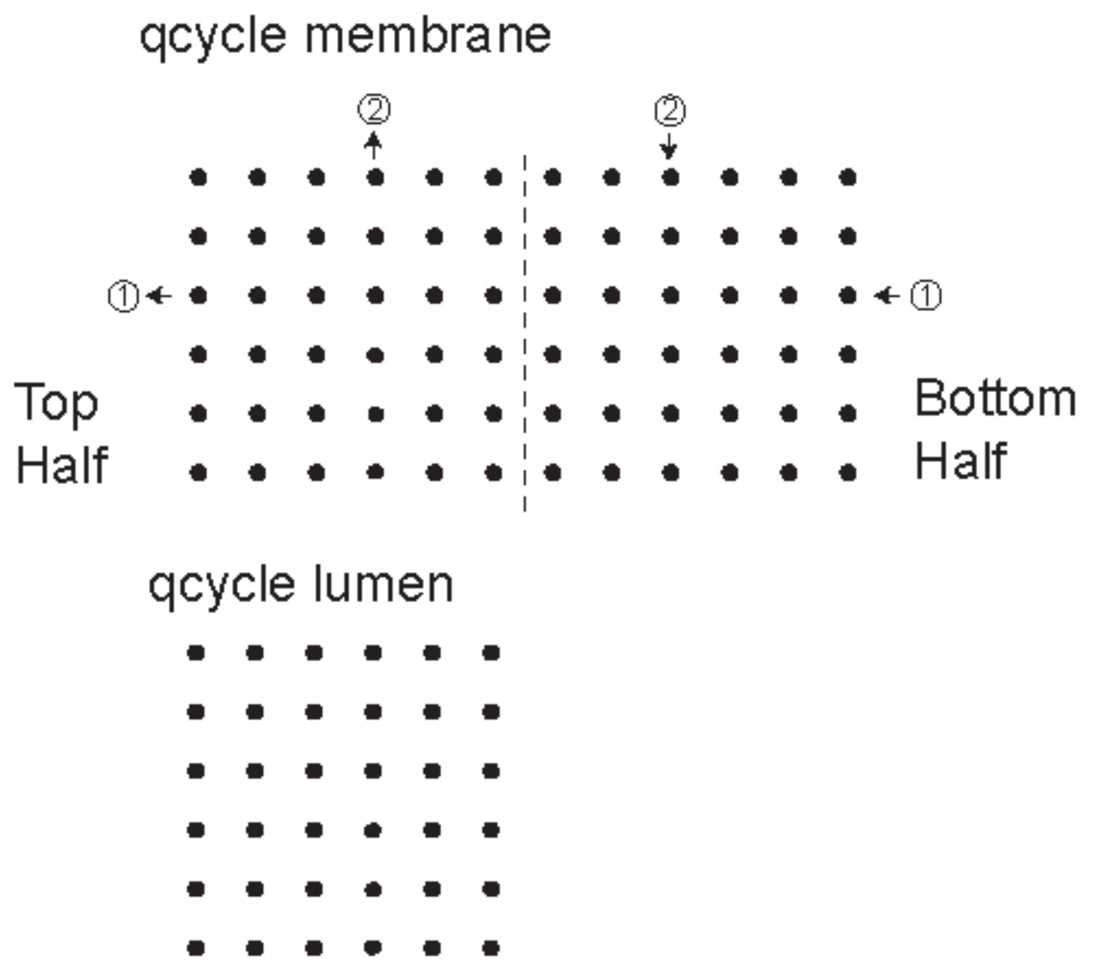
Q-cycle model of electron transport in a cytochrome b6f monomer

Low potential chain

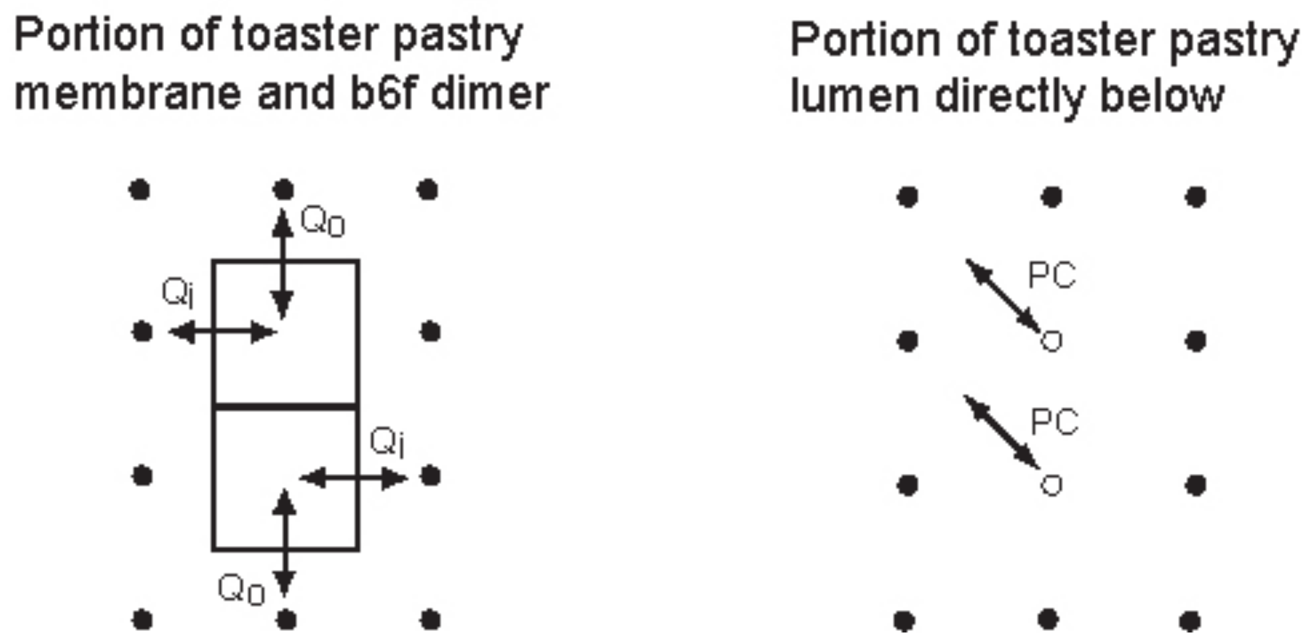
High potential chain



Toaster pastry model of thylakoid



Model of cyt b6f in the thylakoid membrane

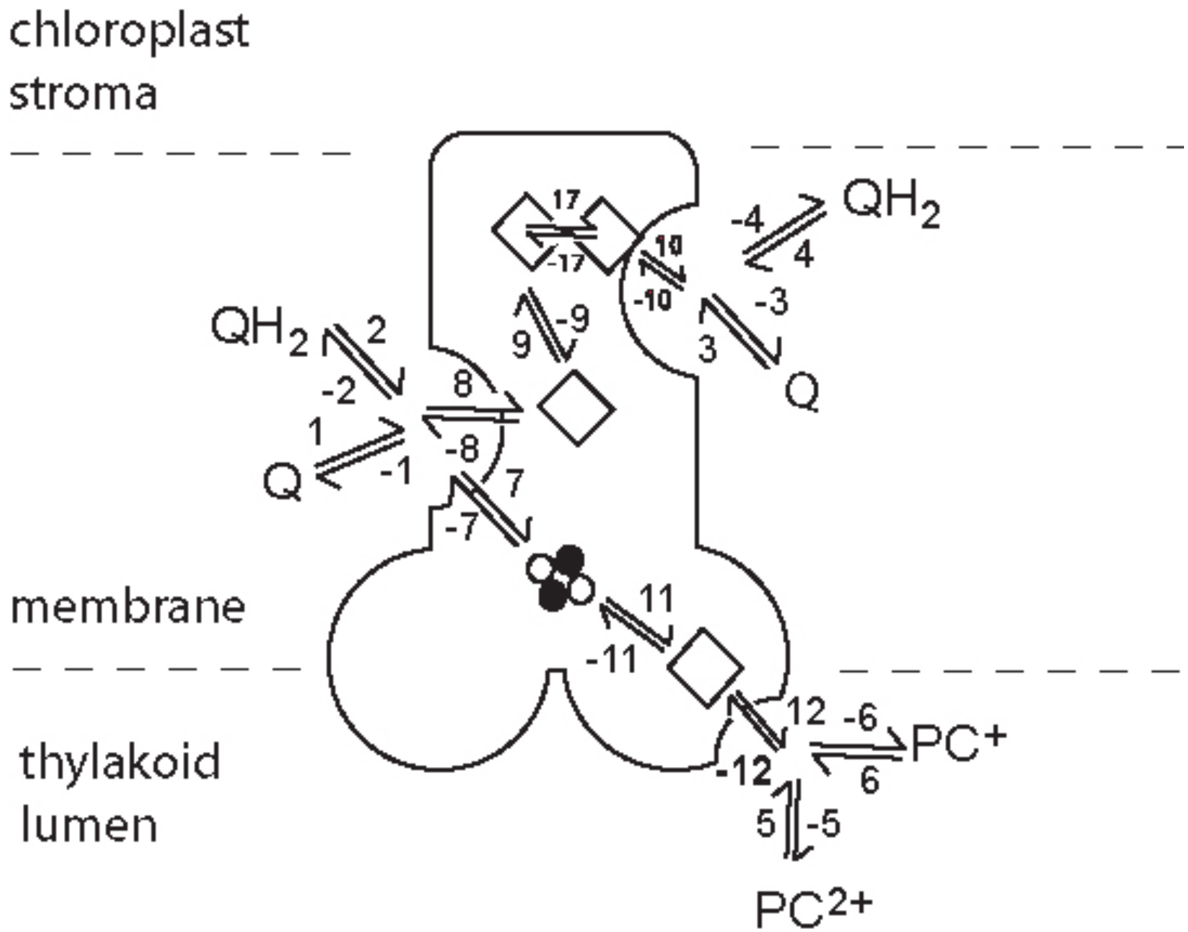


Simulation of a simplified Q-cycle

Model of cyt b6f:

3 substrate binding sites

5 internal redox sites



Gillespie's Direct Method

Suppose M different events could occur at time t_0. Let $a_i \; \Delta t$ be the probability that event i will occur during a short time interval Δt. Let $a = \sum a_i$. Let r_1 and r_2 be random numbers uniformly distributed on the unit interval. Then the random time until the next event is

$$\tau = \frac{1}{a} \ln \frac{1}{r_1}.$$

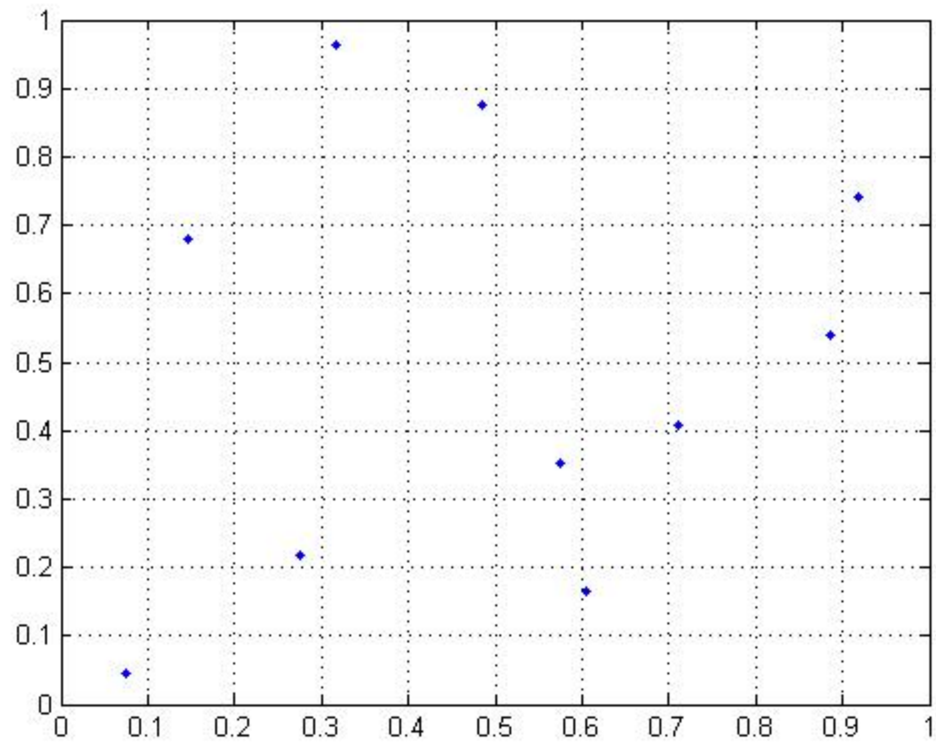
τ is exponentially distributed with density: $a \exp(-a\tau)$. The event j is chosen that satisfies

$$\sum_{i=1}^{j-1} a_i \leq a\, r_2 < \sum_{i=1}^{j} a_i.$$

Event j is implemented. The process is then repeated.

Latin hypercube sampling

- Partition *n* parameter intervals into *m* subintervals each.
- Choose a random number in each subinterval
- Randomly pair subintervals for different parameters.
- Example for n=2 and m=10



Use of Latin hypercube sampling to find parameter sets that fit the data

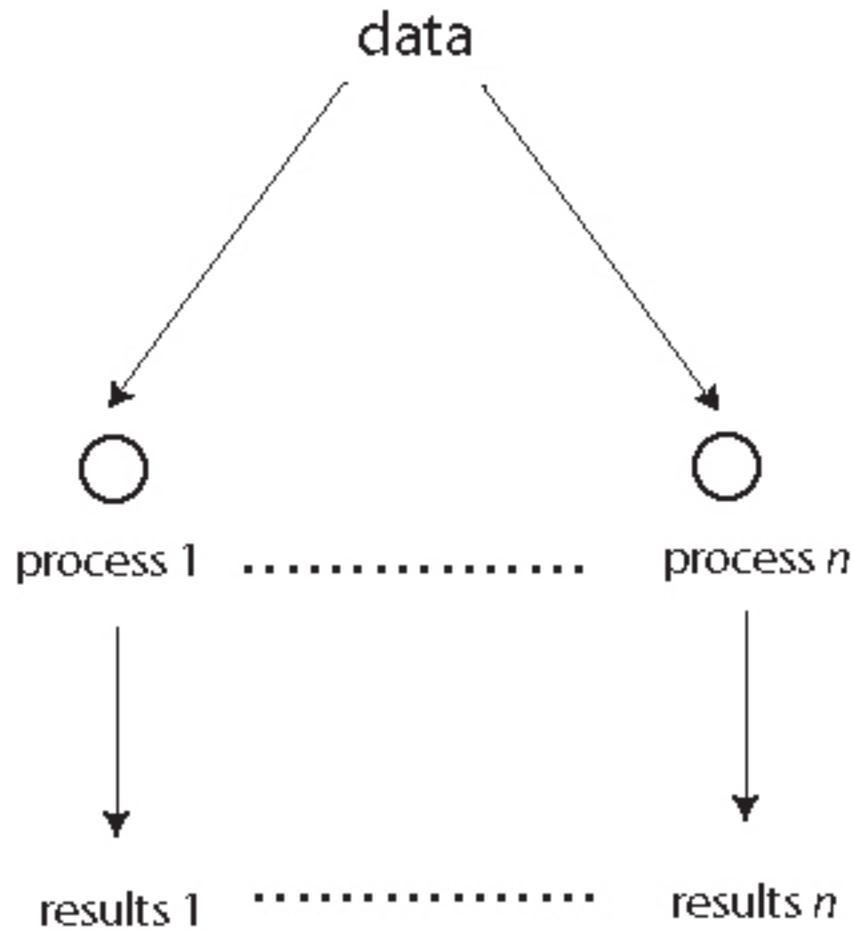
- Sample points are chosen from hyper-cubes in 15 or 22 dimensional parameter space.
- Each dimension is partitioned into m subintervals. A random no. is chosen in each. These are randomly associated to generate *m* points in parameter space. We used $m=256$ and $m=8192$.
- A Monte-Carlo simulation is performed for each sample point and a root mean square deviation (χ^2) between the simulation and experimental data is computed. The sample point corresponding to the lowest χ^2 value is chosen as giving the best fit.

Parallel processing on a Beowulf cluster

Results *i* depend only on data, process rank *r*, and no. of processes *N*.

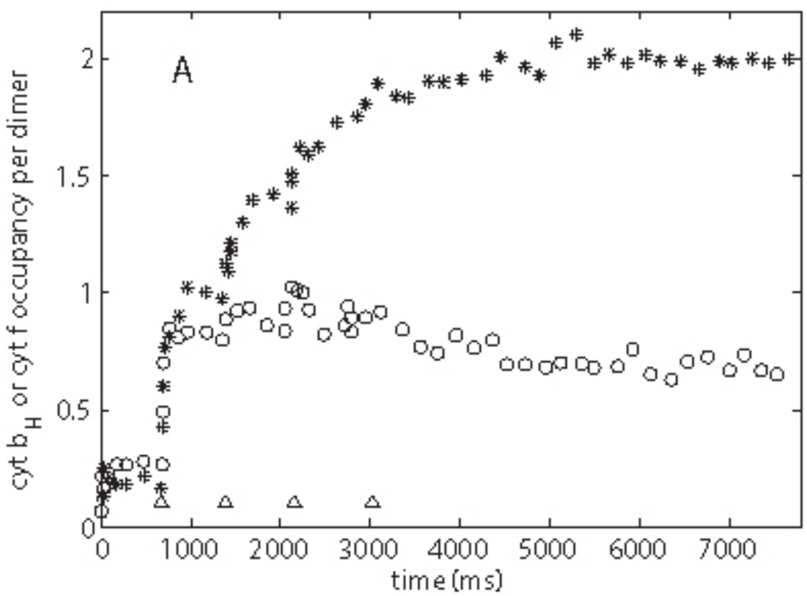
Sets of 64 simulations are averaged together to give a trajectory.

Each processor runs a sequence of sets of simulations.

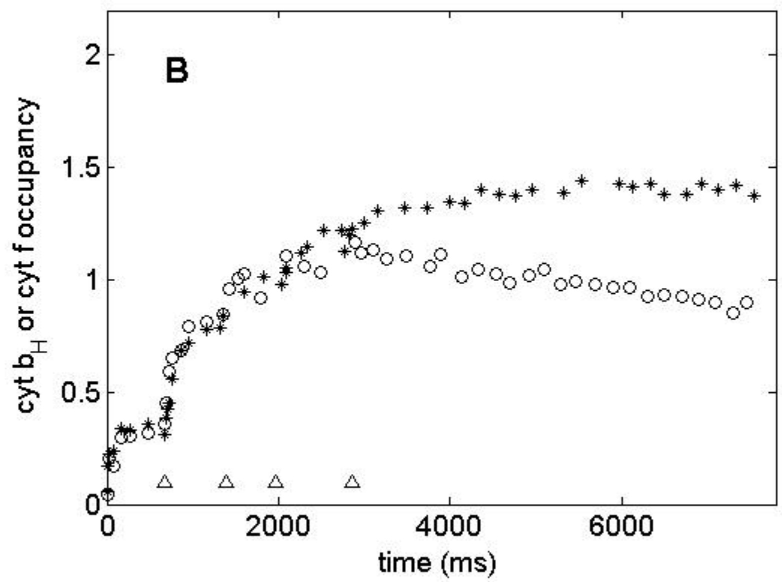


Data of Kramer and Crofts from spinach (BBA, 1993)

A data: without Inhibitor



B data: with Inhibitor

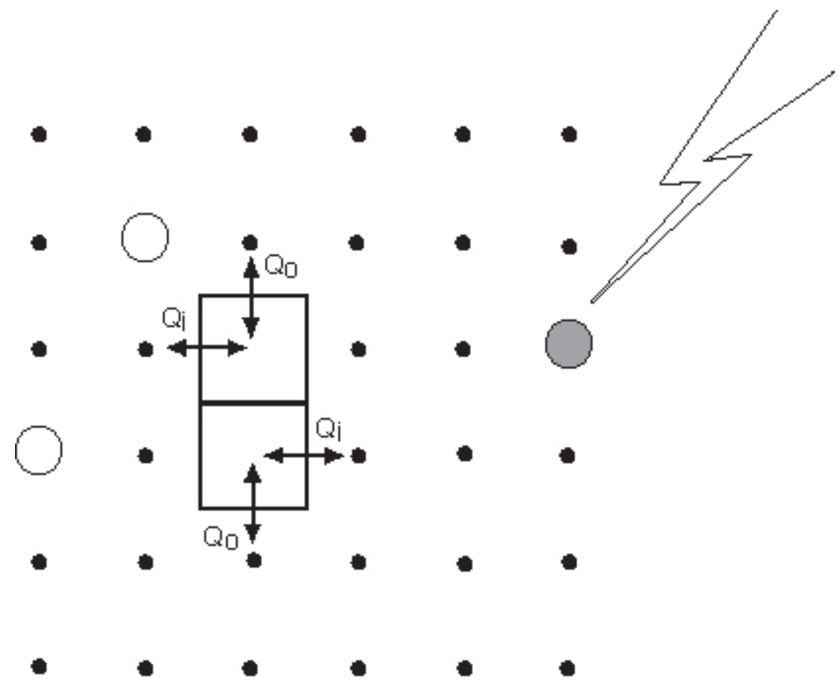


Initial and Auxiliary Conditions

All electron carriers (quinones) and cyt b6f redox centers are initially oxidized.

Experimentally, photosystem II releases a quinol at every other flash of light.

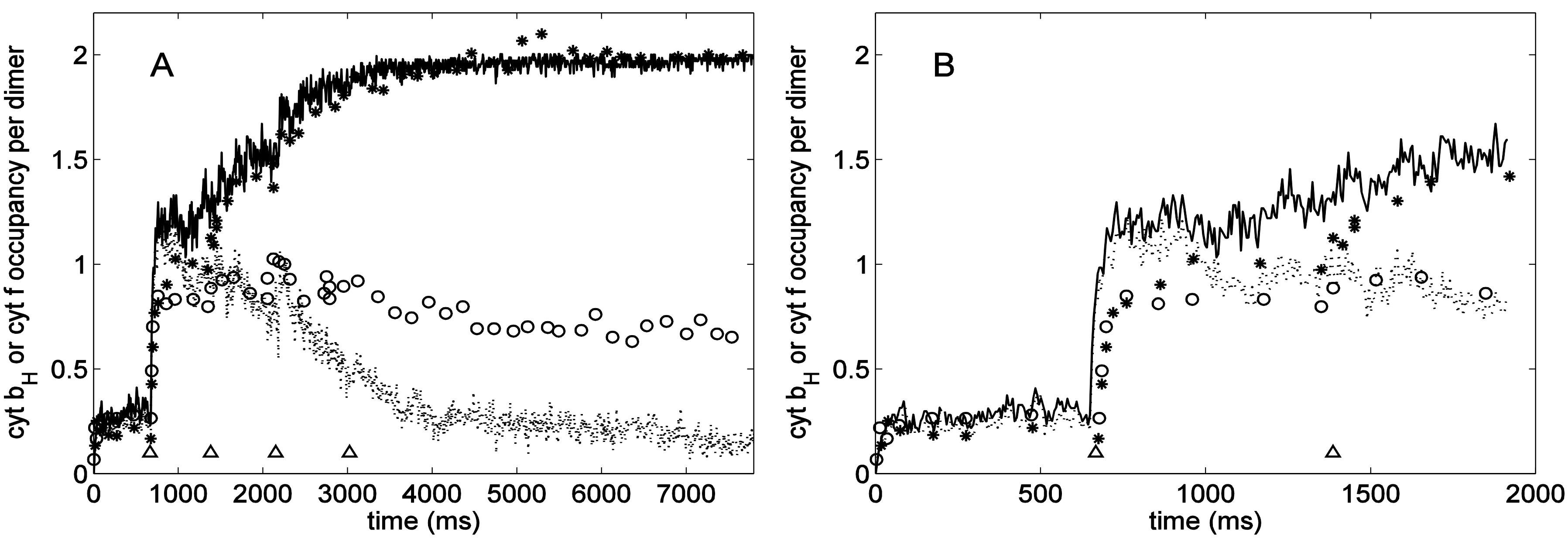
In the simulation, at every second 'flash' (lightning bolt), a quinone is chosen at random and changed to a quinol.



Fit to A data with base rate set
(15 adjustable parameters)

0 < t < 7.8 seconds

0<t<2 seconds



Rate exponents and hypercube

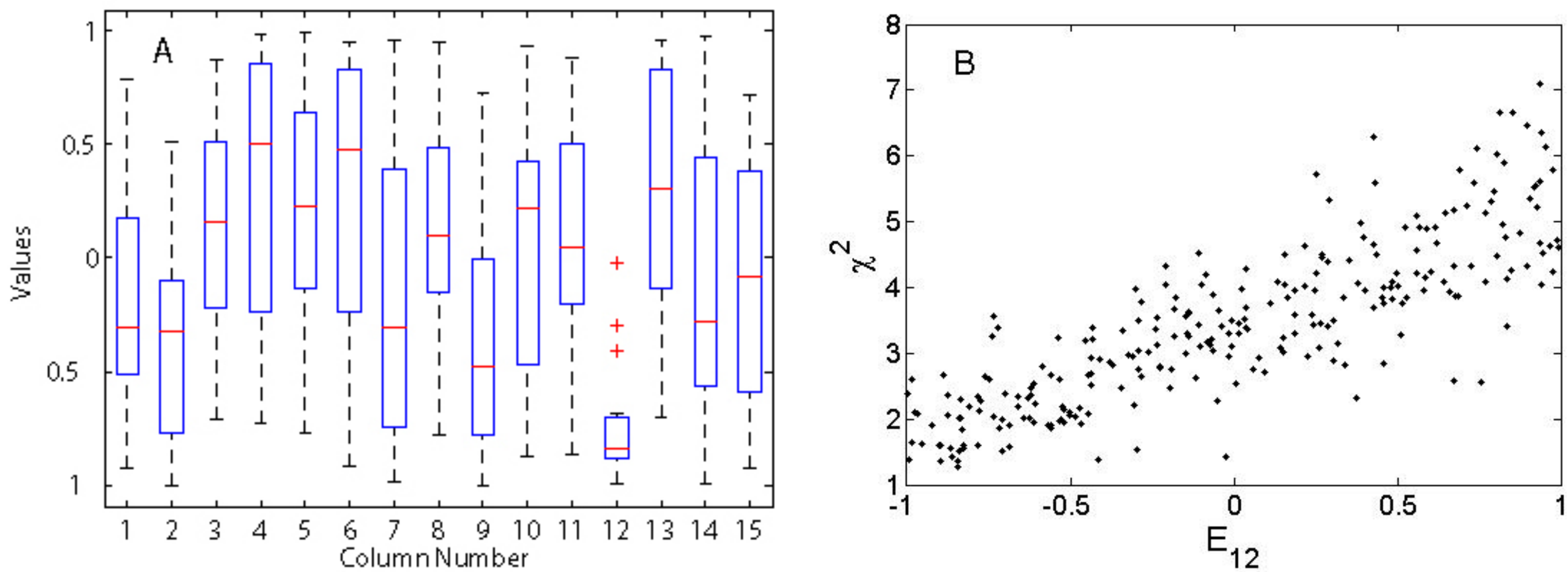
Rate exponents $E_i, i = 1 \ldots 15$ are defined by

$$R_i = B_i 2^{E_i}$$

where B_i is the base value for the i^{th} rate and R_i is a Latin hypercube sample value of the i^{th} rate.

256 samples are generated in the hypercube $[-1\ 1]^{15}$.
Where are the 16 samples that give best fits with the data?

Sensitivity Analysis



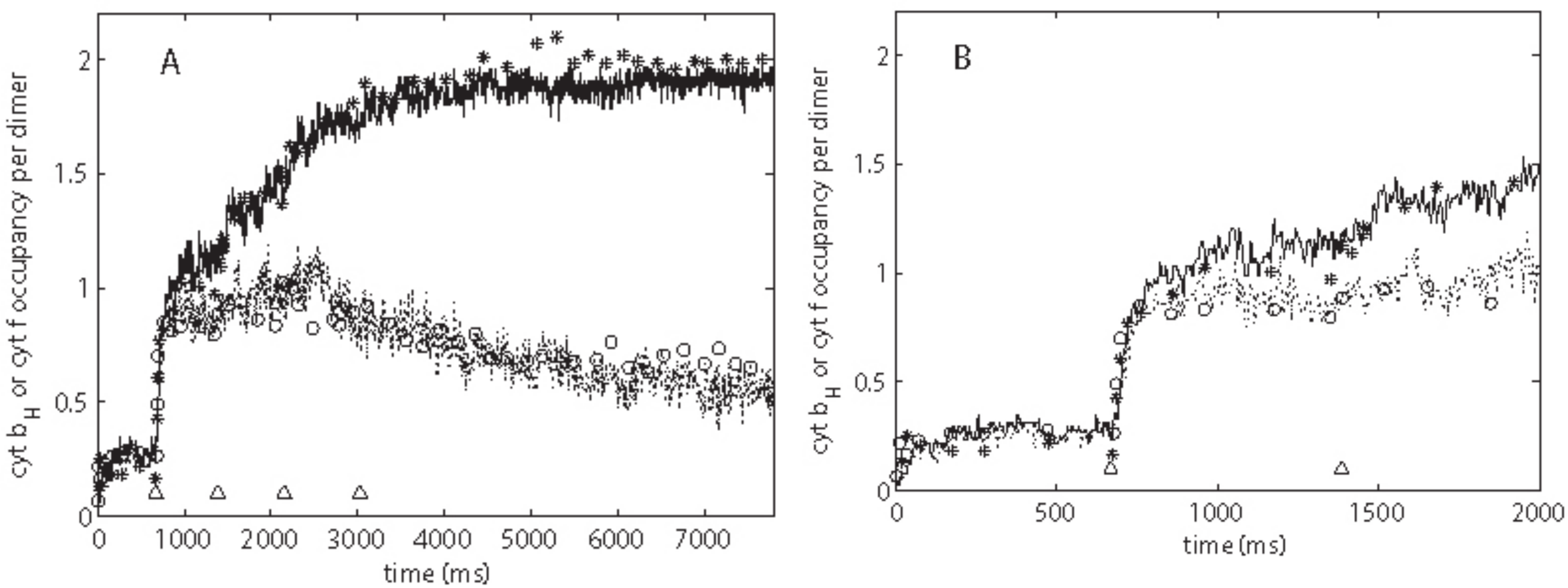
(A) Quartiles of rate exponent E_i as a function of column number i for 16 best fits.

(B) Scatterplot shows distribution of rate exponent E_{12} for all 256 Latin hypercube samples.

Best Latin Hypercube Sample Fit to A Data

$0 \leq t \leq 7.8$ seconds

$0 \leq t \leq 2$ seconds



15 adjustable parameters

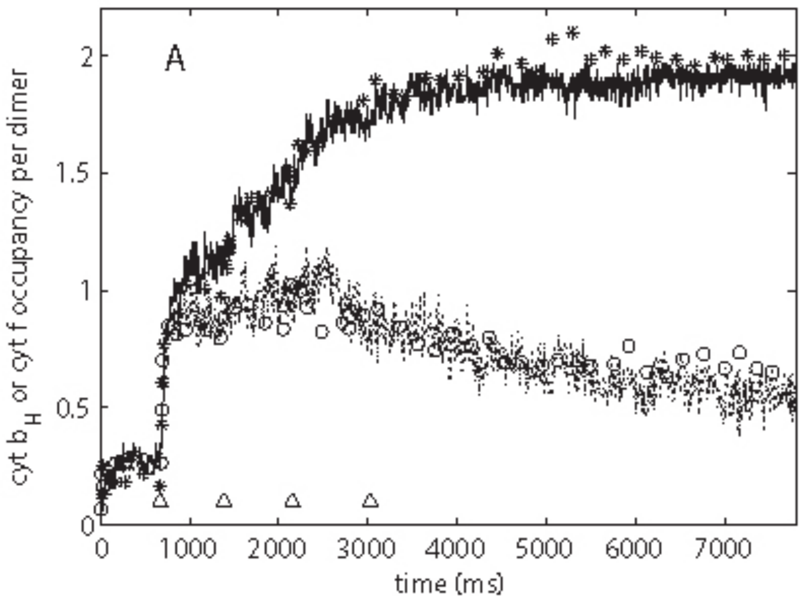
“Give me four parameters and I can fit an elephant, give me five and I can wriggle its trunk”

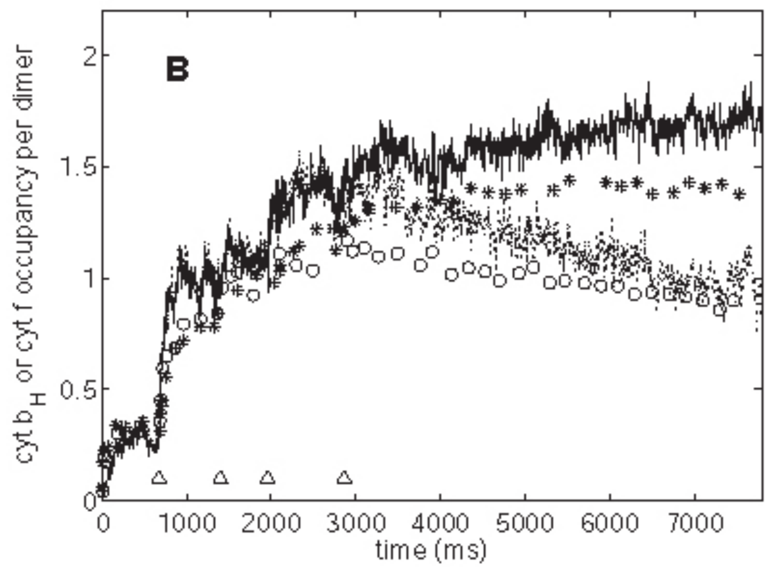
John von Neumann

One parameter fit to the B data

A data: no inhibitor

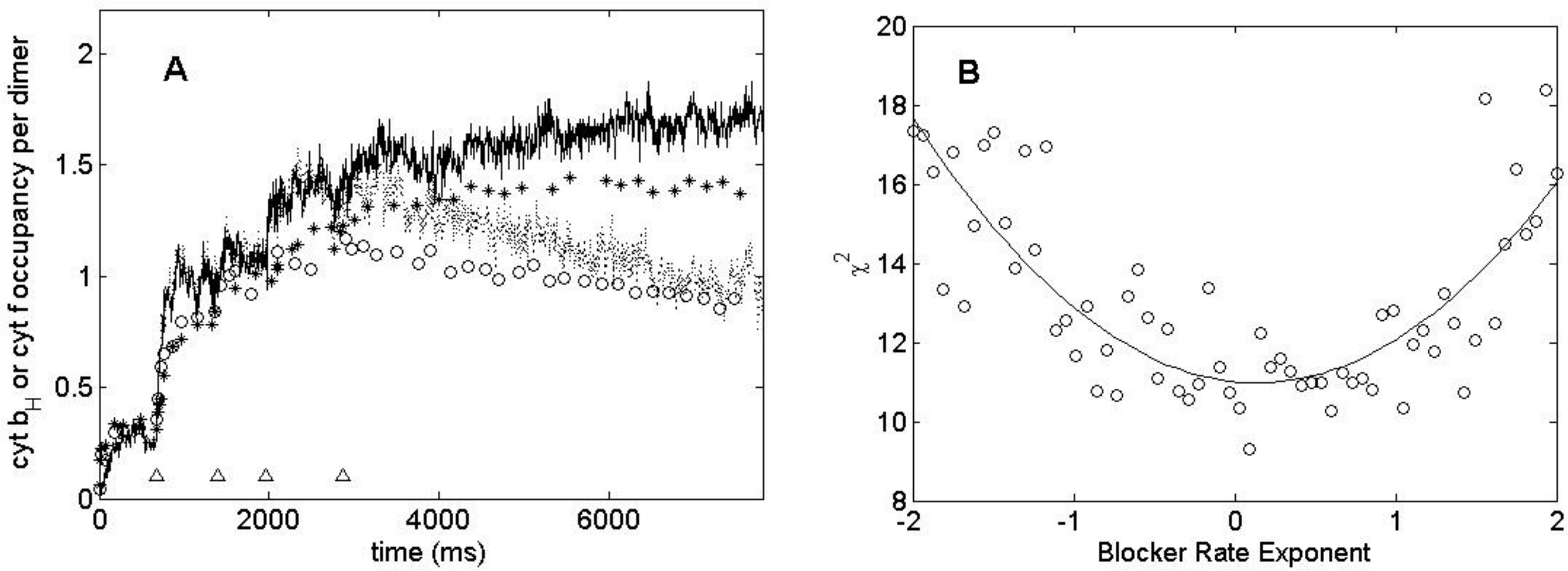
B data: with inhibitor





(A) Fit to A data for $0<t<7.8$ seconds. (B) Fit to B data obtained by starting from fit to A data and optimizing a single reductase site inhibitor release rate.

X^2 is a noisy function of transition rates



(A) One parameter fit to B data.

(B) χ^2 values plotted as a function of inhibitor release rate exponent.

Conclusions

Our Q-cycle model successfully interpolated the fairly complex experimental data sets of Kramer and Crofts.

Using the parameters obtained by the fit to the A data, we were able to (in a weak sense) predict the B data.

Ultimately, a model like this could be strongly validated by predicting the results of further experiments.

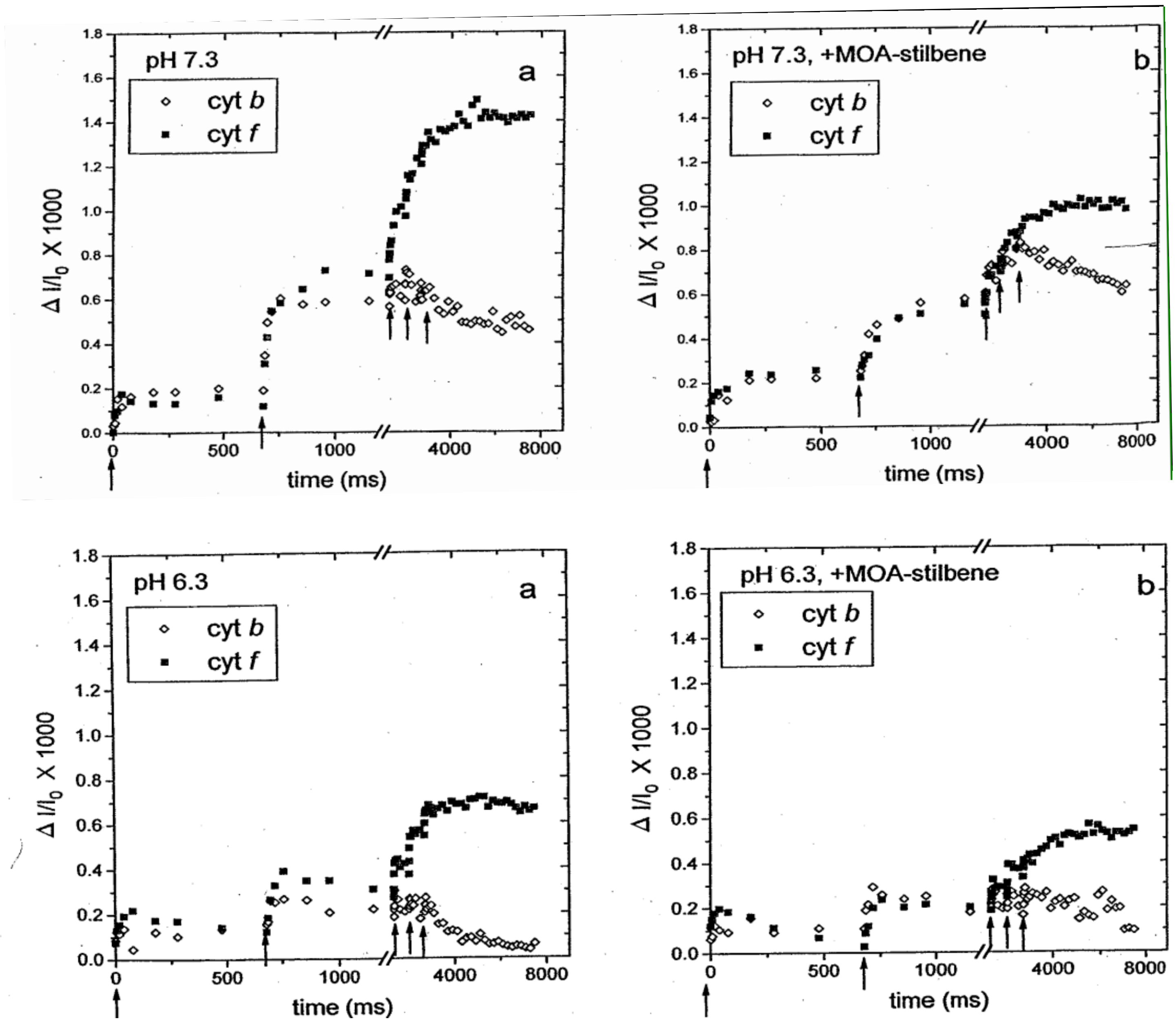
Proposed pH project

Kramer and Crofts (1993) also studied the dependence of the dynamics of their system on pH.

Experiments at
pH 7.3 and
pH 6.3

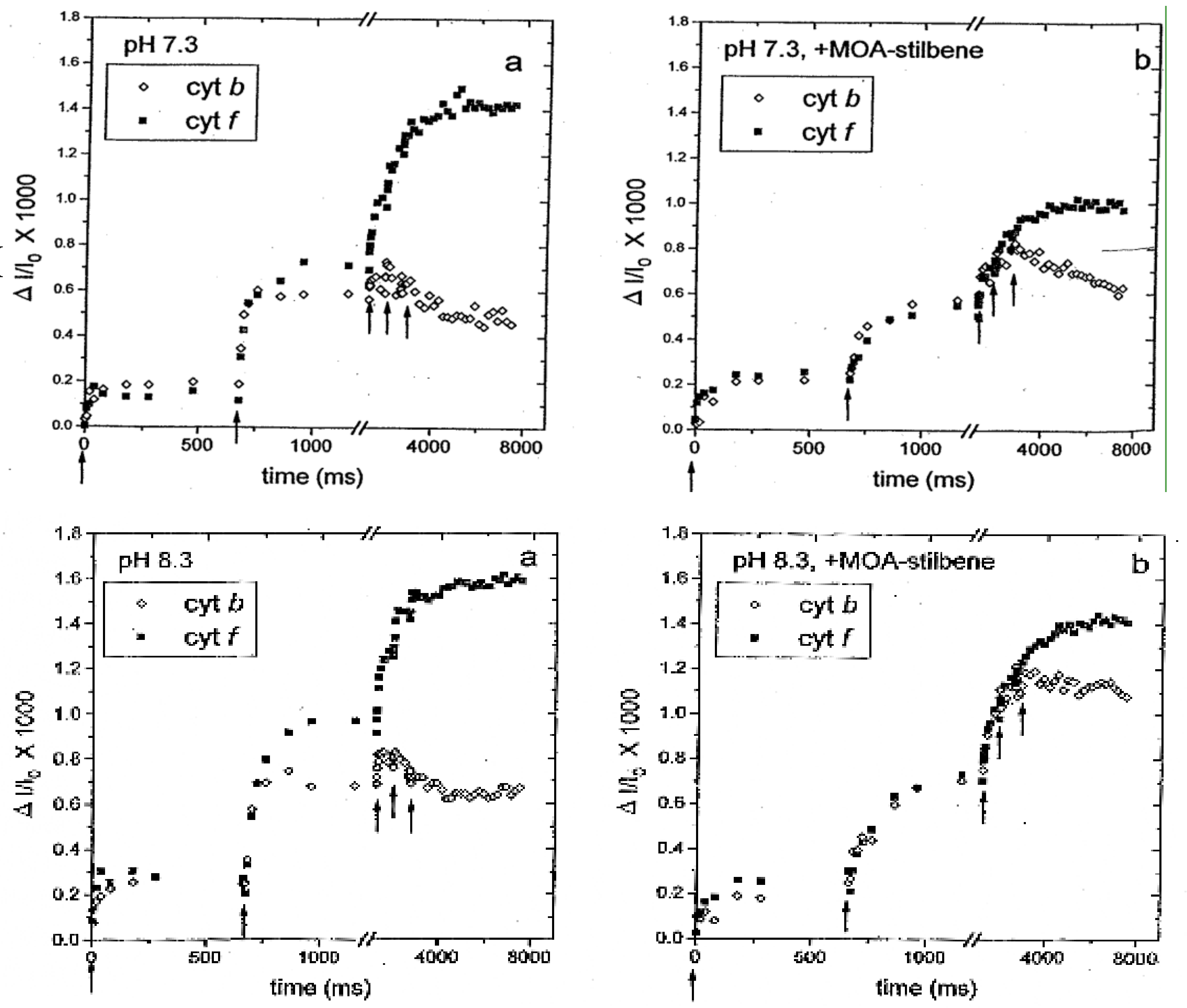
Kramer and Crofts, 1993

Mass action

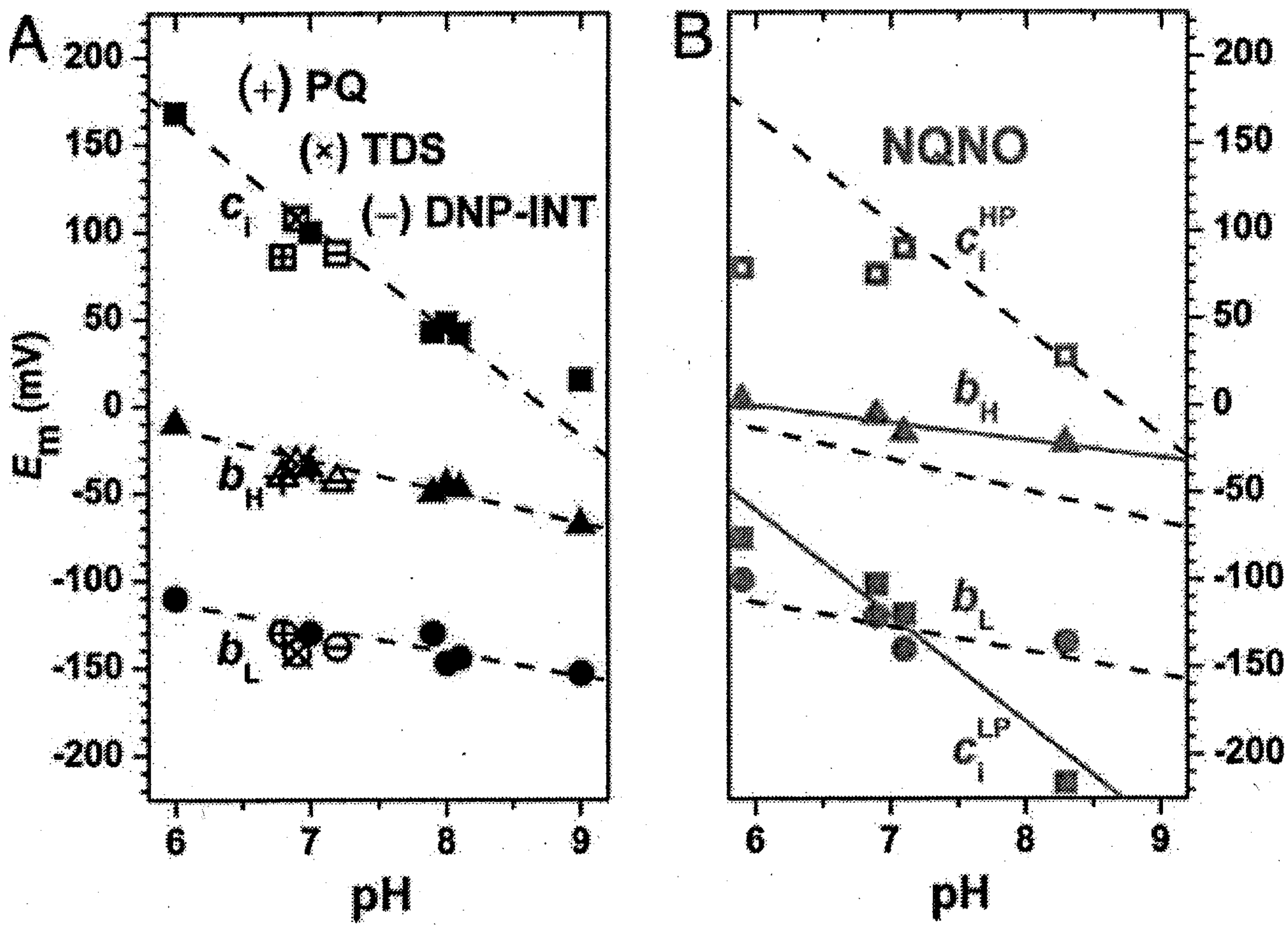


Experiments at pH 7.3 and pH 8.3

Kramer and Crofts, 1993



pH dependence of midpoint potentials (Alric et al. 2005)



Goals

- Use published measurements of the pH dependence of cyt b6f redox site midpoint potentials and the pH dependence of the oxidation of quinol to build simulations that are sensitive to pH.
- Compare with the results of Kramer and Crofts.
- Does one proton or do two protons come off of quinol when the first electron is donated to the FeS cluster?

David Kramer may collaborate with us.

What would a research experience involve?

Learning how to run the simulation program and then doing so for a variety of scenarios.

Other possibilities depending on interests and skills.

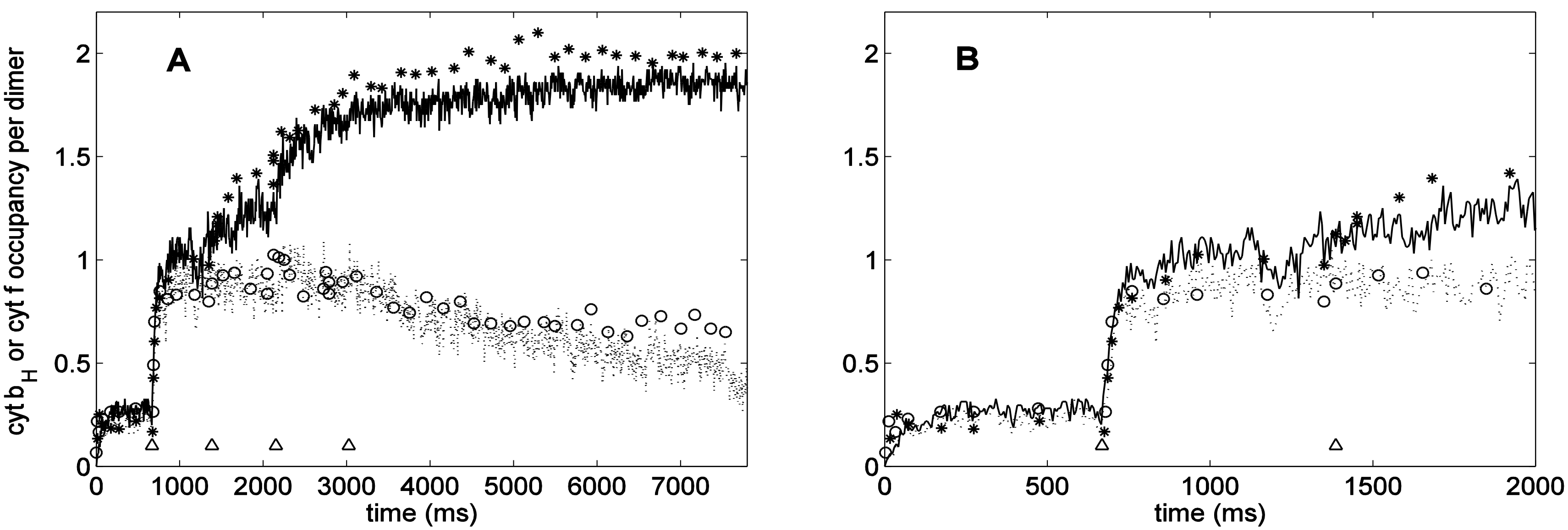
Thank you!

Appendix

Best Latin hypercube sample fit to combined A and B data.

A data: no inhibitor
$0 \leq t \leq 7.8$ seconds

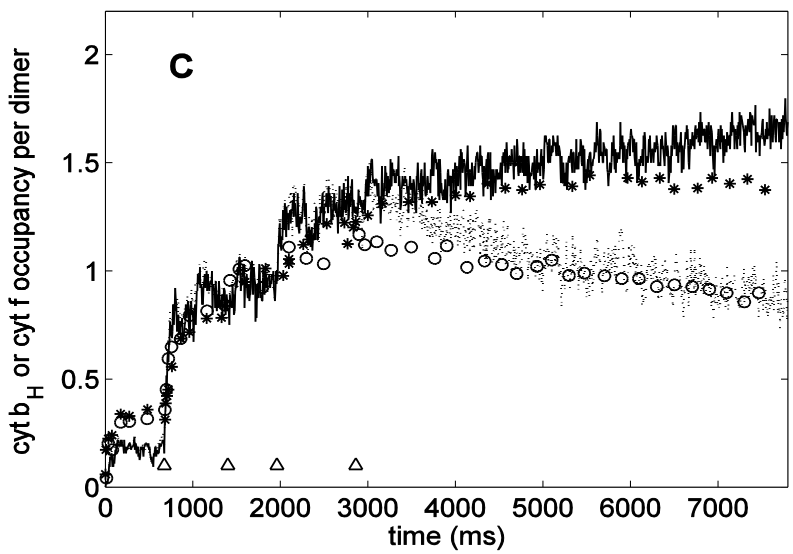
A data:
$0 \leq t \leq 2$ seconds



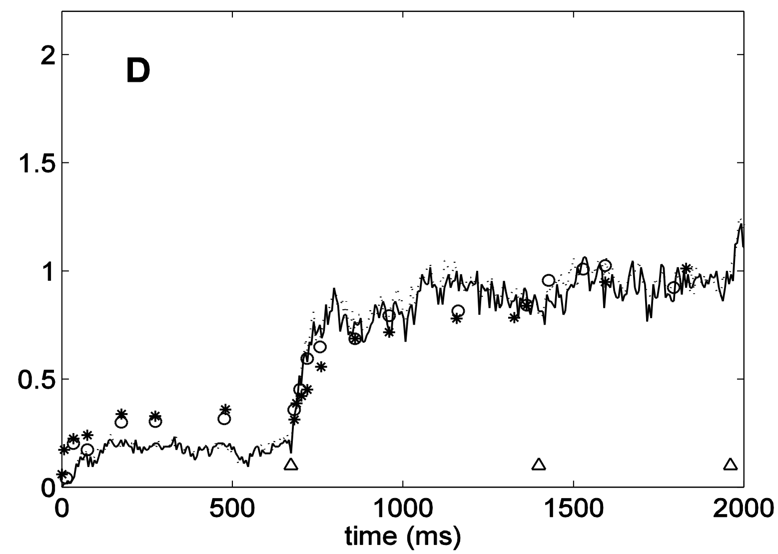
22 adjustable parameters

Best Latin hypercube sample fit to combined A and B data.

B data: initial inhibitor
$0 \le t \le 7.8$ seconds



B data:
$0 \le t \le 2$ seconds



22 adjustable parameters (same as previous slide)

Literature and optimized values of midpoint potentials

Redox site	Lit. value (mV)	Fit to A and B (mV)
Cyt b_H	-50.	-53.
Cyt b_l	-130.	-138.
x_a	100.	83.
x_b	-125.	-132.
Cyt f	350.	337.
R	300.	287.

Implications of fits for Q-cycle model

The Q-cycle model successfully explains the fairly complex experimental data set of Kramer and Crofts.

•A very nice 15 parameter fit to the A data was obtained from an initial base rate set.

•The model fit to the A data successfully 'predicted' a one parameter fit to the B data.

•A good 22 parameter fit to the combined A and B data was obtained, and six of the additional parameters are modest adjustments of the literature values.

Ultimately, a model like this could be strongly validated by predicting the results of further experiments.

Final remarks

Latin hypercube sampling worked well as a method of approximately optimizing very noisy functions of 15 or 22 variables to fit a somewhat complicated set of experimental data. This method should be further investigated.

Monte Carlo simulations are a more powerful method of modeling certain kinds of systems than differential equations. However, the simulations are inherently stochastic. Latin hypercube sampling and parallel computations can provide a practical way to optimize their parameters to fit fairly complicated data sets.

Acknowledgments

Lou Gross suggested Latin hypercube sampling.

Ed Pate made the Beowulf cluster available.

Kevin Cooper advised me on programming the cluster.

Chandana Somayaji and Talitha Anderson were undergraduates who worked on an early version of the qcycle program.