JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Hybrid Genetic Algorithm with an Adaptive Penalty Function for Fitting Multimodal Experimental Data: Application to Exchange-Coupled Non-Kramers Binuclear Iron Active Sites

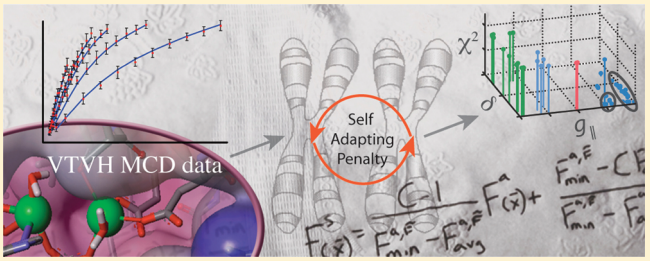
Eric Beaser,[†] Jennifer K. Schwartz,[‡] Caleb B. Bell, III.,[‡] and Edward I. Solomon[‡,*]

[†]Santa Clara, California, United States

[‡]Department of Chemistry, Stanford University, 333 Campus Drive, Stanford, California 94305, United States

Supporting Information

ABSTRACT: A Genetic Algorithm (GA) is a stochastic optimization technique based on the mechanisms of biological evolution. These algorithms have been successfully applied in many fields to solve a variety of complex nonlinear problems. While they have been used with some success in chemical problems such as fitting spectroscopic and kinetic data, many have avoided their use due to the unconstrained nature of the fitting process. In engineering, this problem is now being addressed through incorporation of adaptive penalty functions, but their transfer to other fields has been slow. This study updates the Nanakorrn Adaptive Penalty function theory, expanding its validity beyond maximization problems to minimization as well. The expanded theory, using a hybrid genetic algorithm with an adaptive penalty function, was applied to analyze variable temperature variable field magnetic circular dichroism (VTVH MCD) spectroscopic data collected on exchange coupled Fe(II)Fe(II) enzyme active sites. The data obtained are described by a complex nonlinear multimodal solution space with at least 6 to 13 interdependent variables and are costly to search efficiently. The use of the hybrid GA is shown to improve the probability of detecting the global optimum. It also provides large gains in computational and user efficiency. This method allows a full search of a multimodal solution space, greatly improving the quality and confidence in the final solution obtained, and can be applied to other complex systems such as fitting of other spectroscopic or kinetics data.



INTRODUCTION

Experimental data fitting problems that are qualitatively linear or polynomial in character and that contain a modest number of design variables have long been solved by algorithms such as least-squares or gradient-based methods.[1,2] Yet, traditional techniques fail when discrete variables or highly nonlinear objective functions are involved.[3,4] These type of constraints (such as those found in complex spectroscopic data fitting) often lead to complicated multimodal or noncontinuous systems and require a different approach.

The need for fast, global solvers is motivated by three additional concerns. Foremost are the constraint conditions, imposed either by theory or based on expert knowledge, which complicate the feasible domain. Second is the time required to perform an exhaustive search of a large multivariate solution space using local area solvers. If the system is truly multimodal, i.e., a design space with many local minima, the user would be required to run a polynomial solver an unrealistic number of times to test and compare the results of various initial conditions. This drives the final motivation. Even after a manual multipoint search has been performed, there is no guarantee that the global optimum has been found.

For these reasons, this study examines a hybrid technique combining a constrained genetic algorithm with a sequential quadratic programming[5] model to efficiently explore a global multimodal solution space. This hybrid approach quickly finds the optimal data fit within the complex construct of a noncontinuous objective function, design variables, and their constraints.

Genetic Algorithms (GA) are ideal as they are insensitive to discrete variables, nonlinearity, and, most importantly, multimodal solution spaces.[4,6−9] In this study, the GA method has been coupled with an adaptive penalty function,[10−13] allowing it to be constrained in an application insensitive way. Further, once the GA is near the vicinity of the global minimum, the hybrid solver transitions to a gradient-based optimization approach to directly drive to the optima.

I. DESIGN OF AN ADAPTIVE, HYBRID GENETIC ALGORITHM

Used for a number of decades as a optimization technique, Genetic Algorithms are based on natural selection in biological systems which is reflected in both its name and the mathematical

Received: March 18, 2011
Published: August 08, 2011

terminology describing this method.[14] Stochastic processes, such as natural selection, chromosomal crossover, and genetic mutation, allow the algorithm to search the entire solution space, gradually improving the fitness of the population. Eventually, the Genetic Algorithm converges on an optimal individual whose genome best solves the objective function.

Constrained Optimization. Since Genetic Algorithms are by their nature unconstrained,[9,14] the algorithm must be modified to evolve erroneous genomes toward the feasible solution space.

The constrained optimization problem to fit experimental data to an analytical solution is

$$\begin{aligned} &\text{min}imize \\ &F(\bar{x}) = F[f(\bar{x})] \\ &where \\ &\bar{x} = (x_1, x_2, ..., x_N) \in \mathscr{R}^N \end{aligned} \tag{1}$$

under the constraints defined as

$$\begin{aligned} &g_i(\bar{x}) \le 0, i = 1, ...K \\ &h_i(\bar{x}) = 0, i = 1, ...P \end{aligned} \tag{2}$$

In order to constrain the design variables, GA penalty functions have traditionally been employed to modify the fitness value of each individual. When a solution violates one or more constraints and becomes "infeasible," its fitness is lowered by a certain amount, $P(x)$

$$\begin{aligned} F^a(\bar{x}) &= F(\bar{x}) \qquad if \bar{x} \in \tilde{F} \\ F^a(\bar{x}) &= F(\bar{x}) - P(\bar{x}) \quad otherwise \end{aligned} \tag{3}$$

where $\tilde{F}$ is the feasible search space, F^a is the augmented fitness value, and $P(x)$ is the penalty assigned for the constraint violation.

Penalty functions have been generally applied through two different approaches. The vast majority of these approaches reduce the individual's fitness in proportion to the degree of constraint violation, with each violation adding to the total penalty assessed

$$\begin{aligned} &P(\bar{x}) = \sum_{j=1}^{K} (\lambda_G)[G_j(\bar{x})]^\beta + \sum_{j=1}^{P} (\lambda_H)[H_j(\bar{x})]^\beta \\ &where \\ &G_J(\bar{x}) = \max[0, g_J(\bar{x})] \\ &H_J(\bar{x}) = abs[h_j(\bar{x})] \end{aligned} \tag{4}$$

Here λ_G, λ_H, and β are nonphysical, a priori constraints that can only be determined through trial and error. Depending on the situation, the penalty functions can be proportional, exponential, or even logarithmic.

The problem with these approaches are the collection of nondimensional parameters that must be assigned prior to the start of the optimization run. Unfortunately, these nonphysical constants have no relation to the problem at hand. They must be optimized for each objective function and data set under study. This is a time-consuming and inexact process.[14,15] In addition, these methods do not always balance the benefit of allowing slightly infeasible individuals to remain within the population while allowing the remaining good genes to propagate forward. A penalty function that is too strict will unduly penalize an individual without consideration for its potential promise.

Within the population there may be key elements, "genes," required to evolve to the global optimum. Moreover, these superior genes could be spread among a number of different individuals. To overly penalize an infeasible solution early in the evolutionary history for a slight constraint violation is unwise; a key feasible gene in an overall "infeasible" genome could be mistakenly discarded.

In this study, the genes of feasible and infeasible individuals are instead allowed to mix over the generations. The penalty function evolves over time, slowly increasing the penalty against the infeasible solutions while simultaneously encouraging potentially key genes to spread, even if their overall genome is slightly in violation.

Adaptive Penalty Functions. This study applies an adaptive penalty function originally developed by Nanakorn et al.[13] to solve structural design problems. The use of an adaptive penalty function has two key advantages: 1) it adjusts the required level of penalty in each generation after assessing the fitness of the entire population; 2) the penalty coefficients have clear, physical meanings that do not need to be varied greatly between data sets.[15]

At its core, the algorithm focuses on modifying each individual's raw fitness value each generation. The scheme is as follows

$$F_i^a = F(\bar{x}_i) - P(\bar{x}_i) = F(\bar{x}_i) - \lambda(t)E(\bar{x}_i) \tag{5}$$

where $F(x)$ is the raw fitness value of the individual at generation i, F^a is the fitness of the individual after the penalty. P is the penalty factor by which to adjust the individual's fitness value. $E(x)$ is each individual's error value, which is the sum of their equality and inequality constraint violations. Finally, $\lambda(t)$ is the penalty factor that varies for each generation. In the case of data fitting problems, the raw fitness value is the "goodness of fit" parameter, comparing the closeness of the analytical fit generated from the individual's genome to the experimental data.

To ensure that the optimal solution will have a higher fitness value than all others, both feasible and infeasible, the following condition was imposed

$$F^a(\bar{x}_i) \le \phi(t) F_{avg}^{\bar{F}} \; for \; \forall \; \bar{x}_i \in \bar{U} \tag{6}$$

where U is the infeasible search space, F_{avg} is the average fitness of all feasible individuals, and $\phi(t)$ is a factor of F_{avg} used to adjust the total penalty applied in a given generation. From this, the maximum fitness value of any infeasible individual is set to no greater than $\phi(t)F_{avg}^{\bar{F}}$. Using this principle, the value of $\lambda(t)$ is computed for each generation. Nanakorn[13] then defined $\lambda(t)$ as

$$\lambda(t) = \max\left(0, \max_{\forall x_i \in \bar{U}} \left(\frac{F(x_i) - \phi(t)F_{avg}^{\bar{F}}}{E(x_i)}\right)\right) \tag{7}$$

To scale the fitness values, a modified bilinear scaling approach was used. The rules are as follows: The minimum scaled fitness is equal to zero to avoid negative, nonphysical values. Looking across the population, the average fitness of all feasible individuals is set to 1. The maximum scaled fitness of the best individual is set to "C" times that of the average fitness of the feasible individuals. As shown in eq 6, the probability that an infeasible member is selected to become a parent of the next generation is set to a user defined value (ϕ) times that of the average fitness of the feasible individuals.

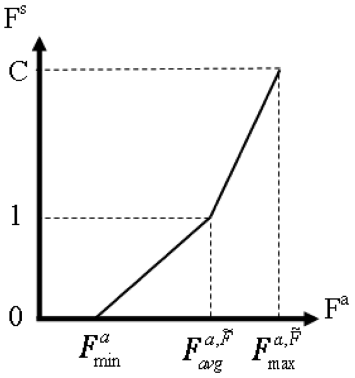


Figure 1. Bilinear fitness scaling for maximization problems with at least one feasible individual.

The user is therefore presented with two a priori gains that are detached from the problem being solved. The constants can be set based on a limited amount of experience using genetic algorithms and do not have to be adjusted for each data set. This, combined with the generation-dependent, gene mutation range described in the Crossover and Mutation section promotes the cross-pollination of genes among many individuals, coaxing out the optimal gene even from slightly infeasible members.

Expansion of Theory for Adaptive Penalty Functions to Treat Minimization Problems. For maximization problems, the theory proposed by Nanakorn et al.[13] is complete. However, the equations cannot be used for minimization in their present form. Typically, to convert from one goal type to the other, the signed value of the fitness function is multiplied by −1. Once done, an optimization routine whose goal is to minimize the fitness value will then search for the smallest negative value, i.e., the maximum score. While the majority of optimization schemes are reversible between minimization and maximization problems, there are two key elements in the Nanakorn[13] theory that prevent this. The first reversibility block is eq 7, $\lambda(t)$. The second is the bilinear scaling approach. Both approaches require that larger scores equate to greater fitness, preventing simple reversal.

The following modifications to adaptive penalty functions used by Nanakorn[13] are proposed to expand the theory to minimization problems:

The conditional eq 6 is modified to

$$F^a(\vec{x}_i) \geq \phi(t)F^{\bar{F}}_{avg} \text{ for } \forall\ \vec{x}_i \in \bar{U} \tag{8}$$

This sets the best fitness of the infeasible individuals to $\phi(t)F^{\bar{F}}_{avg}$. In contrast to the Nanakorn[13] approach of defining the optimal score as the maximum attainable, eq 8 defines the best infeasible fitness as the lowest achieved.

Equation 8 flows into the expression for $\lambda(t)$ in eq 7. To select the best infeasible individual as the true originator of $\lambda(t)$, eq 7 is modified as follows

$$\lambda(t) = \min\left(0, \min_{\forall x_i \in \bar{U}}\left(\frac{F(x_i) - \phi(t)F^{\bar{F}}_{avg}}{E(x_i)}\right)\right) \tag{9}$$

It can be seen that due to the comparison between zero and the inner product in eq 7, a reversed fitness function would not allow $\lambda(t)$ to be computed properly without the changes made in eq 9.

The bilinear fitness scaling model must also be updated. This function converts each individual's fitness into an expectation for selection into the next generation. Here again the original theory assumed that the best score is the fitness with the greatest magnitude, as seen in Figure 1. To switch to minimization problems, this paper adopts and modifies the method proposed by Nanakorn[13] to solve for scaled fitness where no individuals are feasible (ref 13 eq 19).

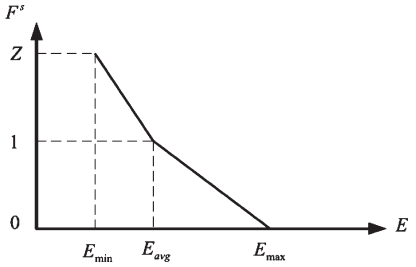


Figure 2. Bilinear fitness scaling for the case where no members are feasible.

When no feasible individuals exist, the population is each given a fitness value equal to the magnitude of their error. This was done to force the population toward feasibility. As seen in Figure 2, the individual with the lowest error was set to the maximum selection expectation of "Z", and the average error was set to a value of one. The individual with the maximum error that generation was assigned a zero expectation of being selected into the next. In the all-infeasible case, solving for $\lambda(t)$ and $\phi(t)$ is not required.

Adapting Figure 2 for minimization problems when there is at least one feasible individual, the lowest (best) feasible fitness is given a selection expectation of C and the average feasible fitness is given an expectation of one. The individual with the largest (worst) fitness after penalties are assessed is given a zero selection expectation. Modifying Nankorn's eq 19 to match Figure 3, the bilinear fitness scaling equations for minimization problems with at least one feasible member are

$$F^S(\vec{x}) = \frac{C-1}{F^{a,\tilde{F}}_{\min} - F^{a,\tilde{F}}_{avg}}F^a(\vec{x}) + \frac{F^{a,\tilde{F}}_{\min} - CF^{a,\tilde{F}}_{avg}}{F^{a,\tilde{F}}_{\min} - F^{a,\tilde{F}}_{avg}} \tag{10}$$

$$\text{where } F^a(\vec{x}) \leq F^{a,\tilde{F}}_{avg}$$

$$F^S(\vec{x}) = \frac{1}{F^{a,\tilde{F}}_{avg} - F^a_{\max}}F^a(\vec{x}) + \frac{F^a_{\max}}{F^a_{\max} - F^{a,\tilde{F}}_{avg}}$$

$$\text{where } F^a(\vec{x}) > F^{a,\tilde{F}}_{avg}$$

$$\text{where}$$

$$F^{a,\tilde{F}}_{\min} = \min(F^{a,\tilde{F}})$$

$$F^a_{\max} = \max(F^a)$$

$$F^{a,\tilde{F}}_{avg} = avg(F^{a,\tilde{F}})$$

In the case where no feasible individuals exist, this study uses the original theory illustrated in Figure 2. Together, eqs 8, 9, and 10 extend the Nanakorn Adaptive Penalty Function theory,[13] allowing the technique to be applied to minimization optimization problems.

Hybrid Method: Integration of GA and Gradient-Based Methods. In direct comparison, the quasi-random walk[14] Genetic Algorithm will require more function calls and therefore

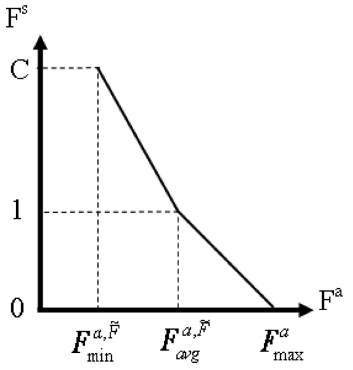


Figure 3. Bilinear fitness scaling for minimization problems with at least one feasible member.

time than an algorithm guided by the fitness derivatives in finding the nearest local minimum.[4,16]

Therefore, once our Genetic Algorithm converges on the region surrounding a potential global minimum, the best genome found to date is transferred to the Sequential Quadratic Programming[17] (SQP) solver. This gradient-based, nonlinear, constrained optimization routine employs a Quadratic Programming Model for the major step direction and a third order polynomial adaptation to Brent's method for the 1D line search.[5] It assumes the multivariate fitness function is quadratic and bases the direction of the next step for the genome (now called design vector) on the locally approximated gradient and Hessian. Starting with the best individual found by the GA, and assuming the best solution will now be found within a small continuous trust region near its starting location, the SQP quickly finds the local minimum within the N-dimensional valley in which it starts, enabling rapid convergence to an optimized solution.

Unfortunately, not all design spaces will have one distinctly optimal solution. If the solution space is extremely multimodal or the solution lies on a constraint boundary, there is value in running the GA→SQP two-step process a number of times. This allows multiple GA initial populations to be generated and be evaluated Monte Carlo style. The resulting set of evolved and potentially globally optimal solutions are reviewed later by the user who remains the true expert in analyzing the complex system. An intuitive graphical method was used in this study to allow selection between the resulting potential solutions.

II. APPLICATION OF THE ADAPTIVE, HYBRID GENETIC ALGORITHM TO A COMPLEX PHYSICAL CHEMISTRY PROBLEM

The appeal of this application insensitive, adaptively constrained hybrid method is it can efficiently fit highly complex chemical problems. This method has been applied to a problem of specific interest, the fitting of Variable Temperature, Variable Field (VTVH) Magnetic Circular Dichroism (MCD) spectroscopic data collected on magnetically coupled, binuclear non-Kramers Fe(II) systems. These systems have four elements of complexity: they are nonlinear, noncontinuous, partially integer based, and are highly multimodal.

VTVH MCD Data Fitting for Non-Kramers Systems. Binuclear nonheme iron enzymes are found in a variety of biological systems and are used to catalyze a number of critical reactions, including H-atom abstraction, desaturation, electrophilic aromatic substitution, ferroxidation, oxidative ring cleavage, and hydroxylation.[18] Insight into the geometric and electronic structure as well as the mechanisms of these metalloenzymes could allow design of more efficient catalysts, utilization of the enzymes in alternative applications such as bioremediation, and a better understanding of diseases in systems where these enzymes malfunction. The active reduced form of these systems, a coupled Fe(II)Fe(II) non-Kramers (integer spin) system, has proven very difficult to study due to its inaccessibility to normally applied spectroscopic methods and the complicated nature of its electronic structure. Fortunately, a methodology using circular dichroism (CD) and MCD has been developed that allows us to probe the electronic and geometric structure of the individual iron sites and the coupling between the iron centers due to bridging ligands. This methodology has been described in detail in other references.[18,19]

The Objective Function. Saturation magnetization data are collected by monitoring the MCD intensity at a representative, fixed wavelength over a range of temperatures (typically 2 to 25 K) and fields (typically 0 to 7 T). The collected VTVH MCD data must then be fitted to the following analytical equations

$$\Delta\varepsilon = \sum_i \left[(A_{satlim})_i \left(\int_0^{\pi/2} \frac{\cos^2\theta \sin\theta}{\Gamma_i} g_{||i}\beta H\alpha_i d\theta - \sqrt{2}\frac{M_z}{M_{xy}} \int_0^{\pi/2} \frac{\sin^3\theta}{\Gamma_i} g_{\perp i}\beta H\alpha_i d\theta \right) + B_i H\gamma_i \right] \quad (11)$$

where

$$\Gamma_i = \sqrt{\delta_i^2 + (g_{||i}\beta H\cos\theta)^2 + (g_{\perp i}\beta H\sin\theta)^2}$$

$$\alpha_i = \frac{e^{-(E_i - \Gamma_i/2)/kT} - e^{-(E_i + \Gamma_i/2)/kT}}{\sum_j e^{-(E_j - \Gamma_j/2)/kT} + e^{-(E_j + \Gamma_j/2)/kT}}$$

$$\gamma_i = \frac{e^{-(E_i - \delta_i/2)/kT} + e^{-(E_i + \delta_i/2)/kT}}{\sum_j e^{-(E_j - \beta_j/2)/kT} + e^{-(E_j + \delta_j/2)/kT}}$$

For each transition monitored, the relative intensities of the MCD data are plotted against $\beta H/2kT$ as shown in Figure 4, with each curve (isotherm) taken at a different temperature. In a non-Kramers system, the low temperature isotherms are offset from those at higher temperatures resulting in a spreading or 'nesting' of the curves. This nesting behavior is due to the rhombic zero-field splitting (δ) of the non-Kramers doublet ground state which arises from nonlinear field-induced mixing between the doublet sublevels of the coupled binuclear system.[20] The ground state parameters for each doublet ($g_{||}$, δ, A_{satlim}, B_i, $M_z/M_{x,y}$, $g_\perp$) and the energy splitting between doublets, E_i, determine the rate of saturation and degree of nesting observed in the VTVH MCD data. Consequently, the MCD intensity can be calculated by modeling the electronic structure of the system as a series of ground and excited state doublets or singlets, as described for non-Kramers systems in eq 11.[18−22]

This equation allows for the contribution of C-term intensity, $(A_{satlim})_i$, and the effects of a linear B-term, B_i, from field induced mixing between states and the presence of thermally excited sublevels of the ground state. E_i is the energy of the i-th excited state, and the energy of the ground state is defined as zero. The Boltzmann population over all states has been included in both the C-term and the B-term intensities as the factors α_i and γ_i,

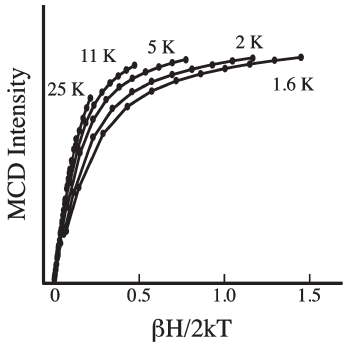


Figure 4. Relative intensity of VTVH MCD data plotted against $\beta H/2kT$. Each curve represents a different isotherm, taken from 1.6 to 25 K.

respectively. H is the applied magnetic field, k is the Boltzmann constant, and M_z and $M_{x,y}$ are the transition dipole moments for the directions indicated. During the experiment, $\Delta\varepsilon$, H, and T are measured; however, most other parameters are generally unknown.

If the energy spacing between doublets is large (>35 cm^{-1}), then an excited state doublet is not likely to be significantly populated before 25 K, the maximum temperature in the experiment, and the system has six unknown parameters ($g_\parallel$, δ, A_{satlim}, B_i, $M_z/M_{x,y}$, $g_\perp$). As multiple excited states are populated, the quantity of fit parameters grows. For example, there are 20 unknowns for the three state doublet, doublet, doublet model (D, D, D).

To make the objective function into a single variable optimization problem, the Goodness of Fit (GoF) parameter, χ^2 is calculated and used as the primary determinate of fitness[23]

$$\mathrm{GoF} = \frac{\chi^2}{(\text{degrees of freedom})} = \frac{\sum_j \left(\frac{\mathrm{calc}(j) - \mathrm{meas}(j)}{\mathrm{error}(j)}\right)^2}{(\#\text{ data pts - }\#\text{ floated parameters})} \quad (12)$$

where calc(j) is the calculated intensity of eq 11 and meas(j) is the measured MCD intensity level at that temperature and field strength. GoF values are dependent upon the experimental error (standard deviation) of individual data points; therefore, while GoF values are valuable for comparing different fits to a particular data set, they cannot be compared between different sets.

Physical Background to VTVH MCD Constraints. For the system under study, A_{satlim} only scales the magnitude of the C-term MCD intensity (note, data are usually normalized to one.) Thus this variable has no qualitative effect on the saturation behavior, and, therefore, a strong constraint is not necessary. Similarly, B term contributions, given as a percentage of A_{satlim}, scale linearly with magnetic field. It will add to or subtract from the intensity of VTVH MCD data, depending on the sign of the B-term intensity relative to C-term. This will have a subtle effect on the saturation of the signal at high field. In both cases, bounded constraints are adequate to fit these parameters.

$M_z/M_{x,y}$ and $g_\perp$ are only considered in some systems and so can initially be constrained to zero. Their effects are subtle, and the level of noise in most data sets will prevent accurate determination of these parameters. If the standard deviation is of high quality however, these parameters can typically be varied between 0 to 1 for $g_\perp$, and −1 to 1 for $M_z/M_{x,y}$.

Most difficult to constrain are $g_\parallel$ and δ for each doublet state. A single high-spin Fe(II) center has an $S = 2$ ground state that will split with a $g_\parallel \approx 8$. If the system contains two exchange coupled Fe(II) centers, these will interact through bridging ligands to give possible ground states with $M_S = \pm 4, \pm 3, \pm 2, \pm 1$, and 0 that will split by $g_\parallel$'s of approximately 16, 12, 8, 4, and 0, respectively. Doublets with larger effective $g_\parallel$ values tend to exhibit a smaller δ. For mononuclear Fe(II) systems δ tends to range from ∼1−7 cm^{-1}; this range can be further narrowed depending upon coordination geometry of the iron.[19] For Fe(II)Fe(II) exchange coupled systems δ can range from ∼0.1−7 cm^{-1}, and acceptable ranges for each $g_\parallel$ can be obtained using the proper spin Hamiltonian for these systems.[18]

In any system, it is possible for $g_\parallel$ to deviate from these integer values; however, without constraint most fitting methods will quickly move into an infeasible solution space. Applying classical constraints, such as logarithmic barriers,[24] to this parameter eliminates possible solutions that may lie close to the feasibility boundary or trap the optimization at the given boundary condition. Therefore, an adaptive penalty function proved to be an excellent alternative for this system.

Implementation Method for Adaptively Constrained Genetic Algorithms. Fitting VTVH MCD data provides a useful example of how to solve highly constrained GA problems as three different types were required

Type I: Bounded Constraints (δ, A_{satlim}, B_i, $M_z/M_{x,y}$, $g_\perp$).

$$\begin{matrix} g_i(\vec{x}) \le 0, i = 1, \ldots K \\ h_i(\vec{x}) = 0, i = 1, \ldots P \end{matrix} \quad \text{from Eq 2} \quad (13)$$

Type II: Discrete/Stochastic Constraints ($g_\parallel$).

$$P(z \in dz)/dz = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-1/2x^2}$$

$$\textit{where}: \quad \left\{ \begin{array}{l} \mu \in Z | \mu = [0, 4, 8, 12, 16] \\ \mu = \left\{ \begin{array}{ll} \mu_{i+1} = 4 & \mu_i = 0 \\ \mu_{i+1} = 12 & \mu_i = 16 \\ \mu_{i+1} = \mu_i \pm 4 & \textit{otherwise} \end{array} \right. \end{array} \right\}$$

$$\sigma^2 = 2/3 \quad (14)$$

Type III: Ascending Order Constraints (Energy, E_i).

$$\left\{ \begin{array}{l} E \in \mathscr{R} \wedge E \ge 0 \\ \textit{where}: \\ E_1 = 0 \\ E_{i+1} > E_i \end{array} \right\} \quad (15)$$

Three key GA subfunctions must be modified to incorporate the proposed constraints: a) Initial Population Creation, b) Fitness Scaling/Selection, and c) Crossover/Mutation.

a). Initial Population Creation. For this study, all variables were encoded as floating point numbers as standard binary mutation algorithms will not work on $g_\parallel$ and the energy parameter due to the complex constraints described below.

Bounded Initial Population Creation. For the simply bounded, Type I variables ($a \le x \le b$), the values of these genes

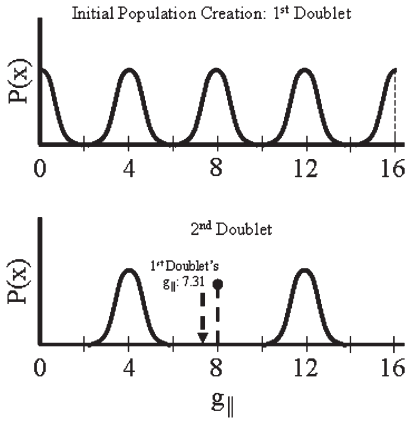


Figure 5. Constraints for creation of initial $g_{||}$ population for 1st and excited state doublets.

are selected using a uniform distribution constrained between the user selected lower and upper bounds.

Initial Population Constraint of $g_{||}$. For constraint type II, the $g_{||}$ variable, the rules are as follows. The first doublet for each individual is confined to values centered normally about 0, 4, 8, 12, or 16. For each subsequent doublet, its $g_{||}$ value is normally centered ±4 away from the previous doublet's integer value (determined from spin state diagrams generated via the spin Hamiltonian for $S = 2$ exchange coupled systems).[18]

To create the initial $g_{||}$ population, a univariate multimodal Gaussian random number generator is employed. This algorithm produces numbers centered about 0, 4, 8, 12, or 16 with a standard deviation that ensures 99% of the values will be ±2 of each designated mean. Once the first state for each individual has been created, each subsequent $g_{||}$ will use the same algorithm. However, these subsequent states are constrained such that the probability distribution will be bimodally centered ±4 from the previous doublet's $g_{||}$ level as seen in Figure 5 (lower). If the user constrains a doublet to a have a $g_{||}$ of a certain value, the algorithm will ensure that the surrounding spin states are suitably constrained, using a modified rejection method similar to the population creation function.

Initial Population Constraint of Energies. The energy of the first spin state in the system is set to zero. For each additional spin state (doublet or singlet), the energy state's value must be larger (constraint type III). The algorithm must also consider situations where the user has constrained the energy level of one of the singlets or doublets, creating implicit constraints on the surrounding states. Here, a modified rejection method is applied. For every individual, a numerical value is randomly generated for each spin state energy level. For example, three energy values will be randomly generated and assigned per individual in a Doublet, Doublet, Doublet (D,D,D) Spin State model. These values are generated using a uniform distribution, constrained from zero to the user defined max allowable energy value. Once the energy values have been generated for the population, each person's energy values are evaluated for order violations. To do so, each spin state's energy value is successively examined to see whether it is greater than that of the previous spin state. If all of the randomly assigned energy values are increasing order for an individual, their gene sequence is saved. If not, the values of their energy levels are randomly regenerated. The cycle continues until a feasible set of energy levels have been developed for each individual in the population.

b). Fitness, Raw Error, and Selection. Individuals are given an expectation for selection into the next generation based on their fitness. In the case of the VTVH MCD problem, an individual's fitness is the Goodness of Fit of the resultant data fit (eq 12). After the individual GoF scores are computed, the raw error is assessed using the method outlined below.

Three areas of error are assessed: minimum and maximum boundary violations, energy level ascending sequence, and $g_{||}$ rules. The magnitude of each error is summed together to determine the total raw error for that individual.

- The error for bounded constraints is the absolute distance from the defined limits
- For energy level errors, each spin state N and N+1 pair are successively examined with respect to each other. If the N+1 energy level is less than the energy of spin state N, the error is the magnitude of that relative violation.
- For each doublet, the subsequent N+1 doublet is examined. If its $g_{||}$ value does not follow the ±4 rule, the error is the distance of the N+1th state from the closest allowable $g_{||}$ integer.

The errors are summed to give the total error for an individual. Using each individual's fitness and raw error, the composite fitness then is determined with the Nanakorn adaptive penalty function method[13] and expectation for selection is assessed. Parents for the next generation are selected using stochastic universal sampling (SUS),[25] a technique similar to a roulette wheel, where each individual's slot size is equal to its computed expectation.

Elite Individuals, Crossover, and Mutation. A user-defined number of elite individuals are selected from the population, and their genome is passed on without modification to the next generation. The remaining individuals create children by combining their genes or undergoing random mutation of their own genome. The former process, called crossover, selects genes and gene sets from each parent when creating the child's genome. If a parent is selected for mutation, all of their genes will be stochastically changed, if not fixed by the user a priori. For this study, the number of elite children and the mutation vs crossover probabilities were fixed. It has been reported that meta-optimization of these parameters in parallel with the main optimization problem improves convergence speed.[26] However, to isolate the benefits of the adaptive penalty function technique, this multiprocess approach was not included.

Crossover. For those parents selected for crossover, there are two modes for selecting which genes to insert into the child's genome. For simply constrained Type I variables, δ, A_{satlim}, B_i, $M_z/M_{x,y}$, and $g_{\perp}$, a uniform crossover technique was used. In this, each gene has an equal probability of being selected from either parent and inserted into the child's genome.

For the Type II and III variables, each parent's energy and $g_{||}$ sequence was considered a genome set. As there is a specific constraint order, selecting the first spin state $g_{||}$ value from one parent and the second spin state $g_{||}$ value from the other typically results in infeasible solutions. While it is beneficial to mix up the genes and try new combinations, this is the job of mutation function, not crossover. Therefore, each of the sequences are copied from a randomly selected parent into the child's genome without modification. In keeping with the uniform crossover technique, the parent selected to provide their energy genome block is selected independently of the parent providing its $g_{||}$ values.

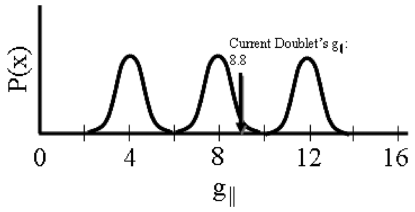


Figure 6. Allowable mutation regions for a hypothetical $g_{\|}$ (type II constraint).

Mutation. For those individuals selected for mutation, the goal is to bring diversity into the population without producing individuals that violate the Type I, II, or III constraints.

For the simply bounded, Type I constrained δ, A_{satlim}, B_i, $g_{\perp}$, M_z/M_{xy} variables, each gene's new value is constrained to a normal distribution about the parent's value, bounded by the user defined minimum and maximum limits. In addition, each gene's standard deviation is a function of the percentage of elapsed generations compared to the limit. At the start of a genetic algorithm evolution, the genes are allowed to fluctuate across the full allowable range of values. As the user set generation limit approaches, the standard deviation used for each Gaussian distribution is decreased, enabling convergence toward a quasi-optimal solution.

The Type II constrained $g_{\|}$ mutation function maintains the ± 4 rule for each mutation parent as follows. The first doublet is allowed to mutate about the closest allowable $g_{\|}$ integer (0,4,8,12,16) as well the $g_{\|}$ integer four above and below. As described in the Population Creation section, a multimodal Gaussian distribution is employed to select the new $g_{\|}$ value. The standard deviation for each distribution is set such that there is a 3σ probability that the new $g_{\|}$ will fall within ± 2 of any of the three possible means (see Figure 6). Each subsequent doublet's $g_{\|}$ value will be multimodal distributions about the previous spin state's newly selected value so to maintain solution feasibility. For example, for a two doublet system where the original first double value is 8.8, the new first double $g_{\|}$ value is selected from normal distributions centered at 4, 8, and 12. If 12.2 is selected, then the second doublet $g_{\|}$ will be selected from distributions centered at 8, 12, and 16.

For the Type III (ascending order constrained energy), each spin state's energy level is generated anew in a normal distribution, centered about the parent's energy values. For speed, each spin state's energy level is picked in order, with its lower bound constrained by the previously selected energy value. While the variables are no longer selected independently as with the initial population approach, this algorithm avoids the costly and possibly nonconvergent rejection method every generation. The standard deviation used in each normal distribution follows the same generation based shrinkage rules described in the Type I Mutation section, staying within the local region as the population limit approaches.

Ultimately within each new generation two of the individuals become elite children, a certain percentage combine in crossovers, and the remaining are mutated. The cycle starts anew until the user defined termination criteria are met, and the final sun-optimal GA solution is further converged by SQP.

SQP Considerations. In Sequential Quadratic Programming, only equality and inequality constraints can be set, leaving a need to find a different method to constrain $g_{\|}$. Since the design vector starts in a feasible trust region near the local optimum, a cubic penalty function was used. This function applies an increasing penalty to the fitness function GoF's raw value depending on how far the gene is away from the allowable $g_{\|}$ integer. The maximum distance is ± 2, and the maximum penalty is assigned by expert input. As an illustrative example, a GoF of $\sim$0.2 is desired, the max cubic penalty is set to 0.5 and if the current $g_{\|}$ is +2 away from 4, the composite penalized GoF will set to 0.7. This is 3.5$\times$ worse than the raw GoF value which is a large relative penalty, forcing SQP to find solutions that stay very close to the allowable $g_{\|}$ integers. From this, a reasonable level of cubic penalty can be assigned based on the requirement to keep $g_{\|}$ close integer values.

III. VALIDATION OF THE ADAPTIVE GA METHOD AND COMPARISON TO PREVIOUS SIMPLEX METHOD

Three steps were taken to validate that the new GA method can find the correct global optimum and the solution to eq 11 given an MCD data set.

First, we compared chi-squared values calculated by each code's fitness function given identical inputs. We achieved agreement between the methods, caveated by the numerical integration discussion below.

Second, using the simplex code,[27] we generated two sets calculated intensities using different known inputs. Then, we took these generated data sets and had the GA fit each of them 100 times. The GA was able to find the correct answers 70% to 87% of the time depending on which data set was run. The remaining fits were degenerate cases where the optimizer got caught in a nonphysical local minimum.

In some of the variables there were some small offsets between the inputs used to generate the curves and the resulting fitted values. However, the small offsets were approximately identical for all of the correct answers. For example, the first doublet $g_{\|}$ was set to 8 in the simplex method and the GA calculated 8.41 $\pm$ 1.83e-3 69% of the time for data set 1. In the second data set, $g_{\perp}$ was set to 1, and the GA found 0.88 $\pm$ 5.69e-4 75 times out of 100 (see Table SI 20).

The differences between the generated spectra inputs and the calculated fit values can be attributed to the GA method using a variable-step quadrature integrator and the simplex method using a fixed four-step trapezoidal method. When integrating eq 11's nonlinear $\cos^2$, $\sin^2$, and $\sin^3$, a more sophisticated integrator will produce slightly different answers than the original. This in turn steers the optimizer to a slightly different answer. However, given the highly multimodal character of the eq 11, the repeatable nature of calculating the fitting parameters of known spectra validates the GA's ability to reliably find the global optimum.

Finally, we tested the GA method against a complex, real world problem to confirm the GA could locate physical solutions within noisy experimental data (see Tables S1−S4).

To test this method, VTVH MCD data collected on biferrous loaded enzyme *myo*-inositol oxygenase (MIOX)[28,29] was analyzed. Fits to these data using the Simplex method found three potentially physical solutions. [These three solutions can be distinguished using additional experimental information and fitting these data to an independent set of spin-Hamiltonian equations.[19] This will be published separately.] This process required over a month of user time to manually develop input sets and sort through the raw results. In comparison, the Hybrid GA technique searched a much broader N-dimensional space (3×10^6 possibilities) and found those same three potential physical solutions with slightly better χ^2 values (Table 1). The GA fitting runs were completed in less than 37 hours of computation time (one processor, 2.5 Ghz),

Table 1. Comparison of Physically Viable Solutions Found by the Simplex and Adaptive Genetic Algorithm Methods[27]

		Simplex Method			Hybrid Genetic Algorithm Method		
	model	solution 1: D, D	solution 2: D, D	solution 3: S, S, D	solution 1: D, D	solution 2: D, D	solution 3: S, S, D
	X^2	0.28	0.32	0.38	0.27	0.25	0.19
spin state 1	δ	2.0	2.0		2.3	3.9	
	$g_\parallel$	7.8	7.8		7.8	8.1	
	A_{satlim}	1.15	1.23		1.27	1.46	
	B_i	6.41	4.54	−1.09	4.74	3.18	−4.27
	E	0	0	0	0	0	0
	$g_\perp$	0	0		0	0	
	$M_Z/M_{X,Y}$	0	0		0	0	
spin state 2	δ	0.26	0.10		0.14	0.06	
	$g_\parallel$	12.7	4.0		11.7	4.2	
	A_{satlim}	0.75	1.24		0.57	0.94	
	B_i	−4.2	0.45	1.27	−1.32	2.18	6.29
	E	8.0	8.3	0.0	10.6	4.0	0.2
	$g_\perp$	0	0		0	0	
	$M_Z/M_{X,Y}$	0	0		0	0	
spin state 3	δ			1.01			0.03
	$g_\parallel$			16			16
	A_{satlim}			1.21			1.21
	B_i			0.23			−79.87
	E			1.1			0.9
	$g_\perp$			0			0
	$M_Z/M_{X,Y}$			0			0

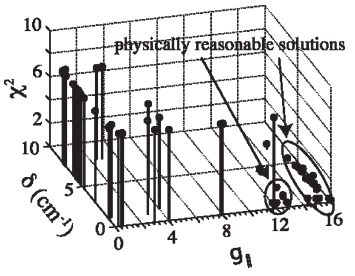


Figure 7. Cluster analysis of 100 Monte Carlo runs. MIOX model assumed singlet, singlet, doublet (S,S,D) spin state order.

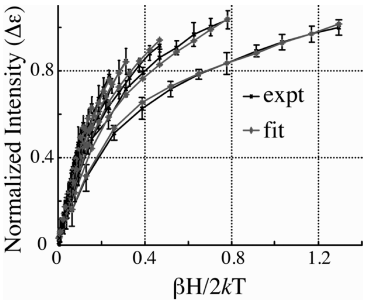


Figure 8. Potential MIOX fit using hybrid GA technique (S,S,D spin state model).

and the best solutions were extracted in less than an hour of user analysis. The χ^2 improvement may be attributed to using a SQP solver in our hybrid solver, which is superior to the Simplex method in finding the closest local minima.

By taking a Monte Carlo approach to the initial conditions and running the hybrid solver repeatedly, cluster analysis plots such as shown in Figure 7 and S12−S19 allow the user to rapidly distinguish the viable physical solutions from purely mathematical ones. Figure 8 shows an overlay of the best fit for the S, S, D (singlet, singlet, doublet) spin state model with experimental data. Figures S7 and S8 show the equivalent plots for the other two physically reasonable solutions.

■ CONCLUSIONS AND APPLICABILITY TO OTHER SYSTEMS

This application had now been validated for the complicated fitting of binuclear Fe(II) VTVH MCD data, requiring at least 13 interdependent variables (Figure 8). As the user interface allows for easy adjustments to the parameter boundaries, it can also be used to fit VTVH MCD data for mononuclear Fe(II) as well as copper-oxo triplet and many other systems.

While other GA approaches have been applied to various spectroscopic fitting problems,[30,31] few systems outside the engineering realm have used adaptive penalty functions. For example, in the fitting of Mössbauer data,[32] the use of a GA provided ballpark initial guesses − yet additional refinement was required using a commercially developed fitting program.

Systems that have previously struggled with or avoided GA methods due to the unconstrained nature of the fitting process may find that an adaptive hybrid method provides an ideal solution for complex multimodal solution spaces. For soft or nonlinear boundary situations, an adaptive penalty function, which modifies the fitness of the population relative to the others,

dx.doi.org/10.1021/ci2001296 |J. Chem. Inf. Model. 2011, 51, 2164–2173

promotes diversity and the expression of genes that could lead to the optimal data fit. And with the expansion of the Nanakorn theory included in this paper, this adaptive penalty approach may now be applied to both maximization and minimization problems.

The adaptive hybrid GA approach has been shown to quickly and efficiently find the multiple local minima in the complex search space, both with respect to computational and user analysis time. Most importantly, the ability to fully search a multimodal solution space greatly improves the quality and confidence in the final solution obtained. These characteristics may allow it to be used in a wider array of chemical and engineering problems, such as fitting EPR, Mössbauer, X-ray absorption, spectroscopic, and kinetics data.

■ ASSOCIATED CONTENT

Supporting Information. First is the method validation bounds table which details each of the data analysis validation runs including constraints, parameter settings, and performance statistics. Second, we have included plots of the two MIOX data sets collected by MCD used for algorithm validation. Third are measured vs GA calculated VTVH data plots are displayed for each of the most likely MIOX solution candidates. Finally, cluster graphs of each of the different spin state model results are displayed, showing the spread of possible solutions within a 100 GA run set. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: edward.solomon@stanford.edu.

■ ACKNOWLEDGMENT

Dr. Elizabeth Pavel originally wrote the Goodness of Fit function used in this program. Yeonju Kwak generated Simplex-based test spectra to help validate the algorithm. We thank NSF MCB-0919027 for support of this research

■ REFERENCES

(1) Gill, P. E.; Murray, W.; Wright, M. H. In *Practical Optimization*; Academic Press: London, UK, 1981; pp 133−259.

(2) Sun, W.; Yuan, Y. In *Optimization theory and methods: nonlinear programming*; Springer: New York, NY, 2006; pp 119−146.

(3) Lampinen, J.; Zelinka, I. Mixed Integer-Discrete-Continuous Optimization By Differential Evolution, Part 1: the optimization method. In *MENDEL'99*, Proceedings of the 5th International Mendel Conference on Soft Computing, Czech Republic, June 9−12; Technical University of Brno Brno, Czech Republic, 1999; pp 71−76.

(4) Haupt, R.; Haupt, S. E. In *Practical Genetic Algorithms*; John Wiley & Sons, Inc.: New York, 1998; pp 2−64.

(5) Conn, A.; Gould, N.; Toint, P. A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds. *SIAM J. Numer. Anal.* **1991**, *28*, 27.

(6) Tanese, R. Distributed Genetic Algorithms. In Proceedings of the 3rd International Conference on Genetic Algorithms, Morgan Kaufmann Publishers Inc.: San Francisco, 1989; pp 434−439.

(7) Goldberg, D. E. In *Genetic Algorithm in search optimization and machine learning*; Addison-Wesley: New York, 1989.

(8) Mitchell, M. In *An introduction to genetic algorithms*; MIT Press: Cambridge, MA, 1998; pp 155−178.

(9) Vose, M. In *The simple genetic algorithm: foundations and theory*; MIT Press: Cambridge, MA, 1999; pp 11−43.

(10) Coello, C. Use of a self-adaptive penalty approach for engineering optimization problems. *Computers in Industry* **2000**, *41*, 113–127.

(11) Coit, D. W.; Smith, A. E.; Tate, D. M. Adaptive Penalty Methods for Genetic Optimization of Constrained Combinatorial Problems. *Informs J. Comput.* **1996**, *8*, 173–182.

(12) Tessema, B.; Yen, G. G. A Self Adaptive Penalty Function Based Algorithm for Constrained Optimization. In CEC 2006, Proceedings of the IEEE Congress on Evolutionary Computation, Vancouver, BC, IEEE: New York, NY, pp 246−253.

(13) Nanakorn, P.; Meesomklin, K. An adaptive penalty function in genetic algorithms for structural design optimization. *Computers & Structures* **2001**, *79*, 2527–2539.

(14) Goldberg, D., In, E. *Genetic Algorithms in search, optimization and machine learning*; Addison-Wesley: Reading, MA, 1989; pp 2−15.

(15) Lobo, F. G.; Lima, C. F.; Michalewicz, Z. In *Parameter setting in evolutionary algorithms*; Springer: New York, NY, 2007; pp 1−46.

(16) Shun-Fa, H.; Rong-Song, H. A hybrid real-parameter genetic algorithm for function optimization. *Adv. Eng. Inform.* **2006**, *20*, 7–21.

(17) Boggs, P. T.; Tolle, J. W. Sequential Quadratic Programming. *Acta Numerica* **1995**, *4*, 1–51.

(18) Solomon, E. I.; Brunold, T. C.; Davis, M. I.; Kemsley, J. N.; Lee, S. K.; Lehnert, N.; Neese, F.; Skulan, A. J.; Yang, Y. S.; Zhou, J. Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem. Rev.* **2000**, *100*, 235–349.

(19) Solomon, E. I.; Pavel, E. G.; Loeb, K. E.; Campochiaro, C. Magnetic Circular Dichroism Spectroscopy as a Probe of the Geometric and Electronic Structure of Non-Heme Ferrous Enzymes. *Coord. Chem. Rev.* **1995**, *144*, 369–460.

(20) Zhang, Y.; Gebhard, M. S.; Solomon, E. I. Spectroscopic Studies of the Non-Heme Ferric Active Site in Soybean Lipoxygenase: Magnetic Circular Dichroism as a Probe of Electronic and Geometric Structure. Ligand Field Origin of Zero-Field Splitting. *J. Am. Chem. Soc.* **1991**, *113*, 5162–5175.

(21) Stephens, P. J. Magnetic Circular Dichroism. *Annu. Rev. Phys. Chem.* **1974**, *25*, 201–232.

(22) Bennett, D. E.; Johnson, M. K. The electronic and magnetic properties of rubredoxin: a low-temperature magnetic circular dichroism study. *Biophys. Acta* **1987**, *911*, 71–80.

(23) Pavel, E.; Solomon, E. *Magnetic Circular Dichroism Spectroscopic Studies of Mononuclear Non-heme Iron Sites: Methodology and Applications to Non-heme Enzymes.* Ph.D. Thesis, Stanford University, Stanford, CA, 1997.

(24) Hertog, D.; Roos, C.; Terlaky, T. On the classical logarithmic barrier function method for a class of smooth convex programming problems. *JOTA* **1992**, *73*, 1–25.

(25) Baker, J. E. Reducing bias and inefficiency in the selection algorithm. In *Genetic Algorithms and Their Applications*, Proceedings of the Second International Conference on Genetic Algorithms, Cambridge, Massachusetts, July 28−31; L. Erlbaum Associates: Hillsdale, NJ, 1987; pp 14−27.

(26) Back, T. Parallel Optimization of Evolutionary Algorithms. In *International Conference on Evolutionary Computation*, Proceedings of the Third Conference on Parallel Problem Solving from Nature: Parallel Problem Solving from Nature, Davidor, Y., Schwefel, H.-P., Männer, R., Eds.; Springer-Verlag: London, 1994; pp 418−427.

(27) John James, T. *On convergence of the nelder-mead simplex algorithm for unconstrained stochastic optimization.* Ph.D. Thesis, Pennsylvania State University, State College, PA, 1995.

(28) Charalampous, F. C. Biochemical studies on inositol. V. Purification and properties of the enzyme that cleaves inositol to D-glucuronic acid. *J. Biol. Chem.* **1959**, *234*, 220–227.

(29) Charalampous, F. C.; Lyras, C. Biochemical studies on inositol. IV. Conversion of inositol to glucuronic acid by rat kidney extracts. *J. Biol. Chem.* **1957**, *228*, 1–13.

(30) Spalek, T.; Pietrzyk, P.; Sojka, Z. Application of the genetic algorithm joint with the Powell method to nonlinear least-squares fitting of powder EPR spectra. *J. Chem. Inf. Model.* **2005**, *45*, 18–29.

(31) da Costa, P. A.; Poppi, R. J. Genetic algorithm in chemistry. *Quim. Nova* **1999**, *22*, 405–411.
(32) Ahonen, H.; de Souza Júnior, P. A.; Garg, V. K. A genetic algorithm for fitting Lorentzian line shapes in Mössbauer spectra. *Nucl. Instrum. Methods Phys. Res., Sect. B* **1997**, *124*, 633–638.