

# Determination of a Unique Solution to Parallel Proton Transfer Reactions Using the Genetic Algorithm

D. Moscovitch,\* O. Noivirt,<sup>†</sup> A. Mezer,<sup>†</sup> E. Nachliel,<sup>†</sup> T. Mark,<sup>†</sup> M. Gutman,<sup>†</sup> and G. Fibich\*

\*School of Mathematics, The Faculty of Exact Sciences, and <sup>†</sup>Laser Laboratory for Fast Reactions in Biology, Department of Biochemistry, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

**ABSTRACT** Kinetic analysis of the dynamics as measured in multiequilibria systems is readily attained by curve-fitting methodologies, a treatment that can accurately retrace the shape of the measured signal. Still, these reconstructions are not related to the detailed mechanism of the process. In this study we subjected multiple proton transfer reactions to rigorous kinetic analysis, which consists of solving a set of coupled-nonlinear differential rate equations. The manual analysis of such systems can be biased by the operator; thus the analysis calls for impartial corroboration. What is more, there is no assurance that such a complex system has a unique solution. In this study, we used the Genetic Algorithm to investigate whether the solution of the system will converge into a single global minimum in the multidimensional parameter space. The experimental system consisted of proton transfer between four proton-binding sites with seven independent adjustable parameters. The results of the search indicate that the solution is unique and all adjustable parameters converge into a single minimum in the multidimensional parameter space, thus corroborating the accuracy of the manual analysis.

## INTRODUCTION

The signal of measured dynamics can be reconstructed by a variety of fitting procedures, sum of exponents, polynomials, damped sinusoid function, etc. These curve fittings can be very accurate, hardly missing a stray experimental point, yet their predicting power is limited. The limitations are a direct consequence of the mode of signal reconstruction. It follows the shape of the observation, not being the consequence of the chemistry/mechanism that generates the observation. An alternative method for reconstructing an observation is to deduce, on theoretical ground, the reactions that are involved in the process, to sum them up in a set of coupled nonlinear differential rate equations and to propagate the equations over time with suitable adjustable parameters. A systematic search over a multidimensional parameter space can yield a combination (or combinations) of the rate constants that will satisfactorily reconstruct the observed signal. When a set of parameters that satisfied this demand is encountered, there is no assurance as to whether it is a global minimum in the parameter space or is one of many local minima. Once the set is defined as a global minimum, the rate constants can be physically interpretable. In this article, we apply the Genetic Algorithm for the evaluation of the uniqueness of the rate constants determined for a complex set of reactions where the mechanism is known, but the rate constants have to be determined.

The pioneering experiments of fast perturbation of systems in chemical equilibria were carried out by Eigen (1964). The analysis, however, was based on linearization of an inherently nonlinear process. This shortcoming was amended by

Gutman who introduced the laser-induced proton pulse technique (Gutman and Huppert, 1979), which modeled the relaxation as a nonlinear process (Gutman, 1984). Reaction systems as complex as proteins in solution were pulsed by free protons that react simultaneously with all proton-binding sites present on the protein, thus initiating a sequence of reactions that relax over a wide range of timescale, from the nanoseconds up to milliseconds. The analysis was carried out by reconstruction of the observed signals by numeric integration of coupled, nonlinear ordinary differential equations with linear and quadratic terms. The coefficients appearing in these equations are a combination of rate constants, equilibrium constants, and the concentrations of the reactants. The system was solved by a search within the parameter space, for any combination of adjustable parameters that reconstruct many, independently measured observations (Checover et al., 1997, 2001; Yam et al., 1988; Marantz et al., 2001; Nachliel and Gutman, 2001). In such systems, the search for the value of the unknown parameters is extremely laborious, and the experience and thorough understanding of the chemical nature of the reactants become a crucial element for a successful search. The analysis is also prone to another kind of criticism as summarized by the general belief: “Three parameters are sufficient to draw an elephant and with four it will also dance.” This poses a legitimate question: to what extent can the reconstruction of many experimental signals (all gathered under different experimental conditions) by a single set of kinetic parameters be taken as evidence that it consists as the only solution of the kinetics problem? The main goal of this study is to demonstrate that the search process for the rate constants can be fully automated using the Genetic Algorithm. This can lead to the subjection of large multiequilibria systems to detailed kinetic analysis. For this reason, in this study our goal was not to find a system that

---

Submitted January 12, 2004, and accepted for publication March 9, 2004.

D. Moscovitch and O. Noivirt contributed equally to this study.

Address reprint requests to M. Gutman, E-mail: me@hemi.tau.ac.il.

© 2004 by the Biophysical Society

0006-3495/04/07/47/11 \$2.00

doi: 10.1529/biophysj.104.039925

assures a fast conversion. On the contrary, the algorithm was set to search over a wide range of parameters, to make sure that any local minimum in the parameter space will be detected.

Since the first presentation by Holland (1970), the Genetic Algorithm had attracted a lot of interest and was applied to a multitude of scientific fields, including molecular modeling, polymer design, protein folding, etc. (for review see Leardi, 2001 and references therein). Yet, its application to the analysis of complex chemical processes was rather limited. Attempts to utilize the Genetic Algorithm for solving chemical kinetic problems are not numerous (Filipic et al., 2000; Hongqing et al., 1999; Houck and Kay, 1995; Terry and Messina, 1998), and a system comparable to ours was described by Viappiani et al. (1998). However, the mechanism investigated by Viappiani was rather simple and only two rate constants were searched for.

In this study, we investigated a system with inherent complexity made of seven independent variables, and we tested to what extent the Genetic Algorithm can solve the various proton transfer pathways. The system consists of a proton emitter (pyranine), indicator (fluorescein), and the bicarbonate ion that is made by the solvation of the atmospheric  $\text{CO}_2$ . What is more, the indicator has two proton-binding sites: the oxyanion of the xanthene ring and the carboxylate of the benzene as independent sites, which can also exchange protons among themselves. Of the four reactants, only two are directly observed, whereas the other two have no spectral signature and the dynamics should be concluded from their modulation of the protonation dynamics of the observed reactants. The level of complexity of this system is comparable to the case of studying the protonation dynamics of an indicator attached to a protein (Nachliel et al., 1996; Nachliel and Gutman, 2001; Shimoni et al., 1993; Yam et al., 1988). The rate constants of protonation of these reactants had been measured previously by manual analysis and their pK values are directly measured. To increase the complexity of the system, the concentration of the bicarbonate had to be treated as an adjustable parameter, its concentration varied daily, reflecting the direction that the wind was blowing  $\text{CO}_2$  from the nearby power station.

The search for the uniqueness was carried out using the Genetic Algorithm under conditions that allowed an intensive search over the parameters' space. The first generation consisted of 100 "phenotypes", each having randomly selected values for the adjustable parameters. The convergence was rather slow, and the calculations proceeded well after the fitness function seemed to reach a constant level, to make sure that the solution was stable and had no tendency to drift into a new set of adjustable parameters. What is more, the same set of experimental signals was subjected to repeat analytic runs and the values of the adjustable parameters were subjected to statistical analysis, as to whether the values of each parameter are members of a single population. The calculations were carried out first

with computer-generated signals, made by the integration of the differential rate equations with known values assigned for each adjustable parameter. When the target signals are noiseless, as in the computer-generated signals, the convergence of the Genetic Algorithm was definitely unique and the values derived by the program were practically identical to those used to create the signals. This is an indication that a global minimum for the fitness function in the multidimensional parameters' space can be obtained. Experimental curves are more difficult to analyze due to the electronic noise and experimental uncertainties, and global minimum was not observed. Yet, when more than one signal was subjected to the analysis, the convergence was efficient and each of the adjustable parameters acted as representing a single population of values.

## EXPERIMENTAL SYSTEM

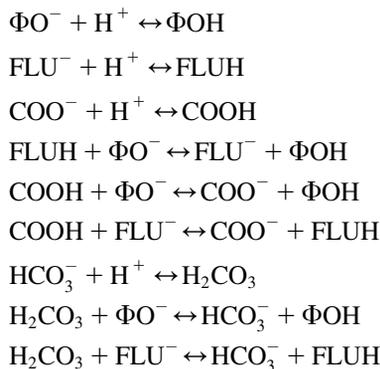
The system selected for study was of moderate complexity, and consisted of four proton-binding residues: 1), a pyranine molecule ( $\Phi\text{OH}$ ), which ejects a proton when excited by a photon; 2), fluorescein (Flu), a pH indicator that has two proton-binding sites, the first of which is the oxyanion of the xanthene ring. The protonation of this site causes a major spectral shift of the indicator; 3), the second proton-binding site of fluorescein is the carboxylate moiety of the fluorescein molecule, a site that differs from the chromophore proton-binding sites by its pK value and whose protonation does not bear any spectral signature in the visible range; and 4), a buffer molecule (BH) with unknown concentration that reacts with the proton but does not generate a measurable signal. In this study, the buffer is the bicarbonate anion,  $\text{HCO}_3^-$ , which is spontaneously generated by the solvation of the atmospheric  $\text{CO}_2$ . When the experiments are carried out at varying pH values, the bicarbonate concentration increases with the experimental pH. A further experimental complication had to be dealt with in this study; due to the proximity of the laboratory to a power station, the  $\text{CO}_2$  content in the air varied daily, depending on the wind's direction. For this reason, the bicarbonate concentration had to be introduced as an adjustable parameter.

The measurements were carried out as described in Checover et al. (1997). Briefly, a 100-mM NaCl aqueous solution was supplemented with pyranine (20  $\mu\text{M}$ ) and fluorescein (10  $\mu\text{M}$ ), equilibrated with the air at two pH values (6.8 and 7.3), and subjected to a train of laser pulses (1–1.5 mJ/pulse; 10 Hz, 355 nm, 3 ns full-width half maximum). The absorption transients were recorded at 458 and 496 nm, where pyranine and fluorescein are, respectively, absorbing. The time resolutions of the measurements were either 30 ns or 300 ns per data point and the readings were converted into concentration units using the differential extinction coefficients 24,000 and 50,000  $\text{M}^{-1}\text{cm}^{-2}$  for pyranine and fluorescein, respectively.

The kinetics measurements were carried out at each pH value at two wavelengths, 458 and 496 nm. During the measurements the pH of the solution was maintained as constant within  $\pm 0.05$  pH unit. The signal/noise ratio of the signals at the maximal amplitude exceeded 100:1. The two vectors (concentrations versus time) were analyzed in tandem.

## THE MATHEMATICAL MODEL

The measured system consisted of the equilibria defined below:



At the zero time point, a certain fraction of  $\Phi OH$  is dissociated by a brief laser pulse, and equal increments of  $H^+$  and  $\Phi O^-$  are generated. The incremental concentration of these reactants relaxes by reaction with all other proton-binding sites present in the solution; the two sites on the fluorescein, the  $HCO_3^-$  and the pyranine anion itself  $\Phi O^-$ . In parallel, the protonated states of each compound can deliver, by a collisional mechanism, a proton to any other site having a higher pK value.

All these reactions proceed in parallel with velocities that are given by the products of the reactants on the left side of the equation times the rate constant minus the product of the reactant on the right side of the equation times the rate constant. As the concentrations of all reactants vary with progression of time, the velocities will vary with time, in accordance with the temporal concentration of each reactant. All these relations are incorporated into a set of differential rate equations having the following form:

$$d[C_i]/dt = -\sum k_{ij}([C_i][C_jH]) + \sum k_{-i-j}([C_iH][C_j]). \quad (1)$$

The velocity of the reactions is given by the general expression as in Eq. 1 where  $C_i$  and  $C_j$  denote the concentrations of the unprotonated state of the reactants, and  $[C_iH]$  and  $[C_jH]$  are the concentrations of their protonated forms.  $k_{ij}$  and  $k_{-i-j}$  are the rate constants of the forward and backward rate constants of the proton transfer reactions with any other reactant ( $j$ ).

We denote by  $V_i$  the deviation in the concentration of reactant  $i$  from its equilibrium level. The initial conditions are:

$$V_1(0) = X_0 \quad \text{and} \quad V_{i>1}(0) = 0.$$

$X_0$  denotes the initial increment of the free proton that was released by the laser pulse and its value depends on the laser pulse energy. The integration of the differential rate equations was carried out using the rate constants of the following reactions as adjustable parameters:

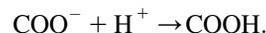
1. The protonation of the pyranine anion ( $k_1$ ).



2. The protonation of the oxyanion of the fluorescein ( $k_2$ ).

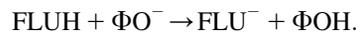


3. The protonation of the carboxylate of the fluorescein ( $k_3$ ).

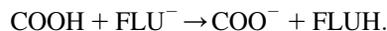


In accordance with the Debye-Smoluchowski equation (Gutman and Nachliel, 1997), the rate constants of reaction 2 and 3 should be the same ( $k_2 = k_3$ ). This equality was imposed as a restriction on the system. Still, as the pK values are not identical, the dissociation rate constants of the two reactions were different.

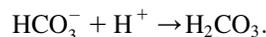
4. The proton transfer from the oxyanion of the fluorescein to the pyranine ( $k_4$ ).



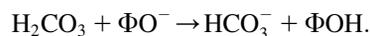
5. The proton transfer from the carboxylate of the fluorescein to the oxyanion of the same molecule ( $k_5$ ).



6. The protonation of the bicarbonate present in the reaction mixture ( $k_6$ ).



7. The rate of proton transfer from the carbonic acid (generated by the protonation of the bicarbonate) with the pyranine anion ( $k_7$ ).



For the sake of simplicity, the proton transfer between the fluorescein and the  $HCO_3^-$  (both having comparable pK values) was ignored.

The proton transfer reaction between the carboxylate and the pyranine was also omitted from the search, the electrostatic repulsion between them is strong and the pK of the carboxylate is low, as a result, the proton dissociation from the carboxylate is sufficiently fast to deplete the COOH state well before an encounter between the COOH and  $\Phi\text{O}^-$  ion.

The last adjustable parameters are the concentrations of the bicarbonate present in the reaction mixture, values that varied with the pH of the solution. The concentration of the pyranine and fluorescein were experimentally measured. The pK values of the two dyes were determined by spectrophotometric titrations and that of the bicarbonate was taken as published (Robinson and Stock, 1959).

### Range of unknown parameters

Each adjustable parameter was allowed to vary within a given range depending on its nature. The rate constants in the thermodynamic-favored direction were allowed to vary from the upper limit set by the Debye-Smoluchowski equation (Gutman and Nachliel, 1997), down to a lower limit, which was set to be four orders of magnitude smaller than the upper limit. The rate constants in the reverse direction were related with the forward ones through the equilibrium constant,  $k_{-i} = k_i/10^{-\text{pK}_i}$ .

The initial magnitude of the perturbation  $X_0$  was estimated directly by back extrapolation of the absorbance at 458 nm during the first 0.5  $\mu\text{s}$  to  $t = 0$ .

Due to the large variation in  $\text{CO}_2$ , the bicarbonate concentration was allowed to vary within a wide range, 5–500  $\mu\text{M}$ .

### THE FITTING PROBLEM

To gain high confidence in the results of the analysis, the fitness function was always calculated for a pair of experimental results measured at the same pH; one is the reprotonation dynamics of the pyranine anion ( $X_i$ ) and the other is the reversible protonation of the fluorescein ( $Y_i$ ). In the following text we shall refer to the two experimental tracings ( $X_i$  and  $Y_i$ ) as one signal.

For any given set of values of the unknown parameters, the ordinary differential equations can be solved using standard numerical subroutines such as Matlab's ODE23s. The level of agreement between the experimental signals and the numerical solutions can be expressed by a fitness function ( $Ft$ ), which is a weighted average of the squares of the differences between the calculated solutions ( $X_i^{\text{calculated}}$ ) and the measured signals ( $X_i^{\text{signal}}$ ), i.e.,

$$Ft(\text{adjustable parameter } 1-7) = \sum_i e^{X_i/X_{\text{max}}} ((X_i^{\text{calculated}} - X_i^{\text{signal}})/X_i^{\text{signal}})^2 + \sum_i e^{Y_i/Y_{\text{max}}} ((Y_i^{\text{calculated}} - Y_i^{\text{signal}})/Y_i^{\text{signal}})^2, \quad (2)$$

where the  $i$  in the summation stands for the value at time  $t_i$ . Our past experience with a manual search for the parameters led us to use the weight functions,  $e^{X_i/X_{\text{max}}}$  and  $e^{Y_i/Y_{\text{max}}}$ , as a measure to make the system more sensitive to these sections of the observed signal where its amplitude is large. Thus, the fluctuations of the signal near the zero level, at the end of the measurement, will make a smaller contribution to the fitness function than the large signal at the early phase of the relaxation.

The fitting problem is not only to find all the parameter combinations that minimize the fitness function. We also have to determine whether the solution is unique. For this reason, the analysis is repeated and the values of each adjustable parameter are subjected to statistical analysis. In the case where the values conform with normal distribution, we can consider them as members of a single population, thus implying that the Genetic Algorithm found the global minimum. Where the values found for one (or more) adjustable parameter appear to be members of more than one population we shall have to conclude that the analysis failed to find a global minimum in the multidimensional parameters' space.

### GENETIC ALGORITHM

In cases of high-dimensional optimization problems with possible nonsmooth fitness function and multiple local minima, a natural choice for optimization of the fitness function is the Genetic Algorithm. In this study each generation consisted of 100 "phenotypes", each of them having a random set of adjustable parameters, selected within the permitted range (see above). The program used these values to reconstruct a signal and to calculate the fitness. At the end of the generation, the best-fit phenotype was cloned and replaced the worst-fitting one. Besides that, the genes were manipulated by the following alternations: two heuristicXover, two arithXover, two simpleXover, four boundaryMutation, six multiNonUnifMutation, two nonUnifMutation, and two unifMutation. All these genetic manipulations are standard procedures and defined in the GAOT program of Matlab (Houck and Kay, 1995).

The fitness function was calculated for all new combinations and their fitness function values were evaluated among themselves and in comparison with the fitness function calculated for the previous generation for the unaltered phenotypes of their parent generation.

Our choice of Matlab as the computational platform is motivated by its portability across different platforms and by the availability of the genetic algorithm toolbox GAOT (Houck and Kay, 1995). To demonstrate the robustness of our methodology, we use the default parameters of GAOT, rather than trying to optimize performance by varying with GAOT's parameters (the selection and termination functions, etc.).

In this study the Genetic Algorithm was searching for the minimum of the fitness function in a seven-dimensional

space. The length of each run, lasting 3000 generations varied among the computers used for the calculations, ranging from 2 to 6 h, depending on the processors.

## RESULTS AND DISCUSSION

### Reconstruction of the experimental signal by the Genetic Algorithm

Fig. 1 depicts a set of reconstructions of experimental signals as they evolve with the number of generations. The frames on the left side depict how the relaxation of the pyranine, from its enforced dissociation, progresses with the number of generations. The curve descends nonmonotonically, yet as it does not inverse its direction, the convergence is rather fast and even after 10 generations, the calculated line resembles the experimental one. The frames on the right reproduce a more complex feature, characterized by a minimum, so that the fitting process calls for more generations before the final shape is attained. The coupling of the two experimental curves into the same fitting process definitely slows the

convergence, although it ensures that all chemical events are equally expressed in the reconstruction of the dynamics.

### Quantitative evaluation of the fitness function

Fig. 2 depicts the reduction of the fitness function as a function of the number of generations for 10 consecutive runs of the same signal. As seen in the figure, the initial values of the fitness function varied widely, from 300 to 900, yet by the end of the run, all converged to a narrow range of  $0.1 \pm 0.05$ . The variation of the fitness function during the process followed different pathways. In some cases, the convergence was fast (*blue trace*), whereas in others the process was slow and  $>1000$  generations were needed to reach the same value that the fast converging run had made within 100 generations.

The convergence of the fitness function implies that, in all runs, the final shapes of the reconstructed curves are very similar. However, this does not mean that the adjustable parameters had converged to a global minimum. It is

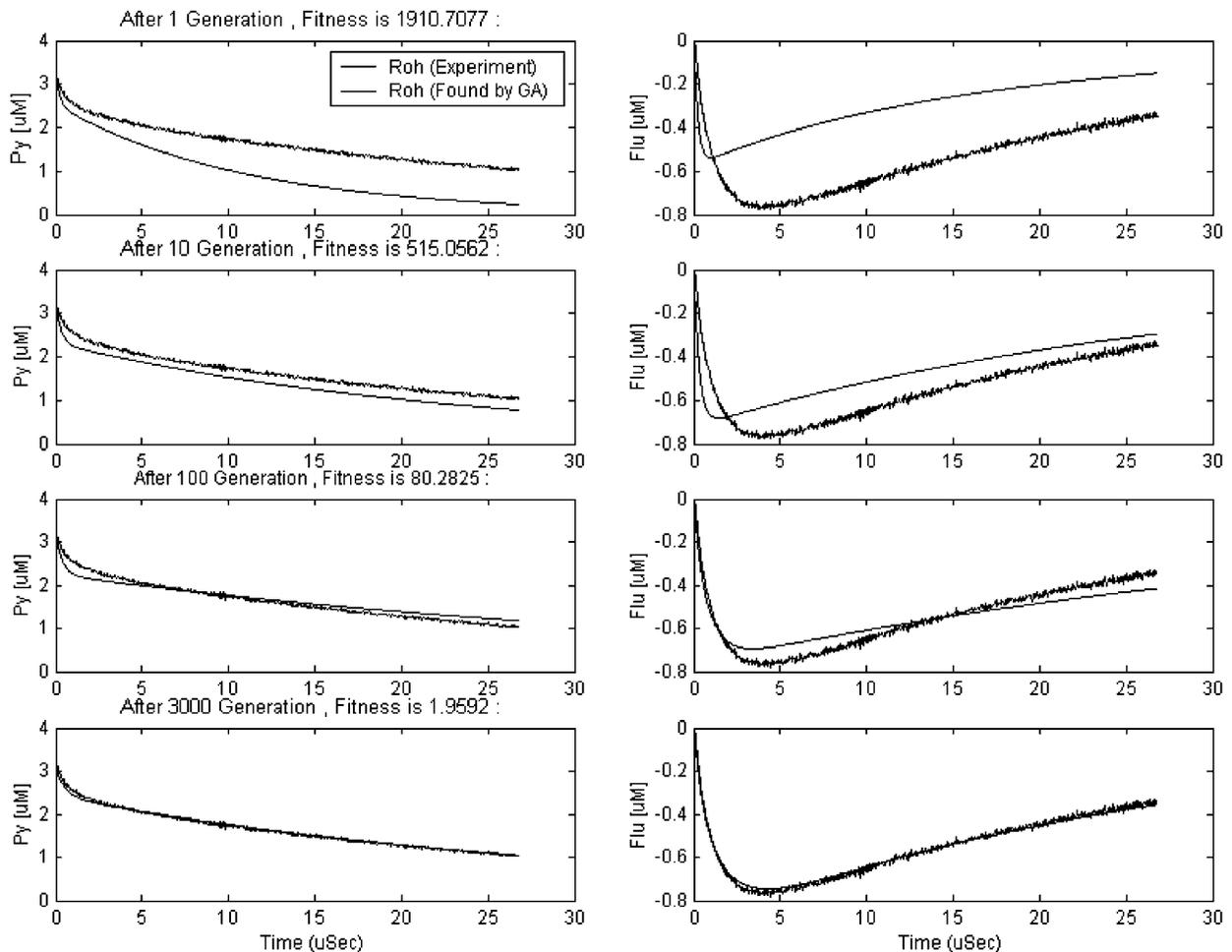


FIGURE 1 Snapshots of the reconstructed dynamics during the evolution of the system. The frames on the left side depict the experimental (*noisy*) and reconstructed (*smooth*) curves of the reprotonation of the pyranine. The frames on the right depict the reversible protonation of the fluorescein chromophore. The generation number and the values of the fitness function are indicated in the figure.

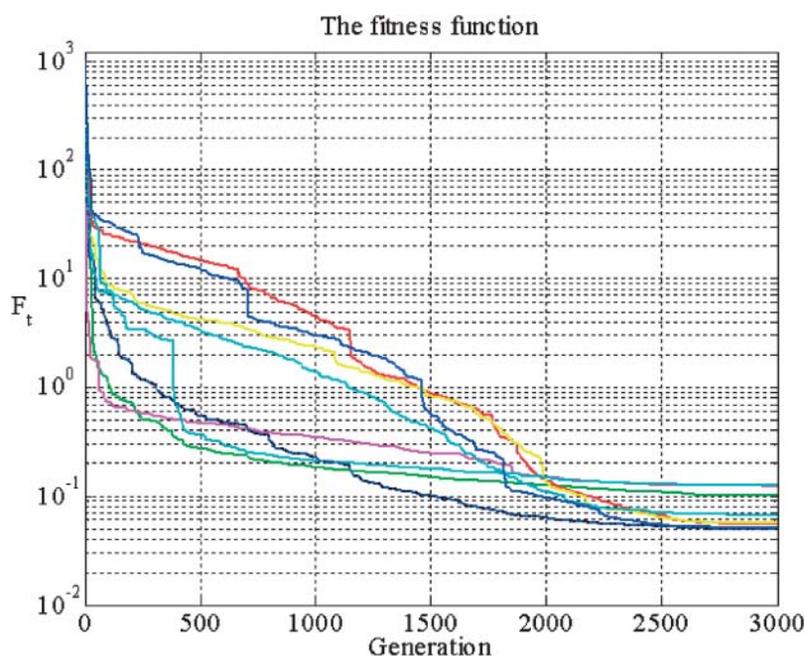


FIGURE 2 The evolution of the fitness function as calculated for eight consecutive simulations of the target signal presented in Fig. 1.

possible that more than one combination of adjustable parameters can yield curves that resemble the experimental signal to the same extent: i.e., there may be few local minima in the parameter space and each of them can reproduce a good macroscopic fit. For this reason, we tested to what extent the Genetic Algorithm can reproduce rate constants whose values are predetermined. Accordingly, a “benchmark signal” was generated by the differential rate equations using a known set of parameters all of which, except one (reaction 4 in Table 1), were determined by the “manual” analysis of experimental results. As a test for the accuracy of the analysis, the value of reaction 4 was deliberately set to be 100 times smaller than the actual rate constant, to verify whether the Genetic Algorithm is capable of determining a slow reaction in the presence of fast ones. These parameters were used to generate noiseless signals (one for the pyranine and one for the fluorescein) that were fed as a “target” for the Genetic Algorithm search. The analysis was repeated 10 times and the divergence of the various parameters from the initial value was calculated.

As shown in Table 1, the values reached by the Genetic Algorithm were within a few percent from the input values. It

is of interest to point out that when the rate constants are initially small, the relative error is large. The relative contribution of the pathways characterized by small rate constants to the overall flux is minor. Consequently, the random search procedure is less sensitive to these reactions and may compensate for their partial flux by small variation of other rate constants. This simple test demonstrates the robustness of the Genetic Algorithm; when the system is free of experimental noise, the calculation converges to values that generated the signal with minimal deviation. Apparently, the shape of the experimental curves bears sufficient information to force the system to find the global minimum in the multiparameter space.

### Analysis of experimental signals

In contrast to the zero-noise input signal, the experimental data are noisy and the convergence can, at most, fit into a band defined by the spread of the experimental points. It is possible that the inherent uncertainty, caused by the width of the target signal, will interfere with the convergence of the fitness function, and end up with more than one set of

**TABLE 1 Comparison between input values of rate constants used to generate signals and the average values derived by 10 independent runs of the Genetic Algorithm**

Reaction		Input	Output	Standard deviation
$\Phi\text{O}^- + \text{H}^+ \rightarrow \Phi\text{OH}$	$k_1$	$5.00\text{E} + 10$	$4.96\text{E} + 10$	$2.88\text{E} + 08$
$\text{FLU} + \text{H}^+ \rightarrow \text{FLUH}^+$	$k_2$	$3.00\text{E} + 10$	$3.01\text{E} + 10$	$7.81\text{E} + 07$
$\text{FLUH}^+ + \Phi\text{O}^- \rightarrow \text{FLU} + \Phi\text{OH}$	$k_4$	$1.00\text{E} + 07$	$1.83\text{E} + 07$	$0.83\text{E} + 07$
Intramolecular proton transfer	$k_5$	$2.00\text{E} + 11$	$1.95\text{E} + 11$	$1.40\text{E} + 09$
$\text{HCO}_3^- + \text{H}^+ \rightarrow \text{H}_2\text{CO}_3$	$k_6$	$2.50\text{E} + 10$	$2.53\text{E} + 10$	$9.40\text{E} + 07$
$\text{H}_2\text{CO}_3 + \Phi\text{O}^- \rightarrow \text{HCO}_3^- + \Phi\text{OH}$	$k_7$	$2.50\text{E} + 08$	$2.81\text{E} + 08$	$1.55\text{E} + 07$

parameters, each of them reconstructing curves that fit within the width of the input signals. The spread of the values can represent a single population, meaning that the convergence had reached a global minimum, which can be either well defined or shallow. Alternatively, the spread may be a combination of local minima, each represented by a different combination of values that all reproduce curves with the shape of the experimental observation. Naturally, the spreading of the results into a family of local minima undermines the utilization of the kinetics analysis of the experimental observations.

Fig. 3 depicts the variation of the adjustable parameters during eight independent runs corresponding to the convergence curves presented in Fig. 2. Although the signal has a relatively low noise level (see the experimental signal in Fig. 1) and the fitness function converged from  $10^3$  down to  $0.1 \pm 0.05$  (Fig. 2), the values of the adjustable parameters exhibit large variations. In two runs, the value of intramolecular proton transfer (Fig. 3 C) had drifted to the upper limit that is allowed for the rate constant, with no noticeable effect on the quality of the fit. Such an observation is consistent with the existence of more than one local minimum in the parameters' space.

### Simultaneous analysis of more than one pair of signals

The divergence of the solutions attained during the reconstruction of a single signal implies that the restrictions imposed on the system were not sufficiently limiting. For this reason, it was investigated whether a simultaneous reconstruction of two pairs of signals, each measured at a different initial condition, would be more restrictive,

forcing the system to converge into one minimum. The argument for this reasoning is that the rate constants are independent of the reactant concentrations; thus, when the results of two experiments that were measured at different reactant concentrations are analyzed simultaneously, the algorithm will search for those values that are concentration independent. The outcome of this strategy is represented in Fig. 4, which summarizes the analysis of two signals, one measured at pH = 6.8 and the other at pH = 7.3. The difference in the initial conditions seems rather small, but sufficient to change the concentrations of the reactants by approximately threefold and make the shape of the signals clearly distinguishable.

There are 10 frames in this composite figure; each presents the convergence of one adjustable parameter along the generation axis. In contrast to the pattern detected in Fig. 3, where some of the parameters drifted far away, in this case the restrictions imposed by the variance of the reactant concentrations were sufficient to enforce a better convergence.

We can classify the adjustable parameters into two categories. The first category comprises those associated with the directly observed reactants, the pyranine and the indicator. The concentration of these reactants is accurately measured and precisely recorded at any time. Accordingly, the convergence of the parameter associated with these reactants is very effective (Table 2).

The other category is that of reactants whose concentration is not predetermined and whose state of protonation is not observable. In this study, the only member of this category is the bicarbonate. When protein dynamics are analyzed, the number of unobserved reactants is larger (Nachliel et al., 2002; Nachliel, and Gutman, 2001). The convergence of the adjustable parameters of the compounds in this category

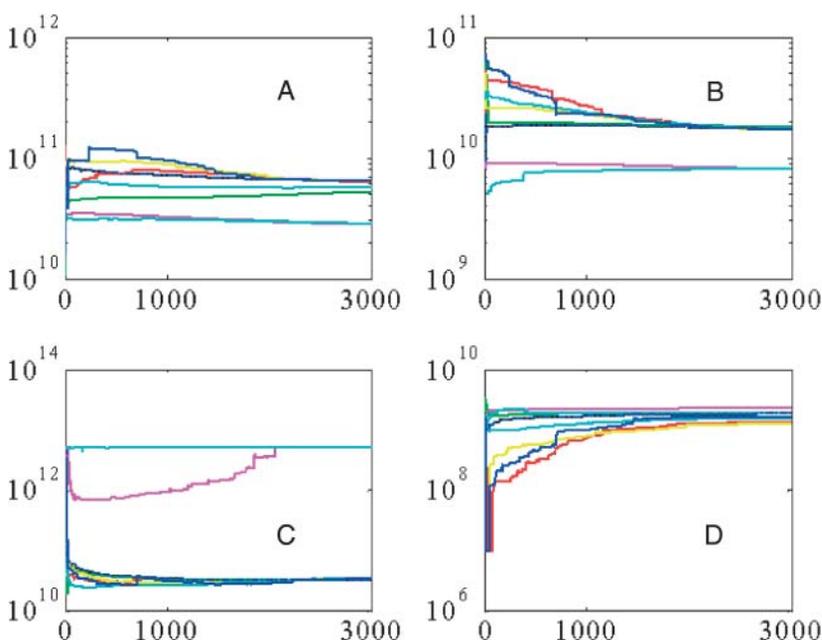


FIGURE 3 The spreading of the values of the adjustable parameters during the repeated analysis of a single pair of experimental signals, measured at pH = 6.8. Panels A–D depict the convergence of the rate constant for the protonation of the pyranine; the protonation of the fluorescein; the intramolecular proton transfer from the carboxylate residue of the fluorescein to the oxyanion; and the proton transfer from the fluorescein to the pyranine.

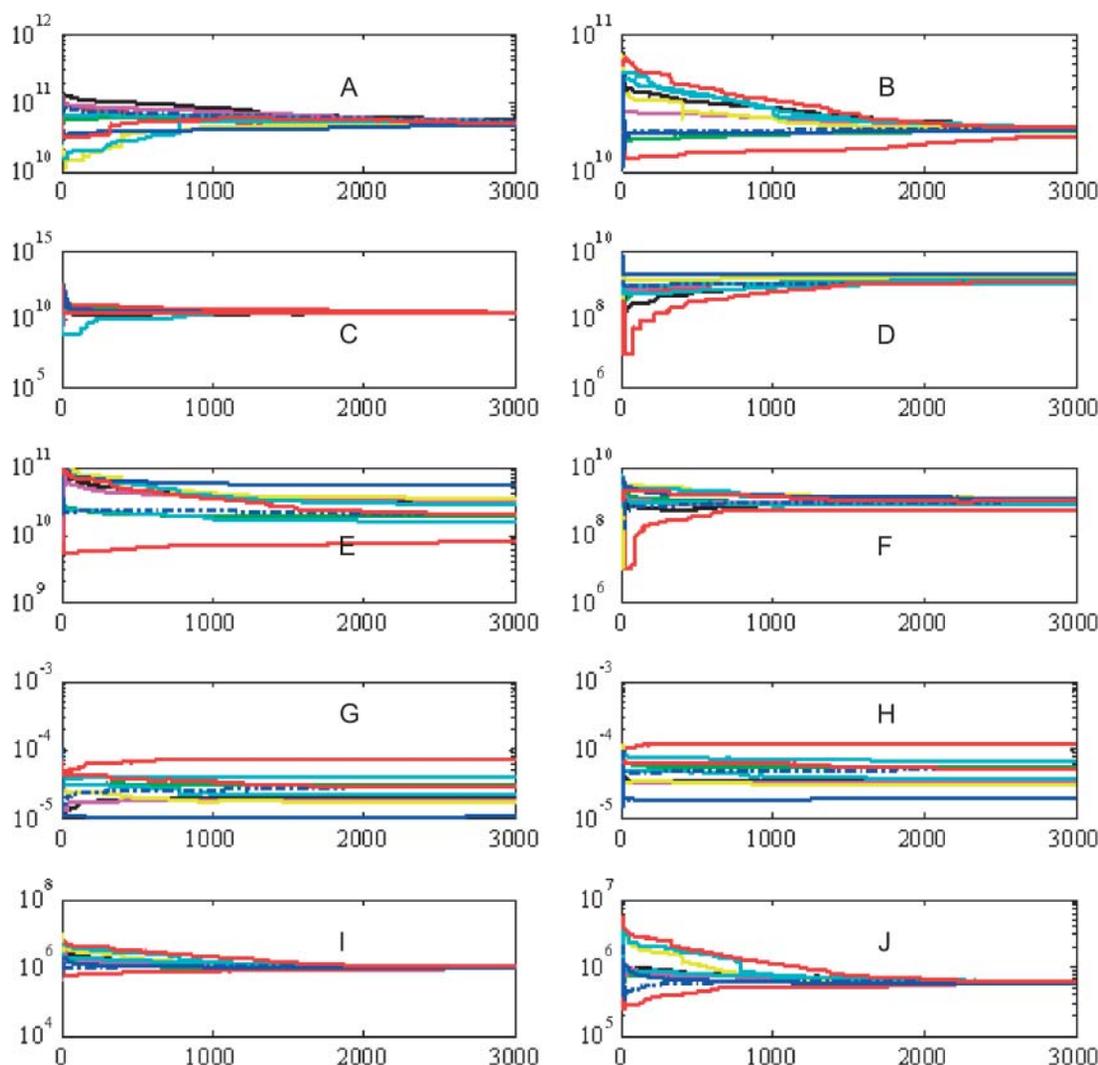


FIGURE 4 The convergence of the parameters during simultaneous analysis of a pair of signals measured at pH = 6.8 and 7.3. Each frame depicts the convergence of a single adjustable parameter: (A) protonation of pyranine; (B) protonation of fluorescein; (C) proton transfer from the carboxylate residue of the fluorescein to the oxyanion; (D) proton transfer from the fluorescein to the pyranine; (E) protonation of the bicarbonate; (F) proton transfer from carbonic acid to pyranine; (G and H) the bicarbonate concentration at pH = 6.8 and pH = 7.3, respectively; and (I and J) time constant of protonation of bicarbonate at pH = 6.8 and 7.3, respectively.

operates within a space having more degrees of freedom than compounds in the first category; the search is carried out both for the concentration and the rate constant of the reactant. As a result, the spread of both values, rate constants and concentrations, are somewhat higher (Table 2). The rate constant of protonation of the bicarbonate was  $2.6 \pm 1.3 \cdot 10^{10} \text{ M}^{-1} \text{ s}^{-1}$  and the concentration of the bicarbonate was  $28 \pm 18$  and  $50 \pm 28 \mu\text{M}$ , depending on the pH of the reaction mixture. In comparison with the other parameters, these values seem to be less accurately determined, yet a close examination of the rate constants reveals that for those runs where the rate constant was high, the derived concentrations of the bicarbonate were low. When the rate constant converged to a lower value, the concentration of the bicarbonate was high. On the other hand, the time constant for

the protonation of the bicarbonate converged to a constant, single value (see Fig. 4, I and J, and the bottom row of Table 2).

It should be stressed that the rate constants derived by the Genetic Algorithm confirm the values that had been determined by the manual search in the parameters' space (see last column in Table 2) (Gutman 1984; Gutman and Nachliel, 1990; Gutman et al., 1985, 2003).

### Statistical evaluation of the solution

The analysis by the Genetic Algorithm is a search process and even when a noiseless signal is recurrently analyzed, the parameters of the best-fitted phenotype of the different runs are not identical and are dispersed (see Table 1). In the case

**TABLE 2** Tabulations of the rate constants as derived from tandem analysis of experimental signals

Reaction	Mean rate constant $10^{10} \text{ M}^{-1}\text{s}^{-1}$	Standard deviation $10^{10} \text{ M}^{-1}\text{s}^{-1}$
$\text{H}^+ + \Phi\text{O}^-$	$k_1$ 5.2870 (5.5 $\pm$ 0.3)*	0.2936
$\text{H}^+ + \text{FLU}^\dagger$	$k_2$ 2.0399 (2.0 $\pm$ 0.2)	0.0983
$\text{FLUH}^+ + \Phi\text{O}^-$	$k_4$ 0.14225 (0.15 $\pm$ 0.05)	0.02676
Intramolecular proton transfer <sup>‡</sup>	$k_5$ 3.05 (3.5 $\pm$ 0.5)	0.3599
$\text{H}_2\text{CO}_3 + \Phi\text{O}^-$	$k_6$ 0.09805 (0.08 $\pm$ 0.03)	0.01826
$\text{H}^+ + \text{HCO}_3^-$	$k_7$ 2.6497 (2.5 $\pm$ 0.3)	1.2749
$\tau (\text{HCO}_3)^\S$	$8 \cdot 10^5 \text{ s}^{-1}$	$0.1 \cdot 10^5 \text{ s}^{-1}$

\*The numbers in parenthesis were determined by independent “manual” analysis of the signals.

<sup>†</sup>The same rate constant applies both for the protonation of the oxyanion of the xanthene rings and the carboxylate of the benzene ring of the fluorescein.

<sup>‡</sup>The rate constant for the intramolecular proton transfer between the carboxylate of the benzene ring to the oxyanion on the xanthene rings of the fluorescein.

<sup>§</sup>The time constant for the reaction between the bicarbonate and the free proton. Calculated for each run as the product of the rate constant times the bicarbonate concentration.

that the target for the fitting is noisy, we needed two signals for tandem analysis. Accordingly the uniqueness of the solution should be derived from statistical analysis.

The normal probability plot (QQplot program of Matlab) is suitable for this kind of analysis. This program displays

a plot of the values of the given parameter (on the y axis) versus theoretical normal distribution of the values as calculated from the whole population. If the distribution is normal, the plot will be close to linear.

The analysis was carried out for all adjustable parameters used for the analysis of the experimental signal (Figs. 5 and 6). Fig. 5 depicts the results of the analysis where the rate constants are the analyzed values. As seen in the figure, the values assigned for each of the rate constants appear to belong to a single normal population. Thus, even if the certainty is not high, and the standard deviation is  $\sim 20\%$  of the mean value, the results are still members of one normal population.

In Fig. 6 we analyzed the distribution of the bicarbonate concentration, a term that we found too difficult to control experimentally and decided to add it as another adjustable parameter. The upper two frames depict the QQplot analysis of the values of bicarbonate concentrations as calculated for the experiments measured in the high (*left*) and low (*right*) pH values. Once again, even though the variance is high, from 10 up to  $\sim 100 \mu\text{M}$ , the values are all members of a single normal population. To better ascertain this conclusion, we also analyzed the time constant of protonation of the bicarbonate, namely the product of the rate constant times the bicarbonate concentration. The lower frames in Fig. 6 confirm that the time constants are also members of a normal population. Thus, each of the

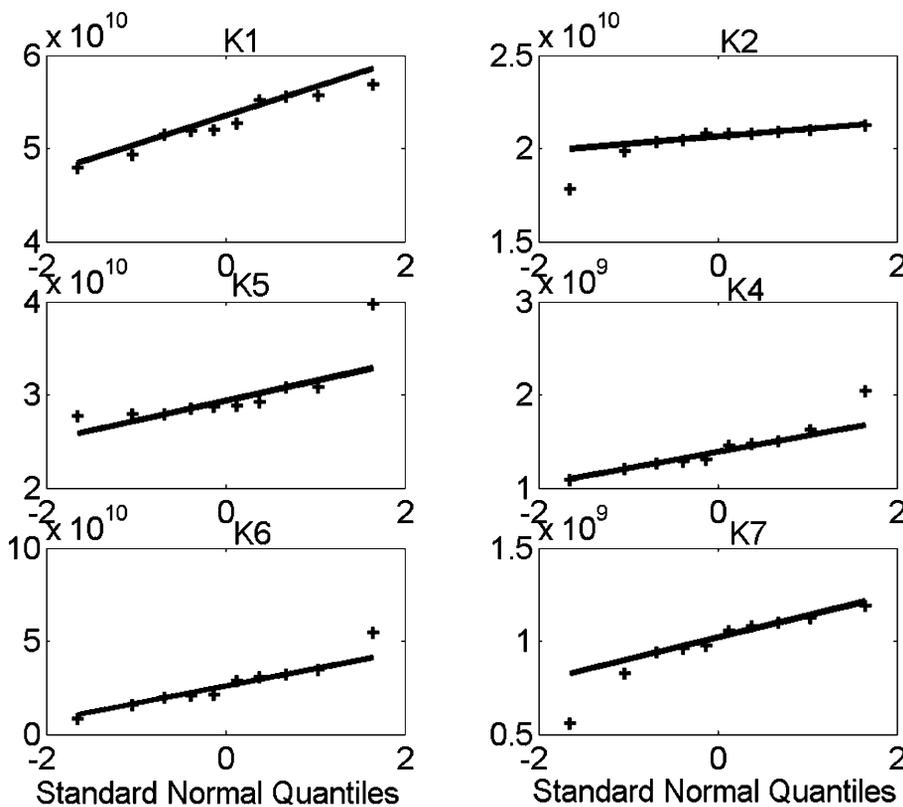


FIGURE 5 The normal probability plot, as calculated by the QQplot program of Matlab, for the rate constants as indicated in the figure. Ordinate, the values of the adjustable parameter; abscissa, the standard normal quantile of the plotted parameter.

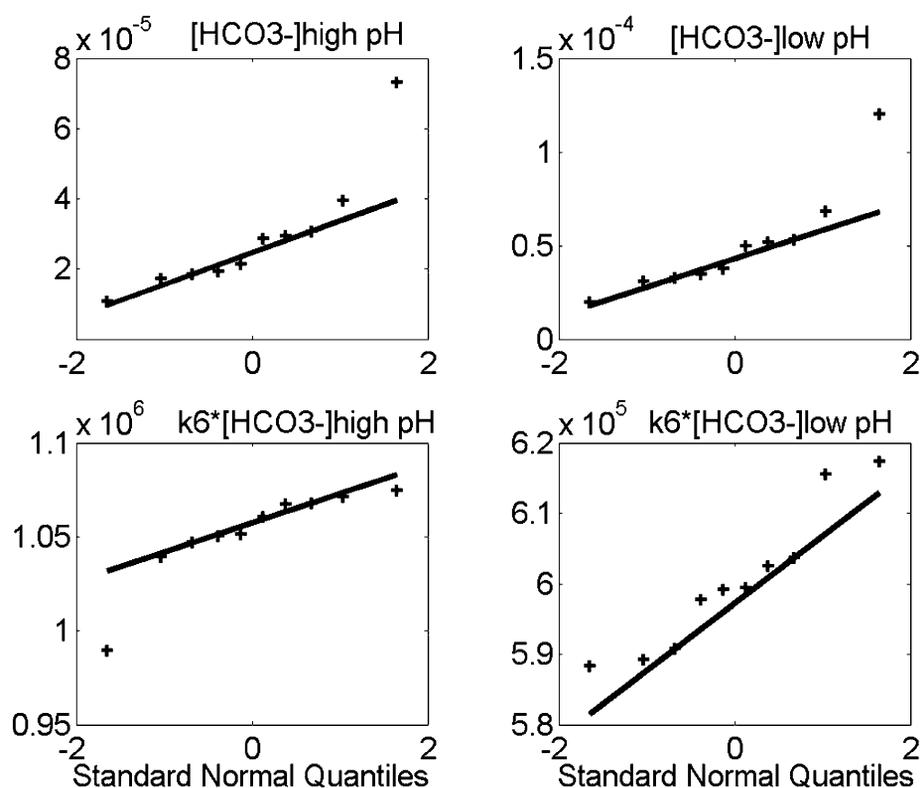


FIGURE 6 The normal probability plot for the concentrations of the bicarbonate as determined for the two experimental systems. The upper panels present the plot as calculated for the concentrations of the bicarbonate and the lower panels present the function as calculated for the time constants (rate constant times concentration) of the protonation of the bicarbonate.

adjustable parameters needed to reconstruct the observation is the only minimum along its axis, and the solution itself represents the global minimum in the multidimensional parameter space.

## CONCLUDING REMARKS

In this study we have shown that the Genetic Algorithm is an applicable mathematical tool to search for a solution of kinetics equations where the rate constants are to be determined. The analysis demonstrated that even when the concentration of one reactant is unknown, the system had successfully converged. The experimental uncertainty in the measured condition, as typical for real biochemistry, necessitated solving two signals at tandem. Under these conditions, a complex system consisting of eight adjustable parameters was successfully solved. The statistical analysis of the rate constants indicated that the spread of the values for each of the adjustable parameters is consistent with the normal distribution of a single population. Thus complex dynamics systems, where many processes progress in parallel, can be solved with the certainty that the parameters represent the global minimum in the multidimensional parameter space.

Finally, we wish to stress that even though this study is an empirical search for the uniqueness, it is versatile and readily applicable for any complex kinetic system, thus it should be

considered as a future standard test, incorporated in kinetic analysis of multireactions systems.

This research is supported by the German-Israeli Foundation for Scientific Research and Development (grant No. I-140-207.98) and the Israel Science Foundation (grant No. 427/01-1).

## REFERENCES

- Checover, S., Y. Marantz, E. Nachliel, M. Gutman, M. Pfeiffer, J. Tittor, D. Oesterhelt, and N. A. Dencher. 2001. Dynamics of the proton transfer reaction on the cytoplasmic surface of bacteriorhodopsin. *Biochemistry*. 40:4281–4292.
- Checover, S., E. Nachliel, N. A. Dencher, and M. Gutman. 1997. Mechanism of proton entry into the cytoplasmic section of the proton-conducting channel of bacteriorhodopsin. *Biochemistry*. 36:13919–13928.
- Eigen, M. 1964. Proton transfer, acid-base catalysis, and enzyme hydrolysis. *Angew Chem. Int. Ed. Engl.* 3:1–19.
- Filipic, B., I. Zun, and M. Perpar. 2000. Skill-based interpretation of noisy probe signals enhanced with a genetic algorithm. *Int. J. Hum. Comput. Stud.* 53:517–535.
- Gutman, M. 1984. The pH jump: probing of macromolecules and solutions by a laser-induced, ultrashort proton pulse-theory and applications in biochemistry. *Methods Biochem. Anal.* 30:1–103.
- Gutman, M., and D. Huppert. 1979. Rapid pH and  $\Delta\mu_{\text{H}^+}$  jump by short laser pulse. *J. Biochem. Biophys. Methods*. 1:9–19.
- Gutman, M., and E. Nachliel. 1990. The dynamic aspects of proton transfer processes. *Biochim. Biophys. Acta.* 1015:391–414.
- Gutman, M., and E. Nachliel. 1997. Time-resolved dynamics of proton transfer in proteinous systems. *Annu. Rev. Phys. Chem.* 48:329–356.

- Gutman, M., E. Nachliel, and E. Gershon. 1985. Effect of buffer on kinetics of proton equilibration with a protonable group. *Biochemistry*. 24:2937–2941.
- Holland, J. H. 1970. Robust algorithms for adaptation set in a general formal framework. *Proceedings of the 1970 IEEE Symposium on Adaptive Processes Decision and Control: XVII*. ACM PRESS, New York.
- Hongqing, C., Y. Jingxian, K. Lishan, C. Yuping, and C. Yongyan. 1999. The kinetic evolutionary modeling of complex systems of chemical reactions. *Comput. Chem.* 23:143–151.
- Houck, C. J. J., and M. A. Kay. 1995. Genetic Algorithm for Function Optimization: A Matlab Implementation NCSU-IE TR 95–09.
- Leardi, R. 2001. Genetic algorithm in chemometrics and chemistry: a review. *ET J.* 15:559–569.
- Marantz, Y., E. Nachliel, and M. Gutman. 2001. Proton-collecting properties of bovine heart cytochrome C oxidase: kinetic and electrostatic analysis. *Biochemistry*. 40:15086–15097.
- Gutman, M., E. Nachliel, A. Mezer, and O. Noivirt. 2003. Gauging of local micro-environment at protein water interface by time-resolved single-proton transfer reactions. *Annals of the European Academy of Sciences*. 1:75–107.
- Nachliel, E., M. Gutman, S. Kiryati, and N. A. Dencher. 1996. Protonation dynamics of the extracellular and cytoplasmic surface of bacteriorhodopsin in the purple membrane. *Proc. Natl. Acad. Sci. USA*. 93:10747–10752.
- Nachliel, E., M. Gutman, J. Tittor, and D. Oesterhelt. 2002. Proton transfer dynamics on the surface of the late M state of bacteriorhodopsin. *Biophys. J.* 83:416–426.
- Nachliel, E., and M. Gutman. 2001. Probing the substrate binding domain of lactose permease by a proton pulse. *Biochim. Biophys. Acta*. 1514: 33–50.
- Robinson, R. A., and R. H. Stock. 1959. *Electrolyte Solutions*. Butterworths Publications Ltd., London, UK.
- Shimoni, E., E. Nachliel, and M. Gutman. 1993. Gaugement of the inner space of the apomyoglobin's heme binding site by a single free diffusing proton. II. Interaction with a bulk proton. *Biophys. J.* 64:480–483.
- Terry, D. B., and M. Messina. 1998. Heuristic search algorithms for the determination of rate constants and reaction mechanisms from limited concentration data. *J. Chem. Inf. Comput. Sci.* 38:1232–1238.
- Viappiani, C., G. Bonetti, M. Carelli, F. Ferrati, and A. Sternieni. 1998. Study of proton transfer processes in solutions using the laser induced pH jump: a new experimental setup and an improved data analysis based on genetic algorithms. *Rev. Sci. Instrum.* 69:270–276.
- Yam, R., E. Nachliel, and M. Gutman. 1988. Time-resolved proton-protein interaction. Methodology and kinetic analysis. *J. Am. Chem. Soc.* 110: 2636–2640.