Comparison of Monte Carlo simulations of cytochrome b_6f with experiment using Latin hypercube sampling

Mark F. Schumaker[1] and David M. Kramer[2]

[1]Department of Mathematics and [2]Institute for Biological Chemistry

Washington State University

Pullman, WA 99164 USA

We have programmed a Monte Carlo simulation of the Q cycle model of electron transport in cytochrome b_6f complex, an enzyme in the photosynthetic pathway that converts sunlight into biologically useful forms of chemical energy. Results were compared with published experiments of Kramer and Crofts (1993). Rates for the simulation were optimized by constructing large numbers of parameter sets using Latin hypercube sampling and selecting those that gave the minimum mean square deviation from experiment. Multiple copies of the simulation program were run in parallel on a Beowulf cluster. We found that Latin hypercube sampling works well as a method for approximately optimizing very noisy objective functions of 15 or 22 variables. Further, the simplified Q cycle model can reproduce experimental results in the presence or absence of a quinone reductase (Q_i) site inhibitor without invoking ad hoc side-reactions.

Key words: Cytochrome b_6f, Latin Hypercube Sampling, Monte Carlo, Photosynthesis, Q cycle

1. Introduction

Inside chloroplasts of higher plants, thylakoid membranes form closed surfaces. They enclose an aqueous thylakoid lumen that has the geometry of a thin sheet. Only a few nanometers separate opposing surfaces of the membrane on either side of the lumen. A single thylakoid membrane may be very elaborately extended within the chloroplast. In some regions sheets may be stacked on top of one another like pancakes and in other regions unstacked. These different regions have different functions in photosynthesis. Dekker and Boekema (2005) review the organization of these membranes. The thylakoid membrane is crowded with proteins. Estimates (Kirchhoff, 2008) suggest that 70-80% of the membrane area consists of proteins that are separated by narrow regions of lipid. Most of these protein molecules participate in the *light reactions of photosynthesis*, transducing sunlight into chemical energy in the form of ATP and other molecules than can be used by different components of plant metabolism.

Fig. 1 shows the major pathway of energy transduction, which is reviewed in, e.g. Ort and Kramer (2009). Sunlight is absorbed by antenna chlorophylls associated with photosystem II, which uses this energy to split water into oxygen, hydrogen ions (protons, H^+), and electrons. For each four photons absorbed, two water molecules are oxidized, forming one molecule of O_2, 4 H^+ and 4 electrons (e^-). The electrons excited in photosystem II are passed from the luminal side to the stromal side of the thylakoid membrane and reduce two plastoquinone (PQ), a small organic molecule that can accept pairs of electrons and protons, to form two plastoquinol (PQH_2). PQH_2 can then diffuse to the quinol oxidase site (Q_o) of cytochrome b_6f

(cyt b_6f), where it is oxidized back to PQ in a process called the 'Q-cycle', described in detail below. The electrons on PQH_2 are (eventually) transferred to plastocyanin (PC). Reduced PC then diffuses through the thylakoid lumen and reduces oxidized P_{700}, the primary chlorophyll redox center on photosystem I. After excitation by light at photosystem I, the electron is transferred across the thylakoid membrane to ferredoxin, and NADPH. The light-driven transfer of electrons through this electron transfer chain is coupled to the translocation of protons from the stroma into the lumen. establishing an electrochemical gradient of protons, termed the proton motive force (*pmf*). The *pmf* drives synthesis of ATP at the chloroplast ATP synthase. Energy stored in NADPH and ATP is used to power the assimilation of CO_2 and other cellular processes.

The coupling of proton translocation to electron transfer is critical for photosynthesis. One proton per electron is released into the lumen during oxidation of water in the oxygen evolving complex of photosystem II. Another proton per electron is taken up from the stroma during reduction of PQ at photosystem II and released into the lumen upon oxidation of PQH_2 at the cyt b_6f complex. Finally, an additional one proton per electron is translocated at the cyt b_6f complex via the Q-cycle, as described below. The Q-cycle is catalyzed by the cyt b_6f complex, which is composed of from 8-11 protein subunits (Smith et al., 2004). Its structure has been solved by X-ray crystallography. Cyt b_6f molecules are dimers with a cross section parallel to the plane of the membrane, and in the membrane, of about $9\text{nm} \times 5\text{nm}$ (Kurisu et al., 2003). The dimer comprises two identical monomers, and each monomer has 3 substrate binding sites and 5 electron transfer cofactors, as illustrated in Fig. 2. There are two distinct PQ/PQH_2 binding sites. The quinol oxidase (Q_o) site functions during normal activity to oxidize

PQH_2. Although the site is large enough to accommodate two functionally-distinct binding 'niches' there is little information about the interactions of substrates or intermediates within these niches (reviewed in Cooley et al., 2008). Therefore, our model treats this site as a single, mono-functional binding site but future modifications can accommodate more complex scenarios. The Q_i binding site normally functions to reduce PQ to PQH_2, and contains the special bound heme, c_i.

The scheme of electron transport in cyt b_6f and the related cyt bc_1 complexes is known as the *Q-cycle* (reviewed in Cape et al., 2006). In the cyt b_6f complex, QH_2 binds to Q_o and is oxidized by both the high and low potential chains in the so-called 'bifurcated reaction' or 'oxidant-induced reduction' that allows the Q-cycle to act as a proton 'pump'. First, PQH_2 at Q_o is oxidized in a one-electron process by the "Rieske" iron-sulfur cluster, resulting in the release of two protons, which are eventually ejected into the lumen, and the formation of a highly reactive semiquinone anion intermediate (reviewed in Kallas et al. 1994; Cape et al., 2007; Cooley et al. 2008; Smith et al. 2004). The iron-sulfur cluster is then oxidized by cyt f, which in turn reduces plastocyanin, which can transfer the electron to photosystem I. In the normal Q-cycle, the semiquinone intermediate formed at the Q_o site reduces oxidized cyt b_L, which in turn reduces cyt b_H and heme c_i. This electron remains on the low potential chain until a second turnover of the Q_o site results in the accumulation of two reduced acceptors. PQ bound at the Q_i site is then reduced by two electrons from the low potential chain, probably via heme c_i, with the uptake of two protons from the chloroplast stroma.

In the Q-cycle, two protons are translocated across the membrane for every electron that passes through the high potential chain. One of these protons is taken up at the Q_B site of PSII, the other at the Q_i site of cyt b_6f ; both protons are released at the Q_o site. The Q-cycle is widely accepted as the basic plan of electron transport in cyt b_6f and bc_1 complexes, but other "bypass" reactions are energetically favored. An important function of the enzyme is to reduce the rate of these unwanted reactions (Muller et al., 2002).

Mathematically, transport of electrons along the high and low potential chains of cytochrome b_6f has the form of coupled 1D exclusion processes with particle baths at either end. Single 1D exclusion processes with somewhat similar boundary conditions may display phase transitions and traveling wave behavior. See, for example, Blythe and Evans (2007) or Derrida (2007). In the Q-cycle, the free energy difference of electrons across the high potential chain also drives transport of electrons across the low potential chain.

Kramer and Crofts (1993) performed experiments on spinach leaf pretreated to remove plastocyanin and oxidize the electron accepting sites of cyt b_6f. Their results are shown in Fig. 3. Electron occupancy of cyt b_H or cyt f per dimer is plotted on the ordinate. Since there are two copies of each electron accepting site per dimer, values on the ordinate should not, in principle, exceed two. However, the experimental signal includes significant noise. Starting at time 0, five pulses of light from a xenon flashlamp were applied to excite photosystem II, causing these enzymes to split water and donate electrons to quinols that were released into the thylakoid membrane. The last 4 of these pulses are indicated by triangles for times greater than 0. The oxidation state of cyt f (asterisks) and cyt b_H (open circles) was deduced by a change in absorbance at specific probe wavelengths. A maximum occupation of 2 on the

ordinate corresponds to electron transfer to (a reduction of) both cyt f or cyt b_H groups on a dimer.

Two sets of experiments were performed. Those shown in Fig. 3B used the same conditions as for Fig. 3A except that the quinol reductase site was initially occupied by a molecule that blocked substrate binding. This inhibitor molecule is slowly released through the course of the experiment, allowing slow turnover of the Q_i site. The two sets of data shown in Fig. 3 will be referred to simply as the A and B data sets below.

Latin hypercube sampling (LHS) was introduced by McKay et al. (1979). Consider a model with n parameters, each defined on an interval of values. LHS partitions each of these *n* intervals into *m* subintervals of equal probability. A single random number x_{ij} is chosen in each subinterval. One such random number is chosen for each parameter (without replacement) to form n-tuples $\boldsymbol{x}_j = \{x_{1j}, x_{2j}, \dots, x_{nj}\}$ representing points in n dimensional space. The m points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ constructed in this way constitute the Latin hypercube sample. Fig. 4 gives an example with $n = 2$ parameter intervals $[0,1]$ partitioned into $m = 10$ subintervals each. A single random number is chosen from a uniform distribution on each subinterval. These are randomly paired to give a Latin hypercube sample of 10 points on the unit square. Helton and Davis (2003) and Helton et al. (2006) give more detailed and general discussions of the Latin hypercube sampling procedure.

LHS has been used extensively in uncertainty and sensitivity analysis. These methods analyze how results $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ are sensitive to variations in inputs $\boldsymbol{x}$. To study the sensitivity, it is better to generate a sample of inputs in a neighborhood of a base value by small variations of

all of the parameters than to vary one parameter at a time (Saltelli et al., 2006). Helton and Davis (2003) gives good insight into the statistical properties of LHS that are advantageous for uncertainty and sensitivity analysis. Helton et al. (2006) compare several sampling methods for sensitivity analysis in their discussion of an example with a 31 dimensional input space. LHS is efficient, at least in part, because it uses a sample from each subinterval of each parameter only once. Below, we use LHS as a method for optimizing the value of a very noisy scalar objective function $f(\boldsymbol{x})$ by sampling values of $\boldsymbol{x}$ in 15 and 22 dimensional spaces. This noisy function gives values of the mean square deviation, denoted χ^2, computed from sets of experimental data and a Monte Carlo simulation that depends on a parameter set $\boldsymbol{x}$.

In the mathematical literature, the word *model* often refers to the function $f(\boldsymbol{x})$. In the biological literature, *model* may refer to a simplified representation of an object or a process that is not mathematical. We often use *model* in this biological sense. Our *mathematical models* are Monte Carlo simulations of a simplified mechanism for electron transport in cytochrome b6f. The simulations give results χ^2 as a function of model parameters (one parameter set is given in Table 1).

In the next section we describe the Monte Carlo simulation of quinone, quinol and cyt b_6f in the thylakoid membrane. Our simulation neglects other proteins in the thylakoid membrane but gives a relatively detailed representation of the Q-cycle. By comparison, Tremmel et al. (2003, 2007) have simulated the protein component of the membrane in detail. Tremmel et al. (2007) included in their simulations a simple model of the Q-cycle and compared their results with experiment in order to determine whether electron transfer from quinol to

the iron-sulfur cluster of cyt b_6f is limited by diffusion of PQH_2 to the Q_o site or by the rate of oxidation of PQH_2 bound at the Q_o site. They used a genetic algorithm to optimize their model parameters to fit the data. We use Latin hypercube sampling to investigate whether this approach is suitable for optimizing parameters of our model of the Q-cycle, and to determine whether the 'standard' Q-cycle model can account for the data of Kramer and Crofts (1993) or if additional modifications or complexities are required. Of particular interest are the kinetics of reduction and re-oxidation of cyt b.

2 Methods

2.1 Model of cytochrome b_6f in the thylakoid membrane

The representation of the thylakoid membrane in our simulation program, qcycle, has a geometry similar to that of a toaster pastry (a strudel with a very thin filling). The simulated thylakoid lumen, corresponding to the filling, is a rectangle. The simulated thylakoid membrane, corresponding to the pastry shell, is wrapped around the lumen. The program qcycle approximates diffusion of plastocyanin in the thylakoid lumen and diffusion of quinone and quinol in the membrane by random walks on square lattices in the simulated domains. Fig. 5 depicts the geometry in the case that the lumen is represented by a 6×6 square lattice. Arrows with numbers indicate how boundary conditions are implemented to model the toaster pastry geometry of the membrane. If the top half of the membrane were a 6×6 lattice as shown, one may imagine the 6×6 lumen just below, and the bottom half of the membrane below that. Corresponding sites on the three layers would be aligned vertically.

Fig. 6 depicts the model of cyt b_6f in the thylakoid membrane. Monomers comprising a dimer are associated with adjacent sites on the membrane lattice. The lattice constant, the distance between adjacent sites on the lattice, is set to 5nm so that the area associated with a pair of lattice sites approximates the membrane area of a b_6f dimer. Transition rates of the quinone and quinol random walks are proportional to their respective diffusion coefficients. Binding to the quinol oxidase sites is modeled by approaches of quinol or quinone from the random walk sites above or below the dimer. Binding to the quinone reductase sites is modeled by approaches from random walk sites to the left or right of the dimer as shown. The plastocyanin binding site of each monomer is modeled by an approach from the lumen site with which it is aligned.

Fig. 2 shows the redox transitions represented by the model of cyt b_6f. There are 5 internal electron accepting sites that are either reduced or oxidized, giving 32 possible internal states. In addition, there are 3 substrate binding sites. The plastocyanin site (lower right) may be empty (no plastocyanin bound) , occupied by reduced plastocyanin, or occupied by oxidized plastocyanin. The quinol oxidase site (Q_o site; at the left) may be empty, occupied by PQH_2 (reduced with two electrons and two protons, i.e. with two hydrogen atoms), by plasto-semiquinone (plastoquinone with one unpaired electron), or by PQ (the most oxidized physiological form). The quinol reductase site (Q_i site; at the upper-right) may be blocked by an inhibitor such as MOA-stilbene, empty, or occupied by PQH_2, plastosemiquinone or PQ. Thus there are $3 \times 4 \times 5 = 60$ possible binding states. Even for our simplified model, there are $32 \times 60 = 1920$ possible cyt b_6f states.

It has been proposed that the binding affinity of substrate PQ and inhibitors to the Q_i site are modulated by the redox state of heme c_i, (Alric et al., 2005). To determine if this proposed feature can explain the observed kinetics for cyt b redox changes, this behavior was treated in our model by letting the off-rates (or unbinding rates) for substrates or inhibitors depend on the heme c_i redox state.

In the general case of our Q cycle model, transitions between the possible b_6f states are governed by 35 distinct rate constants (or intrinsic transition probabilities). However, plastocyanin was removed in the experimental preparations of Kramer and Crofts (1993), making transitions $\pm 5, \pm 6$ and ± 12 inaccessible (transition numbers refer to Fig. 2). The remaining 29 rates are described in Table 1. Most of these rates depend only on the adjacent redox sites between which the transitions are made, however 5 sets of rates are more complicated. The exit (unbinding or off-) rates of quinone and quinol from the quinol reductase site, transitions -3 and -4, depend on whether heme c_i is oxidized or reduced, consistent with the 'redox switch' mechanism proposed by de Lacroix de Lavalette et al.(2009). The rate of transition 10 depends on whether a PQ or plastosemiquinone occupies Q_i before the transition is made. Likewise, that of transition 17 depends on the state of occupation of Q_i. Transition 18 corresponds to the irreversible release of the reductase site inhibitor, and also depends on whether heme c_i is oxidized or reduced. Finally, transition 19 allows electrons to be passed from the low potential chain of one monomer to that of the other, via electron transfer between the two cyt b_L hemes, as has been described for the cytochrome bc_1 complex (Shinkarev and Wraight, 2007) and as expected for the cyt b_6f complexes (Soriano et al. 1999). By symmetry, this transition occurs with the same rate in either direction.

Our model of the cyt b_6f complex is simplified in a number of ways to facilitate development of the methods. Future modifications will allow us to test the consequences of various theoretical models. For example, at present proton transfer reactions at the Q_o and Q_i sites are not included explicitly in our model of cyt b_6f, but are clearly important and play key roles in governing the catalytic properties of the complex. Likewise, we do not at present consider the kinetics of the pivoting of the iron-sulfur protein, which has been proposed to gate electron transfer on the high potential chain (reviewed in Cooley et al., 2008).

Several constraints decrease the number of free parameters. Four of the rates, nos. $+9$, $+11$, $+17a$ and $+17b$, were assumed to be sufficiently rapid as to be non-rate-limiting under any conditions, and thus were treated as constant in the optimization described below. For example, rate $+9$ was treated as a constant and rate -9 as a variable. Part of the rationale is that, in the limit as the rates of two transitions in opposite directions between a pair of states are taken to ∞, while transition rates between neighboring pairs of states remain finite, the resulting process will depend only on the limit of the ratio of the fast rates. The assumption that fast rates may be approximated by the given constants is also supported empirically by the good results obtained below. The four fast rates were set to values that were relatively fast in Table 1, but are perhaps much slower than their true rates. This is helpful both because the number of free parameters is decreased and because the simulation becomes more efficient as the range of time scales simulated is narrowed.

Transition rates are also constrained by the requirement for detailed balance at thermodynamic equilibrium (for example: Hill, 1977). Detailed balance means there is no net

flow of probability between any two states. For example, consider a system with states A and B, and rates $k_{A\to B}$ and $k_{B\to A}$ for transitions between these states. Then detailed balance implies $P_A k_{A\to B} = P_B k_{B\to A}$, where P_A and P_B are the state probabilities at equilibrium. This requirement is nontrivial in the case of the cycle shown in Fig. 7. As described above, the rate of reduction of heme c_i depends on the occupation state of the reductase site and, conversely, binding of a quinone to the reductase site depends on the oxidation state of heme c_i. Beginning with an oxidized heme c_i and an empty binding site, at the upper left hand corner of Fig.7, there are two paths to the state at the lower right. Detailed balance implies that the product of the rates clockwise around the cycle equals the product of rates counterclockwise around the cycle. Solving the resulting equation for k_{3b}^- obtains

$$k_{3b}^- = k_{3a}^- f \quad \text{with} \quad f = \frac{k_{17a}^+ k_{17b}^-}{k_{17b}^+ k_{17a}^-}. \tag{1}$$

Analysis of a similar cycle with quinone replaced by quinol yields $k_{4b}^- = k_{4a}^-$. This simpler result is due to our assumption that reaction 17 proceeds at the same rate, k_{17a}, when the reductase site is empty or occupied by quinol. There is also a cycle of states associated with inhibitor release from the reductase site and this yields $k_{18b}^- = k_{18a}^- f$.

The transition rates are also constrained by midpoint potential measurements. These are values of the ambient electrical potential at which an electron carrier is reduced with probability of ½. Midpoint potentials for all five redox cofactors in cyt b_6f are available from the literature (Kramer and Crofts, 1993; Alric et al., 2005; Cramer et al., 2006). These come from cyt b_6f molecules in three different organisms, so by using them we assume that these molecules are so similar that their midpoint potential values can be used as a consistent set of

constraints. Table 2 shows values of midpoint potentials obtained from the literature and used in the fit to the A data. In the simultaneous fit to the A and B data, the midpoint potentials were treated as free parameters and those values are also shown in Table 2.

To see how midpoint potential values may be used to obtain constraints on transition rates, consider two electron accepting sites, A and B, and an electron transition between them: $A \leftrightharpoons B$. Let $X = A^- \cdot B \cdot C$ be a state of cytochrome b_6f with A reduced and B oxidized and where C denotes a configuration of the other electron accepting sites and the binding sites of the molecule. Let $Y = A \cdot B^- \cdot C$ be the corresponding state with A oxidized and B reduced. Let $Z = A \cdot B \cdot C$ be the state with A and B both oxidized. Then the Nernst equation gives

$$E = E_A^m - \frac{\mathrm{RT}}{\mathrm{zF}} \ln P_X/P_Z \quad \text{and} \tag{2}$$

$$E = E_B^m - \frac{\mathrm{RT}}{\mathrm{zF}} \ln P_Y/P_Z, \tag{3}$$

where E_A^m and E_B^m are midpoint potentials for sites A and B, respectively. R is the gas constant, T the absolute temperature, F Faraday's constant, $z = 1$ is the valence of sites A and B, and P_X denotes the probability of state X. These equations may be solved for P_X/P_Z and P_Y/P_Z and the ratio of the resulting expressions taken to obtain

$$\frac{P_X}{P_Y} = \exp\left(\frac{F}{RT}(E_A^m - E_B^m)\right). \tag{4}$$

Let $k_{A\to B}$ be the transition rate from state A to state B. At thermodynamic equilibrium, detailed balance requires

$$P_X k_{A\to B} = P_Y k_{B\to A}. \tag{5}$$

This may also be solved for P_X/P_Y and the result combined with (4) to obtain the constraint on electron transfer rates:

$$\frac{k_{B\to A}}{k_{A\to B}} = \exp\left(\frac{zF}{RT}(E_A^m - E_B^m)\right). \quad (6)$$

To summarize, there is the following breakdown of the 35 model rates. Six are inaccessible to the experimental system because plastocyanin was removed. Three are determined by detailed balance applied to thermodynamic cycles. Six are determined by midpoint potential measurements. These are set to literatures values in the fits to the A data and they are treated as adjustable parameters in the simultaneous fits to the A and B data. Four are set to constant values because they are fast. One is an inhibitor release rate that is only used in fits to the B data. There are 15 adjustable parameters left for fits to the A data alone. For the simultaneous fits to the A and B data, the inhibitor release rate and the midpoint potentials were also treated as independent variables for a total of 22 adjustable parameters.

2.2 Monte Carlo simulations

All simulations involve a single cyt b_6f dimer on a model membrane with a 6×6 lumen. Initially all cyt b_6f electron carriers are oxidized with no bound substrate molecules. Before any 'flashes', 12 quinones are placed randomly on the membrane lattice. There are two sources of quinol modeling in a simple way two photosystem II reaction centers. They both have the following action. On every second flash, a quinone on the membrane lattice is chosen at random and changed to a quinol. In this way the sum of the numbers of quinones and quinols is conserved. The probabilities of the initial states of the model reaction centers are chosen

randomly and independently, with quinols placed on the lattice at the second and fourth flashes with a probability of 80% and placed on the lattice at the first, third and fifth flashes with a probability of 20%. These probabilities were chosen by eye to match an initial portion of the A data, $0 \leq t \leq 1200$ms, before any attempt to fit the data as a whole. We assume that the probabilities of the initial states of the reactions centers are the same for the A and B experiments. No plastocyanin are placed on the model lumen. The choice of one b_6f dimer and two photosynthetic reaction centers was chosen based on the data of Kirchhoff et al. (2000).

We assumed that the experimental system was not limited by diffusion of quinones and quinols in the membrane, which seems reasonable if photosystem II and cyt b_6f are both in the same small microdomains and is consistent with the conclusions of Tremmel et al. (2007). In other cases, diffusion is thought to become an important limitation (e.g. Kirchhoff, 2008), and the use of the random walk approach will allow future simulations of these effects. The diffusion constants of quinol and quinone were fixed at the fast value of $0.16\ \mathrm{nm}^2\ \mu s^{-1}$. This value was determined by matching the simulations by eye with the initial portion of the data, $0 \leq t \leq 1200\mathrm{ms}$. A much higher value of the diffusion coefficient was initially considered and this was then decreased until the leading edge could not be fit. This reduced value of the diffusion coefficient was then increased by a factor of 16 to obtain the final value used.

A discrete-state continuous-time random walk was simulated using Gillespie's direct method (Gillespie, 1976; 1977). Suppose that at a given time t_0, M different events could happen. These might correspond to movement of a simulated quinol or quinone between adjacent sites on the membrane lattice, binding or release of a quinol or quinone from the

oxidase or reductase sites, or an electronic transition internal to one of the b_6f monomers. Let $a_i\,\Delta t$ be the probability (to first order in Δt) that event i will occur during a short time interval Δt. Let $a = \sum a_i$ be the sum over all of these rates. Let r_1 and r_2 be random numbers uniformly distributed on the unit interval. Then the random time until the next event is chosen as

$$\tau = \frac{1}{a}\ln\frac{1}{r_1}. \tag{7}$$

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

$$\sum_{i=1}^{j-1} a_i \le a\, r_2 < \sum_{i=1}^{j} a_i. \tag{8}$$

Event j is implemented and appropriate variables updated, including the sum a over rates. The process is then repeated by choosing new random numbers r_1 and r_2, etc.

In this way Monte Carlo simulations corresponding to the initial value problem were solved for $0 \le t \le 7.8$s. The average time step is $\bar{\tau} \approx 1\mu s$, so that several millions of steps are required per simulation. Averages over 64 simulations were used to construct the numerical trajectory for comparison with the experimental data. A simple estimate of χ^2 calculated from the average trajectory and the data was minimized using the Latin hypercube sampling technique.

2.3 Estimate of χ^2

An estimate of the variance of the experimental signal, σ_{dat}^2, was obtained from the A data cyt b_H occupancy for $t > 5\ sec$. In this region the probability of occupancy appears to be decreasing approximately linearly as a function of time; see Fig. 3A. A least squares line was

fitted to these data and then the difference between the data and line was considered, subtracting the linear trend. The variance $\sigma_{dat}^2 \approx 0.0013$ was found from this residual. The cyt f occupancy data appears to be almost constant in this region, but its variance may be less typical because cyt f occupancy is near saturation. Similarly, the variance of the cyt b_H results of an early simulation run appeared to decrease approximately linearly as a function of time for $t > 5\ sec$. The linear trend was also subtracted from these results and the residual used to find $\sigma_{sim}^2 \approx 0.0030$. The total variance of the difference between the simulation and data was estimated as $\sigma_{tot}^2 = \sigma_{sim}^2 + \sigma_{dat}^2$. This was used to calculate

$$\chi^2 = \sum_i \frac{(y_i - \hat{y}_i)^2}{\sigma_{tot}^2} \tag{9}$$

where y_i are the data and $\hat{y}_i$ the simulation results and the sum is taken over all results for both cyt f and cyt b. This is a simple estimate of χ^2 statistic; the actual variances of the data and simulations are probably not constant functions of time. However, we are not using this estimate for a statistical test, but only as a function (of the model parameters) to be minimized.

2.4 Parallel processing using Latin hypercube samples

After constructing the base rate set, a domain of possible rates with values within a factor of 2 of the base rates was defined by setting

$$R_i = B_i 2^{E_i}\ , \tag{10}$$

where B_i is the base value for the ith adjustable rate, the rate exponent $E_i \in [-1,1]$ and R_i is the value of the ith rate. Latin hypercube samples of E_i were chosen from the hypercube

$[-1,1]^{15}$ for the fit to data set A and from $[-1,1]^{22}$ for the simultaneous fit to sets A and B. Typical sample sizes were $n = 256$ and $n = 8192$. Uniform distributions were used to sample from the subintervals of $[-1,1]$. The Matlab subroutine lhsu of a public domain package for LHS was used to pre-generate the Latin hypercube samples (Minasny, 2004).

Monte-Carlo simulations were run on a Beowulf cluster, typically using 64 processors simultaneously. The processes were organized in a very simple way as shown on Fig. 8. Processes used a common set of data (including the Latin hypercube samples), the process rank r and the number of processes n. Each processor ran a sequence of sets of simulations, each set averaging together 64 independent simulations using a set of 15 or 22 rates generated from a single LHS sample.

In Fig. 8, results r, for $r = 1 \ldots n$, are sequences of χ^2 values that depend only on the data, the process rank r and the number of processes n. These last two numbers were used to determine the set of Latin hypercube samples for which each processor ran simulations. Otherwise the processes ran independently. At the end of the run, the results files were concatenated and sorted to determine the Latin hypercube samples associated with the lowest values of χ^2, which were further analyzed. The code was written in FORTRAN 77 including Message Passing Interface (MPI) commands. Only 4 distinct MPI commands were required. Programs and data files used for the simultaneous analysis of data sets A and B, with results shown in Fig. 13, are available at www.math.wsu.edu/math/faculty/schumaker/welcome.html.

3 Results

3.1 Preliminary results

We began to construct the base rate set by starting from guesses for parameter values and refining these using several *ad hoc* procedures. These include attempts to fit initial portions of data set A by eye and use of the Nelder-Mead method (Press et al., 1992) to improve the fit, which was used with modest success. The greatest discrepancy between the results of the simulations and experiment was the behavior of cyt b_H redox state at long times, $t \geq 2$ sec. Values of midpoint potentials of redox cofactors that constrain the rate set are available from the literature (Table 2). These values were altered slightly in the initial attempt to fit the data, but were returned to literature values in order to decrease the number of adjustable parameters to 15. This formed the base rate set $\{B_i\}$, for $i = 1, \ldots, 15$. Re-adopting the literature values of the midpoint potentials somewhat increased the discrepancy between the model and experimental values of cyt b_H redox state at long times. Fig. 9 compares an average of 64 Monte Carlo simulations using the base rate set with the A data of Kramer and Crofts. In addition to the difference between the model and experimental values of cyt b_H at long times, there is also a significant discrepancy after the second xenon flash near 700ms.

3.2 Use of Latin hypercube sampling to fit simulations to A data

The initial attempt to improve the fit from that given by the base rate set $\{B_i\}$ generated a set of random rates $\{R_i\}$ chosen according to Eq. (10) using a Latin hypercube sample of rate exponents $\{E_i\}$. Each of the 15 intervals of possible rate exponent values was partitioned into $n = 256$ subintervals and a uniformly distributed random number chosen from each subinterval. These were randomly grouped together to give 256 sample points in $[-1,1]^{15}$. Monte Carlo simulations were run for each of the associated rate sets. A sensitivity analysis

showing how these fits depend on parameter values is given by Fig. 10**.** Panel A is a Matlab box plot showing the quartiles for the rate exponents $\{E_i\}$ as a function of column number i for the 16 best fits to the A data as determined by χ^2. On the box for parameter i, the central mark is the median value of E_i and the edges are the 25th and 75th percentiles. Whiskers extend to the most extreme values not considered outliers. Outliers are plotted individually. If one considers the whole sample of 256 rate sets, ¼ of the sample values in each column would lie between 0.5 and 1, ¼ between 0 and 0.5, etc. Some of the multiplier exponents do seem to be distributed broadly over the entire interval $[-1,1]$, for example those associated with parameter 7. But notice that for parameter 12, the 16 best fits are all associated with multiplier exponents less than 0, and 13 of these are grouped close to -1. These results suggested that the base rate for parameter 12 needed to be decreased. This conclusion was confirmed by the scatterplot shown in panel B, which shows χ^2 values for all 256 sample points as a function of E_{12}. The average value of χ^2 increases on the interval $[-1,1]$, and the three low values of χ^2 for $E_{12} \in [-0.5,0.]$ are clearly outliers with respect to this trend.

After decreasing the base rate for parameter 12 by a factor of 2, Monte Carlo simulations were run for $n = 8192$ Latin hypercube samples. The best fit is shown in Fig. 11. Agreement is generally very good, although the simulated cyt f occupancy is systematically less than the data at large t and the sharp rise of the experimental occupancy values near $t = 0$ is not fully captured. The fit is not perfect but is a big improvement over that given by the base rates. In particular, the large discrepancy between the experimental and simulated cyt b_H occupancy is greatly diminished.

3.3 Optimization of one parameter to fit the B data

The experiments that generated the B data of Kramer and Crofts (1993) were carried out under the same conditions as those that generated the A set except that an inhibitor initially occupied the Q_i site and slowly dissociated over the course of the experiment. If the fit to the A data of Fig. 11 is physically meaningful, we should obtain behavior qualitatively similar to that of the B data by starting from the A data rate set, placing an inhibitor at $t = 0$ on the reductase site in the simulations, and optimizing only its release rate. For simplicity inhibitor release is assumed to be irreversible, giving a new comparison with data involving the optimization of only a single parameter. (The release rate of inhibitor depends on the redox state of heme c_i, as discussed in section 2.1. However, the two release rates are related as described below Eq. (1)). Fig. 12, panel A, shows the fit associated with the lowest χ^2 value computed using blocking rates R calculated from Eq. (10), a pre-determined inhibitor release base rate B and 64 values of the inhibitor rate exponent E uniformly partitioning the interval $[-2,2]$.

Comparing the relationship of the data with the simulation in Fig. 12, panel A, with that of Figs. 3 and 11 shows that optimizing the single degree of freedom modeling blocker release alters the simulation in a qualitatively similar way as the data. This is evidence that the model is capturing a significant aspect of the dynamics of the biological system.

Fig. 12, panel B, shows the 64 χ^2 values as a function of the rate exponent $E \in \{-2, \dots, 2\}$. The value of χ^2 may change by more than one between Monte Carlo simulations using adjacent values of the rate exponent. This variation reflects the stochasticity of the simulations: even tiny changes in a parameter value may change the choice of one of the

hundreds of millions of events that comprise the 64 simulations averaged together to construct an estimate a single value of χ^2, and after a first difference occurs the trajectories modeling two different sets of rates will diverge. Geometrically, Panel B shows samples of a one dimensional section of a χ^2 surface that is a function of all 15 adjustable parameters in the simulation. This is a very noisy function of 15 parameters.

3.4 Simultaneous fit of simulations to the A and B data

If the simulation model is sufficiently accurate physically, and if there are no additional complications, it should be possible to fit both the A and B data sets simultaneously with the same set of rates. We first attempted simultaneous fits using 16 adjustable parameters: the 15 parameters used to fit the data plus the inhibitor release rate. The best fit achieved using Latin hypercube sampling gave a fair approximation to the combined A and B data (not shown).

These runs used literature values of the midpoint potential, obtained from three different papers reporting on three different organisms. Measurement of midpoint potentials is also subject to experimental error. We therefore next included the midpoint potentials in the free parameter set and associated six additional intervals $[-1,1]$ with these new parameters. Midpoint potential values corresponding to the endpoints of these intervals vary by as much as $\Delta V_0 \approx 17.5$mV on either side of the base rate. This value of ΔV is defined by $\exp\left(\frac{zF\,\Delta V_0}{RT}\right) = 2$ because rates determined by midpoint potentials depend on similar exponential factors (see Eq. 6). Inclusion of the inhibitor release rate and the six midpoint potentials gives a model with 22 adjustable parameters.

Parallel simulations using Latin hypercube sampling were used to fit the 22 parameter model to the combined A and B data. The minimum χ^2 value found was lower than that achieved with the literature values of the midpoint potentials. The corresponding fits are shown in Figs. 13. The corresponding rate set is given by Table 1 and the optimized values of the midpoint potentials are given by the right hand column in Table 2. The simultaneous fits to the A and B data may be compared with the 15 parameter fits to the A data alone, given by Fig. 11. At short times, $0 < t < 2000$ms, the fits to the A data are of comparable quality. However, comparison of the $0 < t < 7800$ms fits shows that the discrepancy between the simulated values of the cyt f redox state and the experimental observations for $t > 2000$ms has increased. Comparing Figs. 13 panels A and C shows that the simulated cyt f reduction states fall below the data under the conditions A but above the data under conditions B. Fig. 13 panel D shows that the simulated values are well below those of the data for $0 < t < 700$ms.

4. Discussion

4.1 Simulation-based optimization and Latin hypercube sampling

In simulation-based optimization the objective function is generated by a simulation. This technique has arisen at the interface of operations research and computer science and has been made possible by the increasing power of computers (Fu, 2002). In this paper the objective function is the mean squared error between the data of Kramer and Crofts (1993) and a Monte-Carlo simulation of the Q-cycle model for electron transport in cyt b_6f (Cape et al, 2006). Our optimization technique is to generate a Latin hypercube sample within a finite subset of parameter space and to run simulations at each sample point. The sample point that

gives the minimum value for the objective function is an improved estimate of 'best fit' set of parameters.

Latin hypercube sampling (McKay et al, 1979) is a version of stratified sampling that searches high dimensional spaces more efficiently than random sampling. It has become popular for uncertainty and sensitivity analysis of computationally demanding models (Helton and Davis, 2003, and Helton et al, 2006). The literature discussing LHS in the context of optimization seems to be much smaller than that for uncertainty and sensitivity analysis, but Davey (2008) discusses optimization of a deterministic objective function by a pattern search method incorporating LHS sampling. Even for deterministic functions, search strategies that incorporate a stochastic component help prevent the search from being trapped by local minima. Davey finds that the pattern search method using LHS is much more efficient than a random search. Kleijnen et al. (2005) reviews the design of simulation "experiments" (in which a simulation is repeated many times with different parameter values in order to explore properties of the model) and recommends LHS for exploring the parameter space of simulations involving many parameters and noisy objective functions.

4.2 Approximating simulations by differential equations

As an alternative to simulation-based optimization, it would be convenient to solve a system of differential equations whose solutions gave the mean values of the simulation as a function of time. Solutions would not have the very noisy character of the simulations as illustrated by Fig. 12, panel B, instead they would be a smooth functions of parameters. Solutions with arbitrary transition rates could be compared with simulation results to check the correctness of the

simulation algorithm. They could also be used directly, together with an optimization algorithm, to fit rates to experimental data.

Transitions between the possible states of the cyt b_6f electron accepting sites could be approximated by a Markov process. In particular, transitions between states with an empty binding site and corresponding states with the binding site occupied would be assumed to occur at constant rates, independent of the distribution of substrate molecules in the membrane or aqueous phase. This corresponds to the assumption that diffusion is not rate limiting. If all possible states of occupation are allowed, the Markov chain has $60 \times 32 = 1920$ possible states. The expected values of the state probabilities would be the solution of a system of 1920 linear ODEs. Construction of such a Markov chain solution appears feasible for approximating the mean behavior of the simulations described in this paper, but this approach would rapidly become cumbersome if additional structure were added to the simulation due to a combinatorial explosion of possible states.

4.3 Limitations of model and implications of fits

In many respects, our model is a vastly simplified caricature of the physical system. But it is designed as a "stepping stone" to demonstrate a methodology that would allow more complex versions. The 2D toaster pastry geometry of the thylakoid membrane in the qcycle model is a cartoon of the elaborate 3D geometry and heterogeneity of the thylakoid membrane in higher plants. At most, it represents a rough approximation to a small patch of the physical system. Movements of electrons and protons within the physical enzyme are fundamentally quantum mechanical, and they may conceivably be correlated in ways that cannot be captured by a

Markov model. In the physical system, quinones and quinols must diffuse in the lipid around the protein component of the membrane, which comprises as much as 80% of the membrane area (Kirchhoff, 2008). Transport of quinols and quinones in the physical system may be facilitated by special structures, such as channels, in the protein component. None of this is represented in the simulations. Further, cyt b_6f is just one in a chain of enzymes that control electron transport in the light reactions of photosynthesis. These enzymes are only represented by a distribution of quinones and quinols on the model lattice at the initial time. Our model can only be expected to represent transient reactions involving cyt b_6f that do not involve feedback from other enzymes, possibly in small microdomains in the grana portion of the membrane where one or two photosystem II molecules share a few quinone or quinols with a cyt b_6f dimer (Kirchhoff et al, 2000). However, our simulation may more fully model simplified physical systems, such as lipid vesicles prepared with cyt b_6f and photosystem II only (Rich et al., 1987).

Despite all of the limitations of the qcycle model, fits to the experimental data of Kramer and Crofts (1993) suggest that the model with optimized rates mimics the dynamics of the oxidations states of the cyt b_H and cyt f electron accepting sites in cyt b_6f fairly well under the conditions of the experiment. This result alone shows that the Q-cycle model can account for the data of Kramer and Crofts without invoking bypass reactions (Muller et al., 2002). Minor changes in the redox midpoint potentials compared to the literature values (Table 2) significantly improved the fit to the data. These deviations were between 3 and 17 mV, well within the expected experimental errors for these types of measurements. Moreover, redox midpoint potentials are typically measured using redox mediators under thermodynamic

equilibrium, and may not reflect the working potentials experienced during catalytic turnover. Our results are also consistent with those of Tremmel et al. (2007), who found that electron transfer from PQH_2 to the iron-sulfur cluster of cyt b_6f is limited by the rate of oxidation at the Q_o site. In our simulations diffusion of quinol is not rate limiting and, according to Table 1, quinol binds to the Q_o site more rapidly than it is oxidized by the iron-sulfur cluster (compare the values of k_2^+ and k_7^+).

Even if the diffusion and electron transport mechanisms of our model do represent a fair approximation to the biological system in the context of the experiments, fits to the data of Kramer and Crofts (1993), shown in Fig. 3, with 15 or 22 adjustable parameters, do not guarantee that a unique set of rates will be identified that are good approximations to 'true' values. Spear (1996) discusses the uniqueness of fits of a model system of 5 nonlinear ordinary differential equations with 19 adjustable parameters to observational data. The univariate marginal distribution of parameters associated with good fits to the data extended over the allowed range of values for almost all of the parameters. Spears concludes that the 'best' parameter estimates correspond to an extended and complex region in a high dimensional space.

The interpretation of 'best' fits in our model analysis is also complex but may be clearer than for the study described by Spear. Fig. 10A shows marginal distributions of 15 parameters for the 16 best fits, from an LHS sample of 256 rate sets, of our Q-cycle model to the A data of Kramer and Crofts. Values for parameter 12 are concentrated at one end of the range, and Fig. 10B shows a strong correlation between χ^2 and the rate exponent of parameter 12. The

comparison clearly shows that the best fits in the sample are achieved by values of the rate exponent E_{12} near -1. A further advantage in our study are the B data of Fig. 3, whose values are quite different from those of the A data but that are generated by experiments with only one difference: the presence of a Qi site inhibitor. The qualitative difference between the A and B data are also reflected in a model analysis; the fit shown by Fig. 12A of the qcycle model to the B data is achieved using the parameter set that gave the fit of Fig. 11 and optimizing the value of only a single new parameter modeling the inhibitor release rate. This is an example of a model predicting the outcome of new experiments, which is a paradigm for model confirmation in the natural sciences. However, the fit to the B data shown in Fig. 12A is somewhat crude, and the model should be further tested and refined. Ultimately an optimized rate set could be strongly validated by predicting the results of appropriate further experiments (Frenklach et al., 2003).

Acknowledgements

M.F.S. thanks Lou Gross for suggesting Latin hypercube sampling, Ed Pate for making the Beowulf cluster available and Kevin Cooper for advice on programming on the cluster. He also appreciates the efforts of Chandana Somayaji and Talitha Anderson, undergraduates at WSU who worked on an early version of the qcycle program.

References

Alric J., Y. Pierre, D. Picot, J. Lavergne, F. 2005. Spectral and redox characterization of the heme ci of the cytochrome b6f complex. Proc. Nat. Acad Sci. 102:44, 15860-15865

Blythe, R.A. and M.R. Evans. 2007. Nonequilibrium steady states of matrix-product form: a solver's guide. J. Phys. A: Math. Theor. 40, R333-R441.

Cape, J.L. Bowman, M.K., Kramer, D. M. 2006. Understanding the cytochrome bc complexes by what they don't do. The Q-cycle at 30. Trends Plant Sci. 11, 46-55

Cape, J. L., Bowman, M. K., & Kramer, D. M. (2007) A semiquinone intermediate generated at the Qo site of the cytochrome bc1 complex. Importance for the Q-cycle and superoxide production. *Proc Natl Acad Sci U S A* 104 7887–7892.

Cooley, J. W., Nitschke, W., & Kramer, D. M. (2008) in *The Purple Photosynthetic Bacteria*, eds. Hunter, C. N., Daldal, F., Thurnauer, M. C., & Beatty, J. T. (Springer, Dordrecht, The Netherlands), pp. 451-473.

Cramer, W.A., Zhang, H., Yan, J., Kurisu, G., Smith, J.L. 2006. Transmembrane traffic in the cytochrome b6f complex. Ann. Rv. Biochem. 75, 769-790.

Crofts, A. R. & Wang, Z. (1989) How rapid are the internal reactions of the ubiquinol: cytochrome c_2 oxidoreductase? *Photosynth. Res.* 22, 69-87.

Crofts, A. R. (2004) Proton-coupled electron transfer at the Q_o-site of the bc_1 complex controls the rate of ubihydroquinone oxidation. *Biochim. Biophys. Acta* 1655, 77-92.

Davey K.R. 2008. Latin hypercube sampling and pattern search in magnetic field optimization problems. IEEE Trans. Magn. 44:6 974-977.

Dekker, J.P., Boekema, E.J. 2005. Supramolecular organization of thylakoid membranes in green plants. Biochim. Biophys. Acta 1706, 12-39.

Derrida, B. 2007. Non-equilibrium steady states: fluctuations and large deviations of the density and current. J. Stat. Mech. P07023.

Frenklach M., Packard A., Seiler P., Feeley R. 2004. Collaborative Data Processing in Developing Predictive Models of Complex Reaction Systems. Int. J. Chem. Kin. 36:1 57-66.

Fu M.C. 2002. Optimization for simulation: theory vs. practice. INFORMS J. Comput. 14:192-215.

Gillespie, D.T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comp. Phys. 22, 403-434.

Gillespie, D.T. 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81, 2340-2361.

Helton, J.C., Davis F.J. 2003 Latin hypercube sampling and the propagation of uncertainty in analysis of complex systems. Reliab. Eng. Syst. Safety. 81, 23-69.

Helton, J.C., Johnson, J.D., Sallaberry C.J., Storlie, C.B. 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliab. Eng. Syst. Safety. 91, 1175-1209.

Hill, T.L. 1977. Free energy transduction in biology. Academic Press. New York. p. 6

Kallas, T., (1994) The cytochrome b_6f complex, in: D.A. Bryant (Ed.), The Molecular Biology of Cyanobacteria, Kluwer Academic Publishers, Dordrecht, The Netherlands., pp. 259-317.

Kirchhoff H., Horstmann S., Weiss E. 2000. Control of the photosynthetic electron transport by PQ diffusion microdomains in thylakoids of higher plants. Biochim. Biophys. Acta. 1459:148-168.

Kirchhoff H. 2008. Molecular crowding and order in photosynthetic membranes. Trends Plant Sci. 13 201-207.

Kleijnen J.P.C., Sanchez, S.M., Lucas T.W., Cioppa, T.M. 2005, A user's guide to the brave new world of designing simulation experiments. INFORMS J. Comput. 17:263-289.

Kramer D.M., Crofts A.J. 1993. The concerted reduction of high and low potential chains of the bf complex by plastoquinol. Biochim. Biophys. Acta 1183, 72-84.

Kurisu, G., Zhang, H, Smith, J.L., Cramer, W.A. 2003. Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity. Science 302, 1009-1014.

de Lacrois de Lavalette, A., Barucq L., Alric J., Rappaport F., Zito F. 2009. Is the Redox State of the *c*i Heme of the Cytochrome *b6f* Complex Dependent on the Occupation and Structure of the Qi Site and Vice Versa? J. Biol. Chem. 284: 31, 20822–20829

McKay M.D., Beckman R.J., Conover W.J. 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 21:2 239-245.

Minasny B. 2004. Latin hypercube sampling. Matlab Central, www.mathworks.com/matlabcentral/fileexchange/authors/11803

Muller, F., Crofts, A.R., Kramer, D.M. 2002. Multiple Q-cycle bypass reactions at the Qo site of the cytochrome bc1 complex. Biochem. 41, 7866-7874.

Ort, D. R. & Kramer, D. (2009) Photosynthesis: The Light Reactions. *Encyclopedia of Life Sciences* DOI: 10.1002/9780470015902.a0001309.pub2,

Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P. Numerical Recipes in Fortran. Second. Ed. Cambridge Univ. Press, pp. 402-406.

Rich P., Heathcote, P., Moss D.A. 1987. Kinetic Studies of Electron Transfer in a hybrid system constructed from the cytochrome bf complex and photosystem I. Biochim et Biophys. Acta 892:1, 138-151

Saltelli, A., Ratto M., Tarantola, S., Campolongo, F. 2006. Sensitivity analysis practices: strategies for model-based inference. Reliab. Eng. Syst. Safety. 91, 1109-1125.

Smith, J.L., Zhang, H., Yan, J., Kurisu, G., Cramer, W.A. (2004) Cytochrome bc complexes: a common core of structure and function surrounded by diversity in the outlying provinces, Curr. Opin. Struct. Biol. 14 432-439.

Soriano G.M., Ponamarev M.V., Carrell C.J., Xia D., Smith J.L., Cramer W.A. 1999. Comparison of the cytochrome bc_1 complex with the anticipated structure of the cytochrome b_6f complex: De plus ca change de plus c'est la meme chose. J Bioenerg Biomembr. 31 201–213.

Spear, R.C. (1997) Large simulation models: calibration, uniqueness and goodness of fit. Environ. Modell. Softw. 12:219-228.

Tremmel, I.G., Kirchhoff H., Weis E., Farquhar, G.D. (2003) Dependence of plastoquinol diffusion on the shape, size, and density of integral thylakoid proteins. Biochim. Biophys. Acta 1607:97-109.

Tremmel, I.G., Weis, E., Farquhar G.D. (2007). Macromolecular crowding and its influence on possible reaction mechanisms in photosynthetic electron flow. Biochim. Biophys. Acta 1767: 353-361

Table 1 Rates used in simulations of Fig. 13

Symbol[1]	Value[2]	Description[3]
k_1^+	$5.2\ 10^{-4}$	entrance rate of quinone into oxidase site
k_2^+	$1.5\ 10^{-3}$	entrance rate of quinol into oxidase site
k_3^+	$4.6\ 10^{-3}$	entrance rate of quinone into reductase site
k_4^+	$4.1\ 10^{-5}$	entrance rate of quinol into reductase site
k_1^-	$1.6\ 10^{-3}$	exit rate of quinone from oxidase site
k_2^-	$3.4\ 10^{-3}$	exit rate of quinol from oxidase site
k_{3a}^-	$2.6\ 10^{-6}$	exit rate of quinone from reductase site when c_i oxidized
k_{3b}^-	$1.2\ 10^{-2}$	exit rate of quinone from reductase site when c_i reduced
k_{4a}^-	$2.2\ 10^{-3}$	exit rate of quinol from reductase site when c_i oxidized
k_{4b}^-	$2.2\ 10^{-3}$	exit rate of quinol from reductase site when c_i reduced
k_7^+	$3.3\ 10^{-4}$	rate for $o2 + R0 \rightarrow o1 + R1$
k_7^-	$3.3\ 10^{4}$	rate for $o1 + R1 \rightarrow o2 + R0$
k_8^+	$7.6\ 10^{6}$	rate for $o1 + L0 \rightarrow o0 + L1$
k_8^-	$3.6\ 10^{0}$	rate for $o0 + L1 \rightarrow o1 + L0$
k_9^+	$3.7\ 10^{0}$	rate for $L1 + H0 \rightarrow L0 + H1$
k_9^-	$1.3\ 10^{-1}$	rate for $L0 + H1 \rightarrow L1 + H0$
k_{10a}^+	$1.2\ 10^{-5}$	rate for $x1 + i0 \rightarrow x0 + i1$
k_{10a}^-	$3.1\ 10^{-5}$	rate for $x0 + i1 \rightarrow x1 + i0$
k_{10b}^+	$4.8\ 10^{2}$	rate for $x1 + i1 \rightarrow x0 + i2$
k_{10b}^-	$5.4\ 10^{-2}$	rate for $x0 + i2 \rightarrow x1 + i1$
k_{11}^+	$9.3\ 10^{-2}$	rate for $R1 + f0 \rightarrow R0 + f1$

k_{11}^{-}	$1.3\ 10^{-2}$	rate for $R0 + f1 \rightarrow R1 + f0$
k_{17a}^{+}	$6.5\ 10^{1}$	rate for $H1 + x0 \rightarrow H0 + x1$ when reductase site is empty or occupied by semiquinone or quinol.
k_{17a}^{-}	$3.3\ 10^{-1}$	rate for $H0 + x1 \rightarrow H1 + x0$ when reductase site is empty or occupied by semiquinone or quinol.
k_{17b}^{+}	$4.8\ 10^{-2}$	rate for $H1 + x0 \rightarrow H0 + x1$ when reductase site blocked or occupied by quinone
k_{17b}^{-}	$1.1\ 10^{0}$	rate for $H0 + x1 \rightarrow H1 + x0$ when reductase site blocked or occupied by quinone
k_{18a}^{-}	$6.1\ 10^{-10}$	rate for blocker release from reductase site when heme c_i oxidized.
k_{18b}^{-}	$2.7\ 10^{-6}$	rate for blocker release reductase site when heme c_i reduced
k_{19}	$3.0\ 10^{-5}$	rate for cyt $b_L \leftrightarrow$ cyt b_L intermonomer transition

[1]Values for $k_5^+, k_5^-, k_6^+, k_6^-, k_{12}^+$ and k_{12}^- were not obtained as the transitions were inaccessible in the experiments of Kramer and Crofts (1993).

[2]Values are for the best simultaneous fit to the A and B data. Units are inverse microseconds. Substrate entrance rates are pseudo first order rates assuming a substrate concentration of $Q_0 = 0.01\ \mathrm{nm}^{-2}$.

[3]Notation for electron accepting sites: o oxidase site; i reductase site; R Rieske; f cyt f; L cyt b_L; H cyt b_H; x heme c_i. Numerals refer to the number of high free energy electrons at the site, for example $o2$ refers to a quinol at the oxidase site, $o1$ refers to a semiquinone at that site, and $o0$ refers to a quinone at that site.

Table 2 Midpoint Potential Values used in simulations of Fig. 13

Symbol	Literature Value[1]	Simultaneous fit[1]
E_H^m	-50.	-53.
E_L^m	-130.	-138.
E_{xa}^m	100.	83.
E_{xb}^m	-125.	-132.
E_f^m	350.	337.
E_R^m	300.	287.

[1]Units are millivolts versus standard hydrogen electrode.

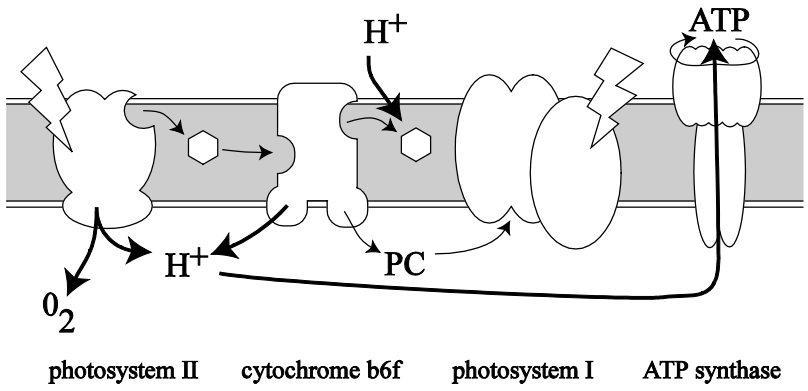

Fig. 1 Light Reactions of Photosynthesis. Four proteins are depicted in the thylakoid membrane. The membrane is shown between two aqueous regions, the chloroplast stroma above and the thylakoid lumen below. Plastoquinols are shown as hexagons entering the Q_o site of cyt b_6f on the left and exiting the Q_i site on the right.

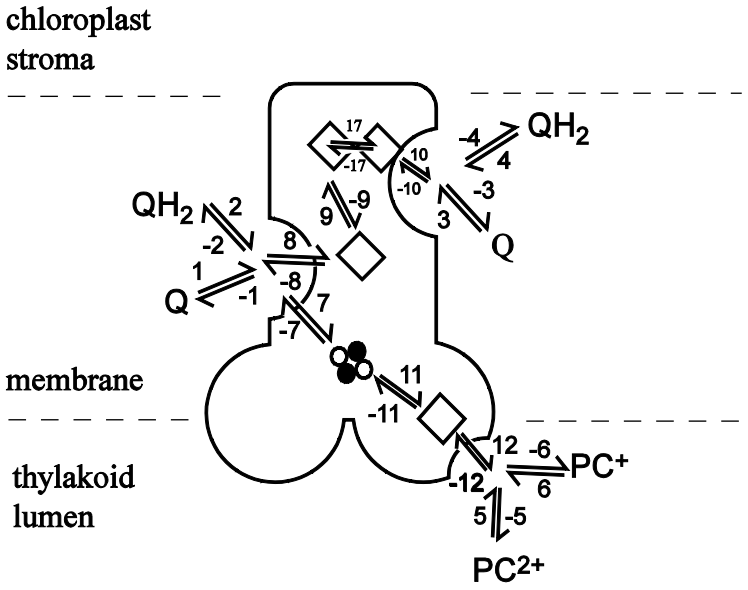

Fig.2 Electron transitions of cytochrome b_6f. The Q_o site is on the left and the Q_i site is on the upper right. The high potential chain allows electrons to migrate from the Q_o site to the plastocyanin binding site at the lower right. From left to right, it includes the Rieske iron-sulfur cluster and cyt f. The low potential chain allows electrons to migrate from the Q_o site to the Q_i site. It includes cyt b_L, cyt b_H and heme c_i.

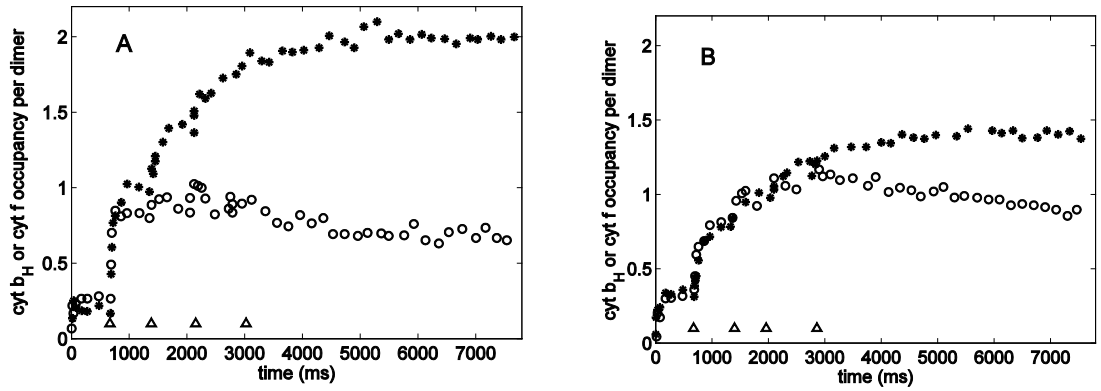

Fig. 3 Data of Kramer and Crofts (1993; fig. 3). A flash of light excites photosystem II at time $t = 0$ and four additional times indicated by triangles. Asterisks give experimental values of cyt f occupancy and circles give experimental values of cyt b_H occupancy. (A) Experiment without quinone reductase site blocker. (B) Experiment with reductase site blocker.

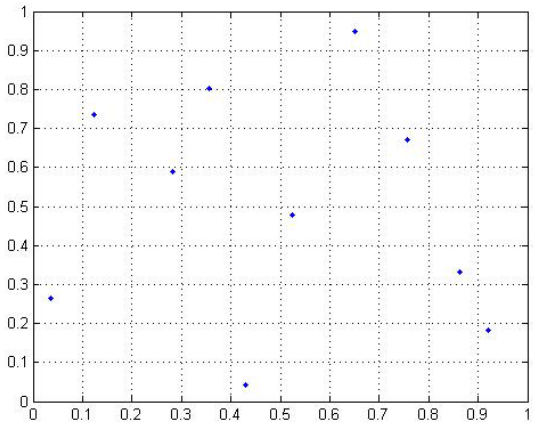

Fig.4 Example of Latin hypercube sampling with 2 parameter intervals partitioned into 10 subintervals each. A single random number is chosen from a uniform distribution on each subinterval. These are randomly paired to give a sample of 10 points on the unit square.

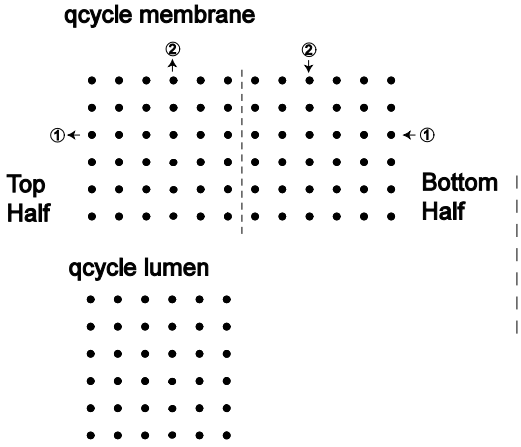

Fig.5 Toaster pastry model of the thylakoid membrane

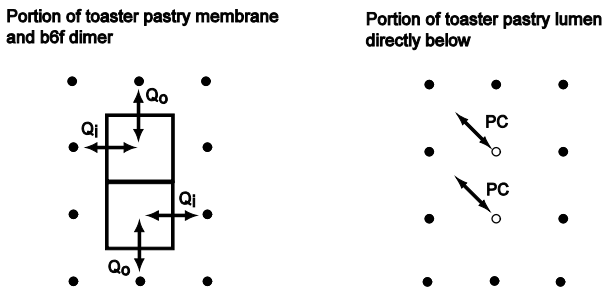

Fig.6 Model of b_6f dimer in the toaster pastry lattice. Quinol oxidase site are denoted Q_o and quinone reductase sites are denoted Q_i.

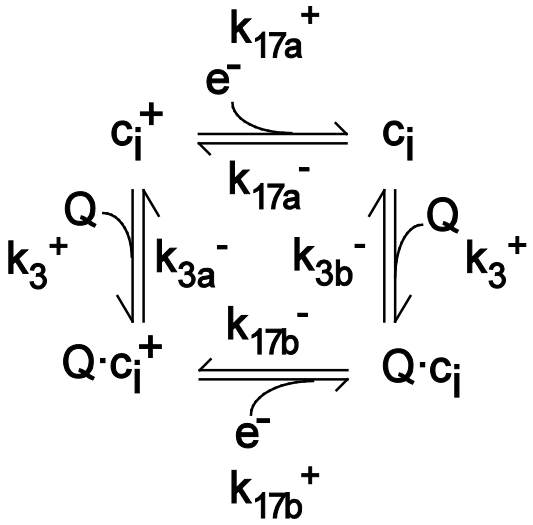

Fig. 7 Thermodynamic cycle used to obtain a relationship between rate constants at the quinone reductase site implied by detailed balance.

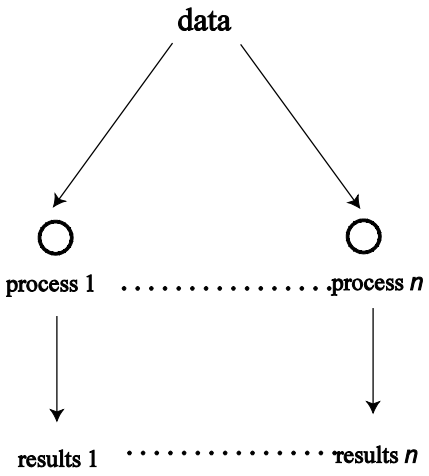

Fig. 8 Scheme of parallel processes running on Beowulf cluster. Processes 1 through n run on different nodes of the cluster, indicated by circles. They depend on a common set of data and produce separate results files, which are later concatenated.

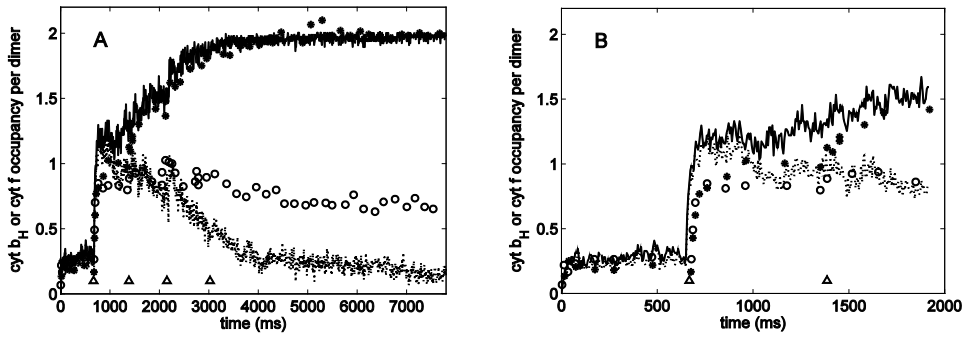

Fig. 9 Base fits to the A data of Kramer and Crofts. Symbols are described in the legend of Fig. 3. The jagged solid curves denote simulated values of cyt f occupancy and the dotted curve denotes simulated values of cyt b_H occupancy. (A) Entire data set: $0 \leq t \leq 7800$ms. (B) First 2000 ms on an expanded scale.

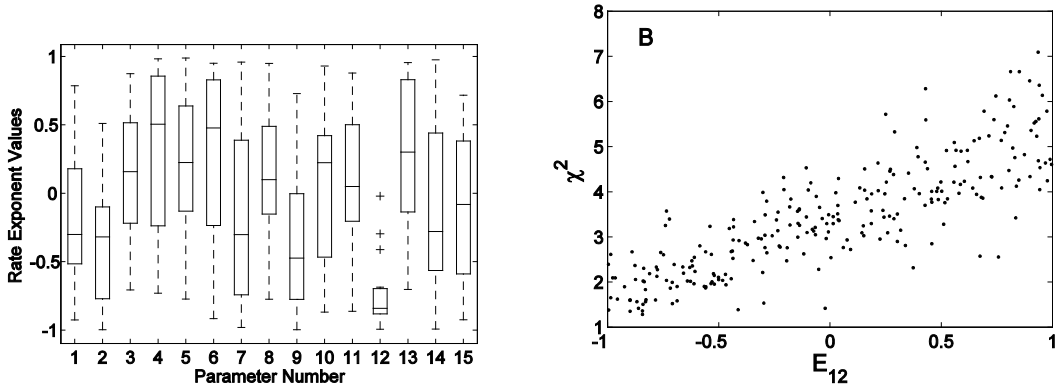

Fig. 10 Sensitivity Analysis of fits to the A data corresponding to a Latin hypercube sample of 256 rate sets. Each rate set gives values for rate exponents of 15 Monte Carlo simulation parameters. (A) Matlab boxplot showing quartiles for rate exponents E_i as a function of parameter number i for the 16 sample rate sets giving the smallest values of χ^2. (B) Scatter plot showing values of χ^2 for all 256 rate sets as a function of E_{12}.

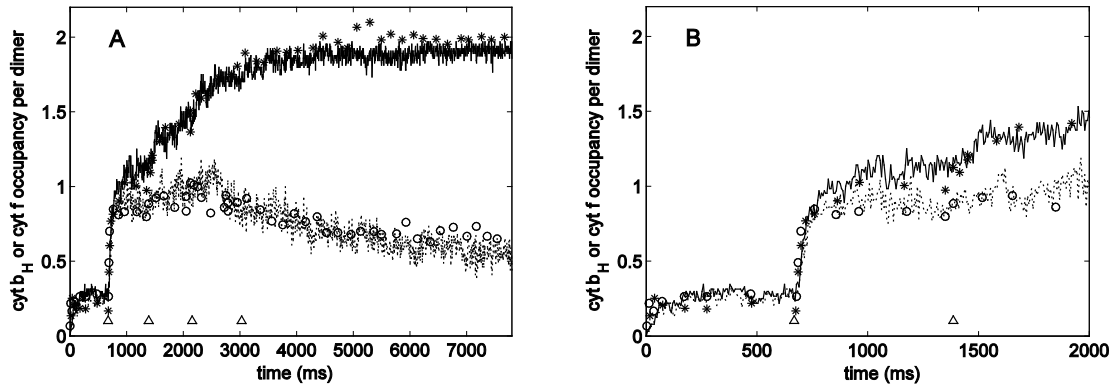

Fig. 11 Best Latin hypercube sampling fit of the 15 parameter model to the A data of Kramer and Crofts (1993). Symbols and curves as in Figs. 3 and 9. (A) Entire data set: $0 \leq t \leq 7800$ms. (B) First 2000 ms on an expanded scale.

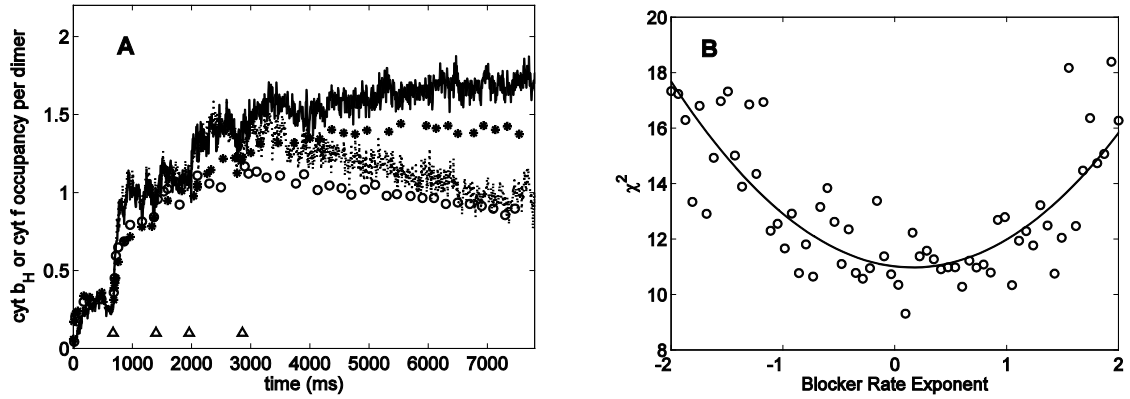

Fig. 12 (A) Fit to B data obtained by starting from fit to A data of Fig. 11 and optimizing a single reductase site inhibitor release rate. Symbols and curves as in Figs. 3 and 9. (B) χ^2 values are plotted as a function of inhibitor release rate exponent. The base rate, corresponding to exponent 0, is close to the rate used to obtain the fit in panel A. The curve shows a least squares fit of a parabola to the χ^2 estimates, and is given to guide the eye.

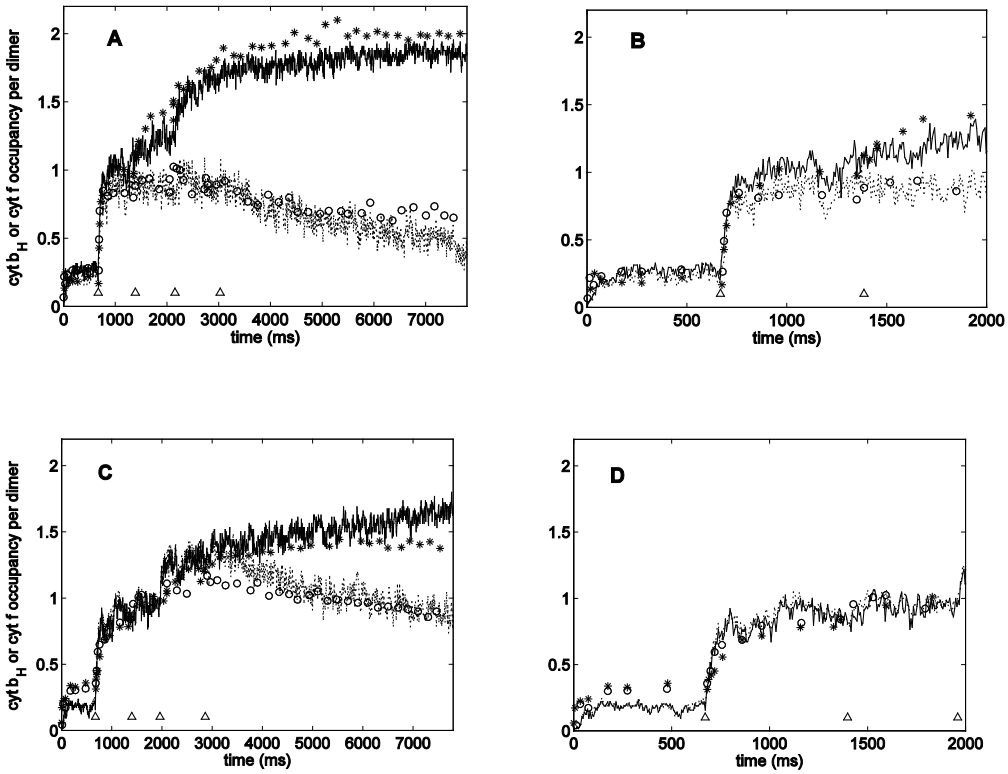

Fig. 13 Simultaneous fit of 22 parameter model to A and B data. Symbols and curves as in Figs. 3 and 9. (A) Fit to A data for $0 \leq t \leq 7800$ms. (B) First 2000 ms of A on an expanded scale. (C) Fit to B data for $0 \leq t \leq 7800$ms. (D) First 2000 ms of B on an expanded scale.