CHAPTER TEN

DynaFit—A Software Package for Enzymology

Petr Kuzmič

Contents

Abstract

Since its original publication, the DynaFit software package [Kuzmič, P. (1996). Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV proteinase. *Anal. Biochem.* 237, 260–273] has been used in more than 500 published studies. Most applications have been in biochemistry, especially in enzyme kinetics. This paper describes a number of recently added features and capabilities, in the hope that the tool will continue to be useful to the enzymological community. Fully functional DynaFit continues to be freely available to all academic researchers from http://www.biokin.com.

BioKin Ltd., Watertown, Massachusetts, USA

Methods in Enzymology, Volume 467
ISSN 0076-6879, DOI: 10.1016/S0076-6879(09)67010-5

1. Introduction

DynaFit (Kuzmič, 1996) is a software package for the statistical analysis of experimental data that arise in biochemistry (e.g., enzyme kinetics; Leskovar *et al.*, 2008), biophysics (protein folding; Bosco *et al.*, 2009), organic chemistry (organic reaction mechanisms; Storme *et al.*, 2009), physical chemistry (guest–host complexation equilibria; Gasa *et al.*, 2009), food chemistry (fermentation dynamics; Van Boekel, 2000), chemical engineering (bio-reactor design; Von Weymarn *et al.*, 2002), environmental science (bio-sensors for heavy metals; Le Clainche and Vita, 2006), and related areas.

The common features of these diverse systems are that (a) the underlying theoretical model is based on the mass action law (Guldberg and Waage, 1879); (b) the model can be formulated in terms of stoichiometric equations; and (c) the experimentally observable quantity is a linear function of concentrations or, more generally, populations of reactive species.

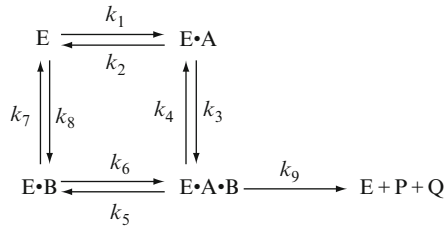
The main use of DynaFit is in establishing the detailed molecular mechanisms of the physical, chemical, or biological processes under investigation. Once the molecular mechanism has been identified, DynaFit can be used for routine quantitative determination of either microscopic rate constants or thermodynamic equilibrium constants that characterize individual reaction steps.

DynaFit can be used for the statistical analysis of three different classes of experiments: (1) the progress of chemical or biochemical reactions over time; (2) the initial rates of enzyme reaction, under either the rapid-equilibrium or the steady-state approximations (Segel, 1975); and (3) equilibrium ligand-binding studies.

Regardless of the type of experiment, the main benefit of using the DynaFit package is that it allows the investigator to specify the fitting model in the biochemical notation (e.g., `E + S <==> E.S --> E + P`) instead of mathematical notation (e.g., $v = k_{cat}[E]_0[S]_0/([S]_0 + K_m)$).

For example, to fit a set of initial rates of an enzyme reaction to a steady-state kinetic model for the "Bi Bi Random" mechanism (Segel, 1975, p. 647) (Scheme 10.1), the investigator can specify the following text in the DynaFit input file:

```
[data]
    data = rates
    approximation = steady-state
[mechanism]
    E + A <==> E.A : k1 k2
    E.A + B <==> E.A.B : k3 k4
    E.A.B <==> E.B + A : k5 k6
```



Scheme 10.1

```
    E.B <==> E + B : k7 k8
    E.A.B --> E + P + Q : k9
[constants]
    k8 = (k1 k3 k5 k7) / (k2 k4 k6)
...
```

The program will internally derive the initial rate law corresponding to this steady-state reaction mechanism (or any arbitrary mechanism), and perform the least-squares fit of the experimental data. This allows the investigator to focus exclusively on the biochemistry, rather than on the mathematics. Using exactly equivalent notation, one can analyze equilibrium binding data, such as those arising in competitive ligand displacement assays, or time-course data from continuous assays.

Importantly, the DynaFit algorithm does not make any assumptions regarding the relative concentrations of reactants. Specifically, it is no longer necessary to assume that the enzyme concentration is negligibly small compared to the concentrations of reactants (substrates and products) and modifiers (inhibitors and activators). This feature is especially valuable for the kinetic analysis of "slow, tight" enzyme inhibitors (Morrison and Walsh, 1988; Szedlacsek and Duggleby, 1995; Williams and Morrison, 1979).

Since its original publication (Kuzmič, 1996), DynaFit has been utilized in more than 500 journal articles. In the intervening time, many new features have been added. The main purpose of this report is to give a brief sampling of several newly added capabilities, which might be of interest specifically to the enzymological community. The survey of DynaFit updates is by no means comprehensive; the full program documentation is available online (http://www.biokin.com/dynafit).

This article has been divided into four parts. The first three parts touch on the three main types of experiments: (1) equilibrium ligand binding studies; (2) initial rates of enzyme reactions; and (3) the time course of enzyme reactions. The fourth and last part contains a brief overview of selected data-analytical approaches, which are common to all three major experiment types.

2. Equilibrium Binding Studies

DynaFit can be used to fit, or to simulate, equilibrium binding data. The main purpose is to determine the number of distinct noncovalent molecular complexes, the stoichiometry of these complexes in terms of component molecular species, and the requisite equilibrium constants.

The most recent version of the software includes features and capabilities that go beyond the original publication (Kuzmič, 1996). For example, DynaFit can now be used to analyze equilibrium binding data involving—at least in principle—an unlimited number of simultaneously varied components. A practically useful *four*-component mixture might include (1) a protein kinase; (2) a Eu-labeled antibody (a FRET-donor) raised against the kinase; (3) a kinase inhibitor, whose dissociation constant is being measured; and (4) a fluorogenic FRET-acceptor molecule competing with the inhibitor for binding. Investigations are currently ongoing into the optimal design of such multicomponent equilibrium binding studies.

2.1. Experiments involving intensive physical quantities

DynaFit can analyze equilibrium binding experiments involving *intensive* physical quantities. Unlike their counterparts, the *extensive* physical quantities, intensive quantities do not depend on the total amount of material present in the system. Instead, intensive quantities are proportional to mole fractions of chemical or biochemical substances. A prime example of an intensive physical quantity is the NMR chemical shift (assuming that fast-exchange conditions apply, where the chemical shift is a weighted average of chemical shifts of all microscopic states of the given nucleus).

We have recently used this technique to investigate the guest–host complexation mechanism in a system involving three different ionic species of a guest molecule (paraquat, acting as the "ligand") binding to a crown-ether molecule (acting as the "receptor"), with either 1:1 or 1:2 stoichiometry (Gasa *et al.*, 2009). This guest–host system involved four components forming up to nine noncovalent molecular complexes, and a correspondingly large number of microscopic equilibrium constants. DynaFit has also been used in the NMR context to determine the binding affinity between the RIZ1 tumor suppressor protein and a model peptide representing histone H3 (Briknarová *et al.*, 2008).

The following illustrative example involves the use of DynaFit for highly precise determination of a protein–ligand equilibrium binding constant.

2.1.1. NMR study of protein–protein interactions

Figure 10.1 (unpublished data courtesy of K. Briknarová and J. Bouchard, University of Montana) displays the changes in NMR chemical shifts for six different protons and six different nitrogen nuclei in the PR domain from a transcription factor PRDM5 (Deng and Huang, 2004), depending on the concentration of a model peptide ligand. The NMR chemical shift data for all 12 nuclei were analyzed in the global mode (Beechem, 1992). The main purpose of this experiment was to determine the strength of the binding interaction. It was assumed that the binding occurs with the simplest 1:1 stoichiometry. A DynaFit code fragment corresponding to Scheme 10.2 is shown as follows:

```
[mechanism]
    R + L <==> R.L : Kd1 dissociation
[responses]
    intensive
[data]
    plot titration
...
```

Note the use of the keyword `intensive` in the `[responses]` section of the script, which means that the observed physical quantity

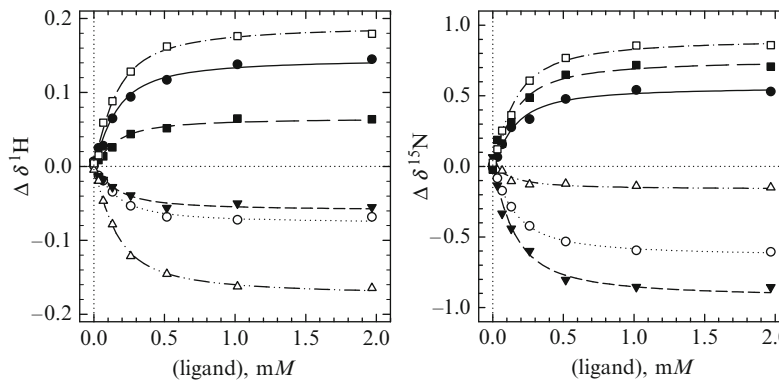


Figure 10.1 NMR chemical shift titration of the PRDM5 protein (total concentration varied between 0.125 and 0.1172 m*M*) with a model peptide ligand. *Left*: ^{1}H-chemical shifts of six selected protons. *Right*: ^{15}N-chemical shifts of six selected nitrogen nuclei. The chemical shifts for all 12 nuclei were fit globally (Beechem, 1992) to the binding model shown in Scheme 10.2.

$$\mathrm{R + L} \underset{}{\overset{K_{d1}}{\rightleftharpoons}} \mathrm{R{\bullet}L}$$

Scheme 10.2

(chemical shift) is proportional not to the quantity of various molecular species present in the sample, but rather to the corresponding mole fractions.

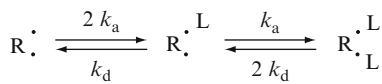
Also note the keyword `titration`, which is used to produce a simple Cartesian plot—with the ligand concentration [L] formally acting as the only independent variable—even though the experiment was performed by gradual addition of ligand to the same initial protein sample. This means that both the protein (titrand) and the model peptide (titrant) concentrations were changing with each added aliquot. It is very important to recognize that, in this case, the experimental data points are not statistically independent, as is implicitly assumed by the theory of nonlinear least-squares regression (Johnson, 1992, 1994; Johnson and Frasier, 1985). However, the practice of incrementally adding to the same base solution of the titrand has been firmly established in protein–protein and protein–ligand NMR titration studies.

The best-fit value of the dissociation equilibrium constant, determined from the data shown in Fig. 10.1, was $K_{d1} = (0.087 \pm 0.007)$ [0.073 ... 0.108] m*M*. The values in square brackets are approximate confidence intervals determined by the *profile-t* method of Bates and Watts (Brooks *et al.*, 1994). Please note that, unlike the formal standard error shown in the parentheses, the confidence intervals are not symmetrical about the best-fit value.

Using the global fit method (Beechem, 1992), the strength of the protein–ligand binding interactions was determined for a number of different nuclei, and the results were highly consistent; the coefficient of variation for the equilibrium was approximately 10% regardless of which chemical shift was monitored.

2.2. Independent binding sites and statistical factors

The most recent version of DynaFit (Kuzmič, 1996) allows the investigator to properly define the relationship between (a) *intrinsic* rate constant or equilibrium constants, and (b) *macroscopic* rate constants or equilibrium constants. This distinction is necessary in the analysis of multiple identical binding sites. As the simplest possible example, consider the binding of L, a ligand molecule, to R, a receptor molecule that contains two identical and independent binding sites (Scheme 10.3).



Scheme 10.3

In Scheme 10.3, k_a and k_d are intrinsic rate constants. The statistical factors ("2") shown in Scheme 10.3 express the fact that there are two identical pathways for L to associate with R, but only one way for for L to associate with RL. Similarly, RL_2 can yield RL in two equivalent ways, whereas RL can dissociate into R + L only in one way. Thus, if we define the first and second dissociation equilibrium constants as $K_1 = [RL]_{eq}[L]_{eq}/[RL_2]_{eq}$ and $K_2 = [R]_{eq}[L]_{eq}/[RL]_{eq}$, then for independent equivalent sites we must have $K_1 = 4K_2$.

In the DynaFit notation, the difference between independent and interacting binding sites can be expressed by using the following syntax:

```
[task]
    data = equilibria
    model = interacting ?
[mechanism]
    R + L <==> R.L : K1 dissociation
    R.L + L <==> R.L.L : K2 dissociation
[constants]
    K1 = ...
    K2 = ...
...
[task]
    data = equilibria
    model = independent ?
[mechanism]
    R + L <==> R.L : K1 dissociation
    R.L + L <==> R.L.L : K2 dissociation
[constants]
    K1 = 4 * K2 ; <== STATISTICAL FACTOR
    K2 = ...
...
```

2.2.1. Interacting versus independent sites on a trimeric enzyme

Błachut-Okrasinska *et al.* (2007) utilized DynaFit for a comprehensive kinetic investigation of mRNA cap analogues binding to the eIF4E regulatory protein (see also Niedzwiecka *et al.*, 2007). From the same laboratory comes a study of the trimeric purine nucleoside phosphorylase (PNP) interacting with nucleoside multisubstrate inhibitors (Wielgus-Kutrowska *et al.*, 2007). A representative equilibrium binding experiment is shown in Fig. 10.2 (raw experimental data (Wielgus-Kutrowska and Bzowska, 2006) courtesy of B. Wielgus-Kutrowska, Warsaw University). The object of the experiment was to determine whether inhibitor binding sites on the PNP trimer are independent or interacting.

Under the given assay conditions, the PNP enzyme is a nondissociative homotrimer. The presence of three separate inhibitor binding sites is

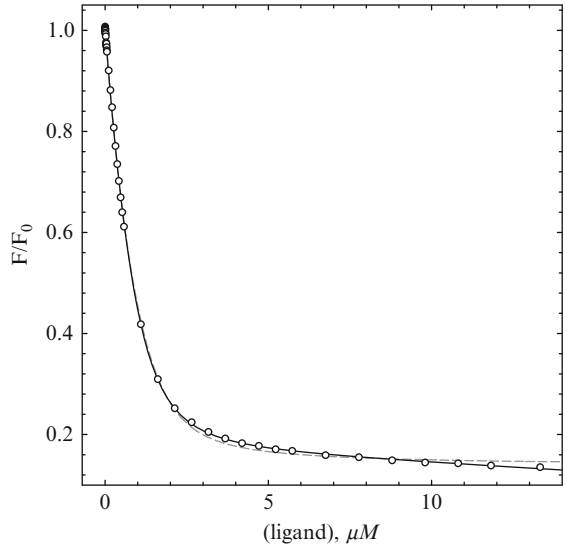


Figure 10.2 Equilibrium titration of the trimeric PNP from *Cellulomonas* sp. (0.47 μM as monomer) with a nucleoside analog inhibitor 2-amino-9-[2-(phosphonometoxy)-ethyl]-6-sulfanylpurine; F/F_0 represents relative fluorescence intensity (PNP plus ligand divided by PNP only). See Wielgus-Kutrowska and Bzowska (2006) for details. *Solid curve*: least-squares fit to the interacting sites model (Scheme 10.4). *Dashed curve*: least-squares fit to the independent sites model, in which equilibrium constants were linked via statistical factors such that $K_1{:}K_2{:}K_3 = 9{:}3{:}1$.

$$\mathrm{P} \overset{K_1}{\rightleftharpoons} \mathrm{PL} \overset{K_2}{\rightleftharpoons} \mathrm{PL_2} \overset{K_3}{\rightleftharpoons} \mathrm{PL_3}$$

Scheme 10.4

represented in Scheme 10.4 by three *association* equilibrium constants, K_1, K_2, and K_3. If the inhibitor sites were genuinely independent, the titration data would fit sufficiently well to an equilibrium binding model where the ratios $K_1{:}K_2{:}K_3 = 9{:}3{:}1$ are strictly maintained. In contrast, if the binding sites are interacting, it would be necessary to relax the fitting model such that the equilibrium constants could attain arbitrary values.

To perform the model discrimination analysis (Myung and Pitt, 2004) using the Akaike Information Criterion (AIC_c) (Burnham and Anderson, 2002), the requisite DynaFit script contains the following text:

```
[task]
    data = equilibria
    model = interacting ?
[mechanism]
    P + L <===> P.L : K1 equilibrium
```

```
    P.L + L <===> P.L.L : K2 equilibrium
    P.L.L + L <===> P.L.L.L : K3 equilibrium
[constants] ; vary independently
    K3 = 1 ?
    K2 = 3 ?
    K1 = 9 ?
...
[task]
    data = equilibria
    model = independent ?
[constants] ; link via statistical factors
    K3 = 1 ?
    K2 = 3 * K3
    K1 = 9 * K3
...
```

As can be seen from Fig. 10.2 (dashed curve), the independent-sites model provides a poor description of the available data. The interacting-sites model (solid curve) produces a much better fit. This result is in agreement with previously published investigations of the same system (Bzowska, 2002; Bzowska *et al.*, 2004; Wielgus-Kutrowska *et al.*, 2002, 2007).

3. Initial Rates of Enzyme Reactions

The study of initial rates of enzyme-catalyzed reactions defines the traditional approach to mechanistic enzymology (Segel, 1975). Earlier versions of the DynaFit software package (Kuzmič, 1996) were suitable for the analysis of initial-rate data under the *rapid equilibrium* approximation (Kuzmič, 2006), where it is assumed that the chemical steps in an enzyme mechanism are negligibly slow in comparison with all association and dissociation steps.

The current version of DynaFit extends the initial rate analysis to the more general *steady-state approximation* (Kuzmič, 2009a). This section introduces the important topic of thermodynamic cycles, which are relevant in steady-state enzyme mechanisms, especially those involving multiple substrates (e.g., kinases or reductases). A simulation study, involving dihydrofolate reductase (DHFR) as a model system, provides an illustrative example.

3.1. Thermodynamic cycles in initial rate models

It is a fundamental fact of thermodynamics that the Gibbs free energy change is independent of any particular path between thermodynamic states. This leads to the idea of a *thermodynamic box* in enzyme kinetic mechanisms (Gilbert, 1999, p. 271).

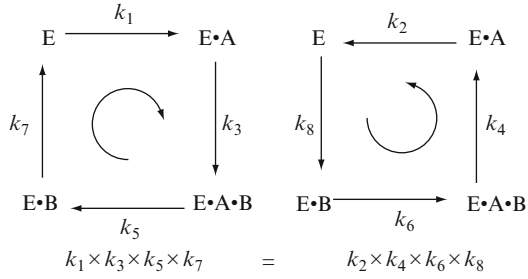
There are numerous logically equivalent ways to express the idea of a thermodynamic box. When expressed specifically in terms of microscopic rate constants, the product of rate constants associated with a set of arrows starting and ending at a given reactant must be the same in both directions (clockwise and counterclockwise, Scheme 10.5). This is equivalent to saying that the overall equilibrium constant associated with any cyclic path through the mechanism must be unity.

In general, we do not usually have advance knowledge of the Gibbs free energy change associated with the uncatalyzed reaction. However, for all nonchemical steps in the mechanism (i.e., noncovalent binding and dissociation of ligands), the overall equilibrium constant for each thermodynamic cycle must be unity. We can use this fact to check on the consistency of a postulated set of rate constant values. In the latest version of DynaFit, we can also use this fact to constrain the values of particular microscopic rate constants.

3.1.1. Steady-state initial rate equation for DHFR

The catalytic mechanism of *Escherichia coli* DHFR is shown in Scheme 10.6 (Benkovic *et al.*, 1988; Fierke *et al.*, 1987). The abbreviations used in Scheme 10.3 are as follows: E is the DHFR enzyme; F and FH are dihydrofolate and tetrahydrofolate, respectively; N and NH are $NADP^+$ and NADPH, respectively; the symbol E_N^{FH} stands for the ternary molecular complex E · N · FH; and the numbers above each arrow represent microscopic rate constant (e.g., "1" stands for k_1). All 22 microscopic rate constants appearing in Scheme 10.6 have been determined in a large number of independent experiment (Table 10.1).

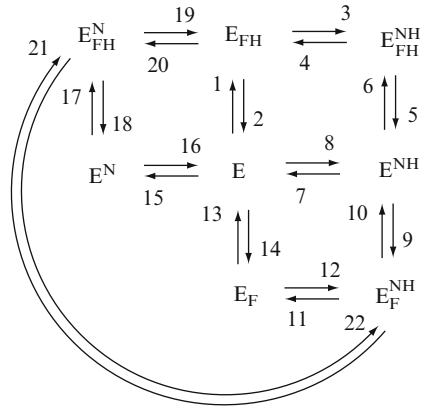
The reaction mechanism in Scheme 10.6 contains six thermodynamic boxes that do not involve the reversible chemical step (rate constants k_{21} and k_{22}). For example, moving clockwise or counterclockwise along the lower right box in Scheme 10.6, we expect that the product $k_8 \times k_9 \times k_{11} \times k_{13} = 27{,}200$ be numerically equal to $k_7 \times k_{10} \times k_{12} \times k_{14} = 28{,}000$. The corresponding equilibrium constant $K = k_8k_9k_{11}k_{13}/k_7k_{10}k_{12}k_{14} = 0.97$



Scheme 10.5

Table 10.1 Microscopic rate constants in the catalytic mechanism of *E. coli* DHFR (Benkovic *et al.*, 1988)

k_1	25	$\mu M^{-1}\ s^{-1}$	k_2	1.4	s^{-1}
k_3	8	$\mu M^{-1}\ s^{-1}$	k_4	85	s^{-1}
k_5	12.5	s^{-1}	k_6	2	$\mu M^{-1}\ s^{-1}$
k_7	3.5	s^{-1}	k_8	20	$\mu M^{-1}\ s^{-1}$
k_9	40	$\mu M^{-1}\ s^{-1}$	k_{10}	40	s^{-1}
k_{11}	1.7	s^{-1}	k_{12}	5	$\mu M^{-1}\ s^{-1}$
k_{13}	20	s^{-1}	k_{14}	40	$\mu M^{-1}\ s^{-1}$
k_{15}	13	$\mu M^{-1}\ s^{-1}$	k_{16}	300	s^{-1}
k_{17}	25	$\mu M^{-1}\ s^{-1}$	k_{18}	2.4	s^{-1}
k_{19}	200	s^{-1}	k_{20}	5	$\mu M^{-1}\ s^{-1}$
k_{21}	950	s^{-1}	k_{22}	0.6	s^{-1}



Scheme 10.6

is indeed very nearly equal to unity. The same is true for all five remaining thermodynamic boxes, including the largest (not counting the chemical step) box defined by the path $E^N_{FH} \rightarrow E_{FH} \rightarrow E^{NH}_{FH} \rightarrow E^{NH} \rightarrow E^{NH}_F \rightarrow E_F \rightarrow E \rightarrow E^N \rightarrow E^N_{FH}$.

The steady-state initial rate equation for DHFR, based on the comprehensive mechanism in Scheme 10.6 and derived by using the King–Altman method (King and Altman, 1956), contains 33 algebraic terms in the numerator, 65 algebraic terms in the denominator, and up to cubic exponents for concentrations. When printed in a page layout required by this volume, the single algebraic rate equation for DHFR would occupy approximately 20 printed pages (results not shown). A quote from Segel's seminal text, discussing the "Bi Bi Random Steady-State" mechanism, is also applicable to DHFR.

> The [initial rate] equation does not describe a hyperbola and, theoretically, the reciprocal plots are not linear, unless one substrate is saturating. [...] The groups of rate constants cannot be combined into convenient kinetic constants [Michaelis constants and inhibition constants]. (Segel, 1975, p. 647)

In DynaFit, we can now represent the same initial rate law, under the steady-state approximation, by entering the following text:

```
[task]
    data = rates
    approximation = steady-state
[reaction] | F + NH <==> FH + N
[enzyme] | E
[mechanism]
       E + FH <==> E.FH : k1 k2
E.FH + NH <==> E.FH.NH : k3 k4
E.FH.NH <==> E.NH + FH : k5 k6
       E.NH <==> E + NH : k7 k8
  E.NH + F <==> E.F.NH : k9 k10
  E.F.NH <==> E.F + NH : k11 k12
        E.F <==> E + F : k13 k14
        E + N <==> E.N : k15 k16
  E.N + FH <==> E.FH.N : k17 k18
  E.FH.N <==> E.FH + N : k19 k20
   E.F.NH <==> E.FH.N : k21 k22
```

The steady-state rate initial law derived internally by DynaFit consists of a system of simultaneous nonlinear algebraic equations evaluated numerically (Kuzmič, 2009a), by using the multidimensional Newton–Raphson method (Press *et al.*, 1992, p. 379). Unlike the traditional algebraic formalism (Segel, 1975), the numerical formalism utilized by DynaFit does not make the simplifying assumption that all reactant and modifier concentrations are essentially infinitely larger than the enzyme concentration.

Given the values of rate constants associated with the mechanism in Scheme 10.6, DynaFit was used to simulate initial reaction rates while varying the concentration of dihydrofolate and $NADPH^+$ (Fig. 10.3). The substrate saturation curves at relatively low dihydrofolate concentrations are expected to display a local maximum, followed by a decrease to an asymptotically saturating value.

Importantly, DynaFit can now properly take into account the presence of thermodynamic boxes in the DHFR mechanism, in order to constrain certain rate constants based on the values of other rate constants. For example, to express the constraint $k_7 = k_8 k_9 k_{11} k_{13} / k_{10} k_{12} k_{14}$, and a similar constraint for rate constant k_2, we would use the following DynaFit input:

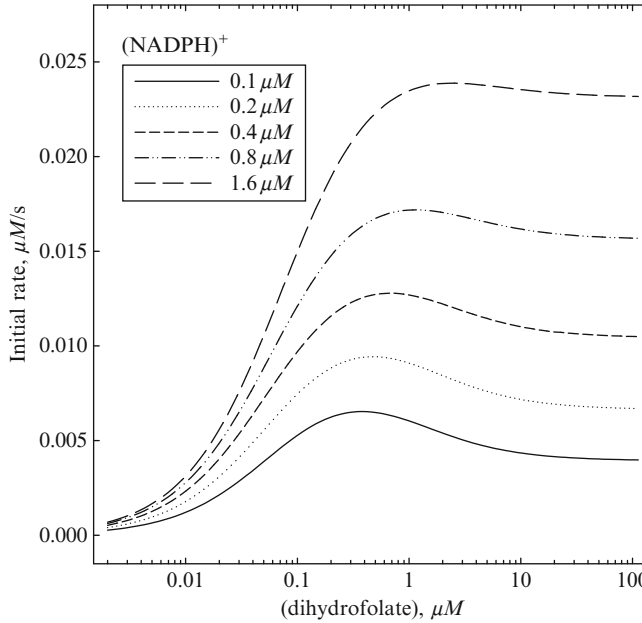


Figure 10.3 Simulated initial reaction rates for DHFR, based on the mechanism in Scheme 10.6 (Benkovic *et al.*, 1988; Fierke *et al.*, 1987) and rate constant values listed in Table 10.1.

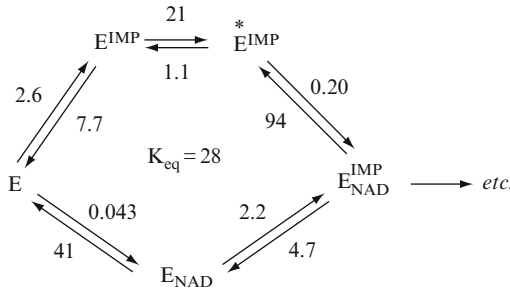
```
[constants]
    k2 = (k1 k3 k5 k7) / (k4 k6 k8)
    k7 = (k8 k9 k11 k13) / (k10 k12 k14)
```

Given that the DHFR mechanism contains six thermodynamic boxes for the noncovalent binding and dissociation steps, and also given that each thermodynamic box involves between 8 and 16 microscopic rate constants, many logically equivalent ways are available to place overall constraints on the kinetic model. Which particular rate constants should be constrained in DynaFit models need to be carefully evaluated on a case-by-case basis.

A casual survey of the published biochemical literature reveals occasional violations of the thermodynamic box rule.

For example, Digits and Hedstrom (1999) presented a kinetic model for inosine monophosphate (IMP) dehydrogenase interacting with IMP and NAD^+ (Scheme 10.7). In Scheme 10.7 (Digits and Hedstrom, 1999), the numerical values of all monomolecular rate constants are in s^{-1} units, and the bimolecular association rate constants are shown in $\mu M^{-1}\ s^{-1}$ units.

The salient feature of the mechanism in Scheme 10.7 is that the enzyme–IMP complex undergoes an isomerization before cofactor binding. Clearly, the overall equilibrium constant for the noncovalent interactions is significantly different from unity, which means that at least one rate constant in



Scheme 10.7

the postulated kinetic model is in error. This error was corrected in a later report (Schlippe *et al.*, 2004), where the kinetic mechanism was further developed using DynaFit.

The explanation for the inconsistency in Scheme 10.7 (L. Hedstrom, personal communication) lies in that the equilibria for formation of the ternary complexes were determined by measuring binding to the binary complexes of an *inactive mutant*. The relevant mutation (Cys to Ala) perturbs IMP binding in the binary complex, so IMP binding to the E.NAD complex is probably also perturbed and therefore unlikely to mimic the wild-type enzyme. Nevertheless, the measured values were utilized in the postulated reaction scheme. As a general warning, when using inactive mutants to infer rate constants in a similar fashion, special attention must be paid to the consistency of thermodynamic boxes.

A similar inconsistency in a noncovalent binding mechanism is present in a DynaFit study of the plasma membrane calcium pump isoform 4b by calmodulin (Penheiter *et al.*, 2003). The recent addition of a thermodynamic box checking feature into DynaFit should prevent similar inconsistencies occasionally cropping up in the published literature.

4. Time Course of Enzyme Reactions

DynaFit (Kuzmič, 1996) was initially developed to process the time course of "slow, tight" (Morrison and Walsh, 1988; Szedlacsek and Duggleby, 1995; Williams and Morrison, 1979) enzyme inhibition assays. In the intervening period, a number of features and capabilities had been added to further facilitate the analysis of reaction dynamics. For example, DynaFit can now be used to analyze "double-mixing" stopped-flow experiments (Williams *et al.*, 2004). Microscopic rate constants can be constrained with respect to statistical factors (see Section 2.2) or thermodynamic boxes (Section 3.1), or defined as fixed ratios where equilibrium constants are

known from independent experiments. This section describes another representative example of such recently added capabilities.

4.1. Invariant concentrations of reactants

Under highly specialized experimental circumstances, or for the purpose of modeling an *in vivo* biochemical system, DynaFit can now be used to simulate or fit experimental data under the assumption that the concentrations of certain reactants remain invariant, even as they participate in the underlying reaction mechanism. The corresponding DynaFit notation is to use the exclamation mark:

```
[concentrations]
    Substrate = 1.2345 !
```

4.1.1. SPR on-chip enzyme kinetics

The invariant concentration technique had been utilized in building a preliminary mathematical model for on-chip kinetics of transglucosidase alter-nansucrase (E.C. 2.4.1.140) from *Leuconostoc mesenteroides* NRRL B-1355 (Clé *et al.*, 2008, 2010). This enzyme catalyzes the transfer of glucose from sucrose to acceptors at their nonreducing ends. In this particular case, the acceptor was a carboxymethyl dextran surface on a surface plasmon resonance (SPR) chip.

When sucrose solution mixed with the transglucosidase enzyme is flowed over the SPR chip, the dextran oligomer chains on the chip's surface are extended with additional glucose moieties, and this process can be monitored by SPR. Importantly, the bulk sucrose concentration does not change over time, because it is being replenished by the continuous flow.

A typical SPR sensorgram of the enzyme-catalyzed extension of a dextran surface is shown in Fig. 10.4 (see Clé *et al.*, 2010 for details). The important portion of the DynaFit script used in this analysis is shown below. Note that the enzyme–sucrose ("S") association is made irreversible in the postulated mechanism. The reasons for choosing this simplified Van Slyke–Cullen kinetic model (Slyke and Cullen, 1914) are explained in a separate report (Kuzmič, 2009b).

```
[task]
    data = progress
    task = fit
[mechanism]
    E + dextran <==> E.dextran : k1 k2
    E.dextran + S ---> E.dextran. S : k3
    E.dextran. S ---> E.dextran + P : k4
[concentrations]
    E = 0.18 ! ; invariant
    S = 11700 ! ; invariant
    dextran = 0.00002
```

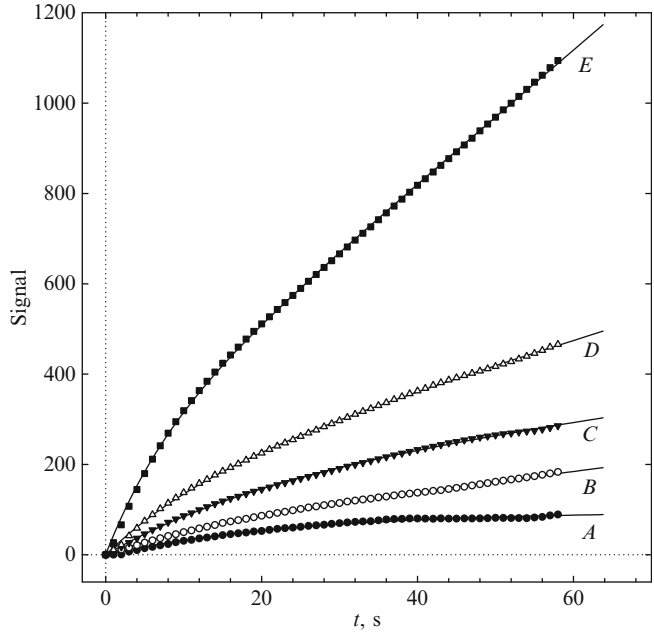


Figure 10.4 SPR sensorgram of the enzyme-catalyzed extension of a dextran surface. Transglucosidase alternansucrase at various concentrations was coinjected with sucrose (11.7 m*M*) over the surface of the SPR chip. Curves A–E: enzyme concentration $[E]_0 = 0.018$, 0.022, 0.03, 0.044, and 0.09 μM, respectively.

The surface catalysis phenomena involved, for example, in starch biosynthesis and in cellulose degradation are still relatively poorly understood. The significance of the on-chip enzyme kinetics experiment is that it can potentially shed light on biologically relevant heterogeneous phase processes.

At this preliminary phase of the investigation, the best-fit values of microscopic rate constants (not shown) were obtained separately for each recorded progress curve. The goal of the ongoing research is to produce a global (Beechem, 1992) mathematical model for the on-chip kinetics.

5. General Methods and Algorithms

This section briefly summarizes selected features and capabilities added to the DynaFit software package since its original publication (Kuzmič, 1996). These general algorithms are applicable to all types of experimental data (progress curves, initial rates, and complex equilibria) being analyzed.

This selection of added features is not exhaustive, but it emphasizes some of the most difficult tasks in the analysis of biochemical data:

- How do we know where to start (the initial estimate problem);
- How do we know whether the best-fit parameters are good enough (the confidence interval problem); and
- How do we know which fitting model to choose among several alternatives (the model discrimination problem).

5.1. Initial estimates of model parameters

One of the most difficult tasks of a data analyst performing nonlinear least-squares regression is to come up with initial estimates of model parameters that are sufficiently close to the true values. If the initial estimate of rate or equilibrium constants is not sufficiently accurate, the data-fitting algorithm might converge to a local minimum, or not converge at all. This is the nature of the Levenberg–Marquardt algorithm (Marquardt, 1963; Reich, 1992), which is the main least-squares minimization algorithm used by DynaFit.

The updated DynaFit software offers two different methods to avoid local minima on the least-squares hypersurface, that is, to avoid incorrect "best-fit" values of rate constants and other model parameters. The first method relies on a brute-force systematic parameter scan, and the second method uses ideas from evolutionary computing.

5.1.1. Systematic parameter scan

To increase the probability that a true global minimum is found for all rate and equilibrium constants, DynaFit allows the investigator to specify a set of alternate initial estimates. The software then generates all possible combinations of starting values, and performs the corresponding number of independent least-squares regressions. The results are ranked by the residual sum of squares.

For example, let us assume that the postulated mechanism includes four adjustable rate constants, k_1–k_4, and that we wish to examine four different starting values (spaced by a factor of 10) for each of them. The requisite DynaFit code would read as follows:

```
[constants]
    k1 = { 0.01, 0.1, 1, 10} ?
    k2 = {0.001, 0.01, 0.1, 1} ?
    k3 = {0.001, 0.01, 0.1, 1} ?
    k4 = {0.001, 1, 1000, 1000000} ?
```

In this case, the program would perform $4^4 = 256$ separate least-squares minimizations, starting from 256 different combinations of initial estimates. In extreme cases, the execution time required for such systematic parameter scans might reach many minutes or even hours by using the currently

available computing technology. However, for critically important data analyses, avoiding local minima and therefore incorrect mechanistic conclusions should be worth the wait.

5.1.2. Global minimization by differential evolution

As an alternate solution to the problem of local minima in least-squares regression analysis, DynaFit now uses the differential evolution (DE) (Price *et al.*, 2005) algorithm. DE belongs to the family of stochastic evolutionary strategy (ES) algorithms, which attempt to find a global sum-of-squares minimum by using ideas from evolutionary biology.

The essential feature of any ES data-fitting algorithm is that it starts from a large number of simultaneous, randomly chosen initial estimates for all adjustable model parameters. The algorithm then evolves this population of "organisms," by allowing only the sufficiently fit population members to "sexually reproduce." In this case, by fitness we mean the sum of squares associated with each particular combination of rate constants and other model parameters (the genotype). By sexual reproduction, we mean that selected population members have their genome (i.e., model parameters) carried over into the next generation by using Nature's usual tricks—chromosomal crossover accompanied by random mutations.

There are many variations on the ES computational scheme, and also a growing number of variants of the DE algorithm itself. The interested reader is encouraged to examine several recently published books and monographs (Chakraborty, 2008; Feoktistov, 2008; Onwubolu and Davendra, 2009; Price *et al.*, 2005) for details. Typically, the number of population members does not change through the evolutionary process, meaning that if we start with 1000 different initial estimates for each rate constant, we also have 1000 different estimates at the end, after a large number of generations have reproduced. Importantly, while we might start with a population of 1000 estimates spanning 12 or 18 orders of magnitude for each rate constant, the hope is that we end with 1000 estimates *all of which* are close to the best possible value.

The performance of the DE algorithm (Price *et al.*, 2005), as implemented in DynaFit, is illustrated by using an example involving irreversible inhibition kinetics of the HIV protease. This particular test problem was first presented in the original DynaFit publication (Kuzmič, 1996), and was subsequently reused by Mendes and Kell (1998) to test the performance of the popular software package Gepasi. The simulation software package COPASI (Hoops *et al.*, 2006), a direct descendant of Gepasi, is also being profiled in this volume.

Figure 10.5 displays fluorescence changes during a fluorogenic assay (Kuzmič *et al.*, 1996; Peranteau *et al.*, 1995) of the HIV protease. The nominal enzyme concentration was 4 nM in each of the five kinetic experiments; the nominal substrate concentration was 25 μM; the inhibitor

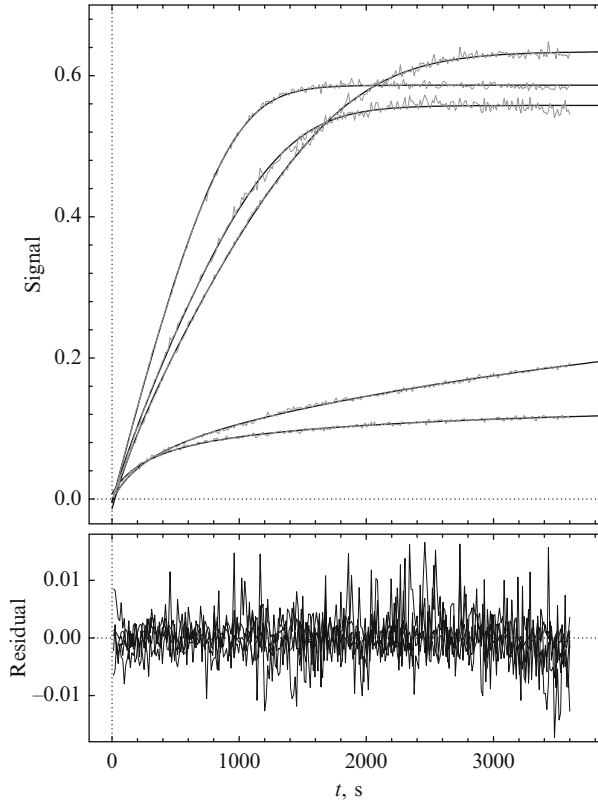
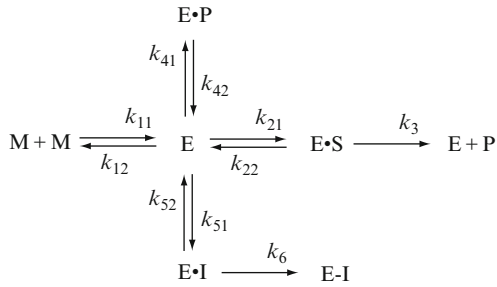


Figure 10.5 Least-squares fit of progress curves from HIV protease in the presence of an irreversible inhibitor. Results of the best-fit were obtained by using the differential evolution algorithm (Price *et al.*, 2005).

concentrations (curves from top to bottom; Fig. 10.5) were 0, 1.5, 3, and 4 nM (two experiments). As is discussed elsewhere (Kuzmič, 1996), each initial enzyme and substrate concentration was treated as an adjustable parameter. The vertical offset on the signal axis was also treated as an adjustable parameter for each experiment separately.

The mechanistic model is shown in Scheme 10.8, where M is the monomer subunit of the HIV protease. The numbering of rate constants in Scheme 10.8 was chosen to match a previous report (Mendes and Kell, 1998). The dimensions used throughout the analysis (see also final results in Table 10.2) were μM for all concentrations, $\mu M^{-1}\ s^{-1}$ for all second-order rate constants, and s^{-1} for all first-order rate constants. The rate constants $k_{11} = 0.1$, $k_{12} = 0.0001$, and $k_{21} = k_{41} = k_{51} = 100$ were treated as fixed parameters in the model, whereas the rate constants k_{22}, k_3, k_{42}, k_{52}, and k_6



Scheme 10.8

were treated as adjustable parameters. To match the Gepasi test (Mendes and Kell, 1998) using the same example problem, each rate constant was constrained to remain less than 10^5 in absolute value. In the course of the DE optimization, rate constants were allowed to span 12 orders of magnitude (between 10^{-7} and 10^5). Each adjustable concentration was allowed to vary within 50% of its nominal value. An excerpt from a requisite DynaFit script input file is shown as follows:

```
[task]
    data = progress
    task = fit
    algorithm = differential-evolution
[mechanism]
    M + M <==> E : k11 k12
    E + S <==> ES : k21 k22
    ES ---> E + P : k3
    E + P <==> EP : k41 k42
    E + I <==> EI : k51 k52
    EI --> EJ : k6
[constants]
    k11 = 0.1
    k12 = 0.0001
    k21 = 100
    k22 = 300 ? (0.0000001 .. 100000)
    k3 = 10 ? (0.0000001 .. 100000)
    k41 = 100
    k42 = 500 ? (0.0000001 .. 100000)
    k51 = 100
    k52 = 0.1 ? (0.0000001 .. 100000)
    k6 = 0.1 ? (0.0000001 .. 100000)
```

Table 10.2 Least-squares fit of HIV protease inhibition data shown in Fig. 10.5: Comparison of the simulated annealing (SA) algorithm as implemented in Gepasi (Mendes and Kell, 1998) and COPASI (Hoops *et al.*, 2006) with the differential evolution (DE) algorithm as implemented in DynaFit (Kuzmič, 1996)

Parameter	SA (Mendes and Kell, 1998)	SA (this work)[a]	DE	SA/DE
k_{22}	201.1	273.1	23.67	11.54
k_3	7.352	6.517	3.922	1.66
k_{42}	1171	1989	128.2	15.51
k_{52}	13,140	11,120	0.00008562	130,000,000
k_6	30,000	4453	0.0004599	9,700,000
$[S]_1$	24.79	24.74	24.65	1.00
$[S]_2$	23.43	23.46	23.37	1.00
$[S]_3$	26.79	26.99	26.99	1.00
$[S]_4$	32.10	20.92	14.39	1.45
$[S]_5$	26.81	17.59	16.04	1.10
$[E]_1$	0.004389	0.005029	0.007484	0.67
$[E]_2$	0.004537	0.004965	0.006568	0.76
$[E]_3$	0.005470	0.005796	0.007116	0.81
$[E]_4$	0.004175	0.004238	0.004221	1.00
$[E]_5$	0.003971	0.003980	0.003396	1.17
Δ_1	−0.00801	−0.00712	−0.00508	1.40
Δ_2	−0.00391	−0.00490	−0.00289	1.69
Δ_3	−0.00896	−0.01395	−0.01354	1.03
Δ_4	−0.01600	−0.01192	−0.00337	3.54
Δ_5	−0.00379	0.00005	0.00777	0.01
Iterations	630,000	1,025,242	–[b]	–[b]
Sum of squares	0.0211024	0.0201911	0.0194526	1.04
Run time (h)	–[c]	16.5[d,e]	1.1[e]	15

[a] Software Gepasi (Mendes and Kell, 1998) ver. 3.30.
[b] Iteration counts in SA and DE are not compatible.
[c] Running time not given in the original publication.
[d] Interrupted.
[e] Intel® Core™2 Duo T7400 microprocessor (2.16 GHz, 667 MHz bus, 4 MB cache).

DynaFit automatically chooses the population size, based on the number of adjustable model parameters, and on the range of values they are allowed to span. In this case, the DE algorithm started with 259 separate estimates for each of the 15 adjustable model parameters (five rate constants, five locally adjusted substrate and enzyme concentrations, and five offsets on the signal axis). A representative histogram of distribution for one of the 15 adjustable

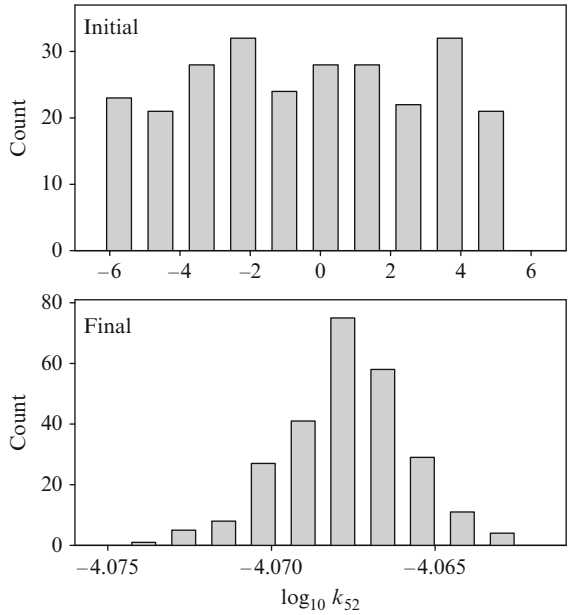


Figure 10.6 The initial and final distribution of the rate constant k_{52} in the differential evolution (Price *et al.*, 2005) fit of HIV protease inhibition data shown in Fig. 10.5. The population contained 259 members.

model parameters (the rate constant k_{52}) is shown in the upper panel of Fig. 10.6. Note that the 259 initial estimates of the rate constant k_{52} span 12 orders of magnitude. The initial random distribution of parameter values is *uniform* (as opposed to Gaussian or similarly bell-shaped) on the logarithmic scale.

The swarm of 259 "organisms," each carrying a unique combination of 15 adjustable model parameters (the genotype), was allowed to evolve using the Darwinian evolutionary principles (selection by fitness; chromosomal crossover during the "mating" of population members; random genetic mutations). After 793 generations, each of the 15 model parameters converged to a relatively narrow range of values, as shown in the bottom panel of Fig. 10.6 for the rate constant k_{52}. The simulated best-fit model is shown as smooth curves in Fig. 10.5. The best-fit values of adjustable model parameters are shown in Table 10.2, where Δ_i is offset on the signal axis for individual data sets.

The simulated annealing (SA) algorithm (Corana *et al.*, 1987; Kirkpatrick *et al.*, 1983) was chosen for comparison with DE, because it appears to be the best performing global optimization method currently reported in the biochemical literature (Mendes and Kell, 1998).

The results listed in Table 10.2 show that the DE algorithm found a combination of model parameters that lead to a significantly lower sum of squares (i.e., a better fit) compared to the SA algorithm. Some model parameters, such as the adjustable substrate concentrations, were very close to identical in both data-fitting methods. Other model parameters, such as the rate constants k_{52} and k_6 that characterize the inhibitor properly, differed by 6–8 orders of magnitude. The SA algorithm had to be terminated manually after approximately 17 h of continued execution, and more than one million iterations. The DE algorithm terminated automatically after 66 min, when defined convergence criteria were satisfied.

We can conclude that, in the specific case of the HIV protease irreversible kinetics, the DE global optimization algorithm clearly performs significantly better than the SA algorithm. However, this does *not* mean that the best-fit DE parameter values listed in Table 10.2 are any closer to the true values, when compared with the SA parameters. In fact, it appears that *neither* set of parameter values should be regarded with much confidence (see Section 5.2). Probably, the only conclusion we can make safely is that very much more research is needed into the relative merits of global optimization algorithms such as DE and SA—specifically, as they are applied to the analysis of biochemical kinetic data.

5.2. Uncertainty of model parameters

Most biochemists are likely to see the uncertainty of kinetic model parameters expressed only as formal standard errors. Formal standard errors are the plus-or-minus values standing next to the best-fit values of nonlinear parameters, as reported by all popular software packages for nonlinear least-squares regression, including DynaFit. However, it should be strongly emphasized that *formal standard errors can (and usually do) grossly underestimate the statistical uncertainty*. For a rigorous theoretical treatment of statistical inference regions for nonlinear parameters, see Bates and Watts (1988).

Johnson *et al.* (2009) recently stated that DynaFit (Kuzmič, 1996) users are provided only with the "standard errors [...] without additional aids to evaluate the extent to which the fitted parameters are actually constrained by the data." This statement is factually false, and needs to be corrected for the record. Since version 2.23 released in January 1997 and extensively documented in the freely distributed user manual, DynaFit has always implemented the *profile-t* search method of Bates and Watts (Bates and Watts, 1988; Brooks *et al.*, 1994; Watts, 1994) to compute approximate inference regions of nonlinear model parameters.

The most recent update to DynaFit adds an *additional* aid to evaluate the extent to which the fitted parameters are constrained by the data. This is a particular modification of the well-established Monte-Carlo method (Straume and Johnson, 1992).

5.2.1. Monte-Carlo confidence intervals

The Monte-Carlo method (Straume and Johnson, 1992) for the determination of confidence intervals is based on the following idea. After an initial least-squares fit using the usual procedure, the best-fit values of nonlinear parameters are used to simulate many (typically, at least 1000) artificial data sets. The idealized theoretical model curves (e.g., the smooth curves in Fig. 10.5) are always the same, but the superimposed pseudo-random noise is different every time. The 1000 slightly different sets of pseudo-experimental data are again subjected to nonlinear least-squares regression. In the end, the 1000 different sets of best-fit values for model parameters are tallied up to construct a *histogram* of the parameter distribution. The range of values spanned by each histogram is the Monte-Carlo confidence interval for the given model parameter.

''Shuffle'' and ''shift'' Monte-Carlo methods A crucially important part of the above Monte-Carlo procedure is the simulation of the pseudo-random noise to be superimposed on the idealized data. How should we choose the statistical distribution, from which the pseudo-random noise is drawn? Usually, it is assumed that the pseudo-random experimental noise has Normal or Gaussian distribution (Straume and Johnson, 1992), and that the individual data points are statistically independent or uncorrelated. If so, the standard deviation of this Gaussian distribution (the half-width of the requisite bell curve) can be taken as the standard error of fit from the first-pass regression analysis of the original data. However, we have recently demonstrated (Kuzmič *et al.*, 2009) that experimental data points recorded in at least one particular enzyme assay are *not* statistically independent. Instead, we see a strong neighborhood correlation among adjacent data points—spanning up to six nearest neighbors.

To reflect the possible serial correlation among nearby data points, DynaFit (Kuzmič, 1996) now allows two variants of the Monte-Carlo method, which could be called the "shift'' Monte-Carlo and "shuffle'' Monte-Carlo algorithms. In both cases, instead of generating presumably Gaussian errors to be superimposed on the idealized data, we merely *rearrange the order* of the actual residuals generated by the first-pass least-squares fit. In the shuffle variant, the residuals are reused in truly randomized order. In the shift variant of the Monte-Carlo algorithm, the order of the residuals is preserved, but the starting position changes.

For example, let us assume that a particular reaction progress curve (such as one of those shown in Fig. 10.5) contains 300 experimental data points. After the first-pass least-squares fit, we could simulate up to 300 synthetic progress curves by superimposing the ordered sequence of residuals. In one such simulated curve, the first synthetic data point would be assigned

residual No. 17, the second data point residual No. 18, and so on. At the end of the ordered sequence of residuals, we wrap around to the beginning (i.e., data point No. 300 − 17 = 283 will receive residual No. 1). In another simulated curve, the first data point would be generated from residual No. 213, the second data point from residual No. 214, and so on.

The practical usefulness of the shift and shuffle variants of the Monte-Carlo method (Straume and Johnson, 1992) is that it avoids having to make assumptions about the statistical distribution (Gaussian, Lorentzian, etc.) of the random noise that is inevitably present in the experimental data. Interestingly, the original conception of the Monte-Carlo method (Dwass, 1957; Nichols and Holmes, 2001) was, in fact, based on *permuting* existing population members, rather than making distributional assumptions.

Two-dimensional histograms The "shift" Monte-Carlo confidence intervals for rate constants k_{22}, k_3, and k_{42} from the least-squares fit of HIV protease inhibition data are shown in Fig. 10.7. The best-fit values of each model parameter are marked with a filled triangle. The rate constant k_3 is characterized by a relatively narrow confidence intervals (spanning from approximately 3 to 9 s^{-1}). In contrast, the Monte-Carlo confidence intervals for rate constants k_{22} and k_{42} not only are much wider (approximately 4 orders of magnitude for k_{42}) but also are clearly bi-modal. The appearance of such double-hump histogram for any parameter is a strong indication that (a) the model is probably severely over-parameterized, and (b) the data could very likely be fit to at least two alternate mechanisms.

In order to better diagnose possible statistical coupling between pairs of rate constants, beyond what conventional Monte-Carlo histograms can provide, DynaFit now produces *two-dimensional histograms* such as those shown in Fig. 10.8. The thin solid path enclosing each histogram in Fig. 10.8 is the *convex hull*—the shortest path entirely enclosing a set of points in a plane. The approximate area occupied by the convex hull is a useful empirical measure of parameter redundancy.

If any two rate constants were truly statistically independent, the corresponding two-dimensional Monte-Carlo histogram plot would resemble a circular area with the highest population density appearing in the center. We can see in Fig. 10.8 the rate constants k_{22} and k_{42} are clearly correlated, as is indicated by the elongated crescent shape of the two-dimensional histogram.

In summary, with regard to assessing the statistical uncertainty of non-linear model parameters, DynaFit (Kuzmič, 1996) has always allowed the investigator to perform the full search in parameter space, using the *profile-t* method (Bates and Watts, 1988; Brooks *et al.*, 1994; Watts, 1994). As a result of such detailed analysis, the investigator often must face the unpleasant fact that the confidence regions for rate constants, equilibrium constants, or derived kinetic parameters (e.g., Michaelis constants) not only are much

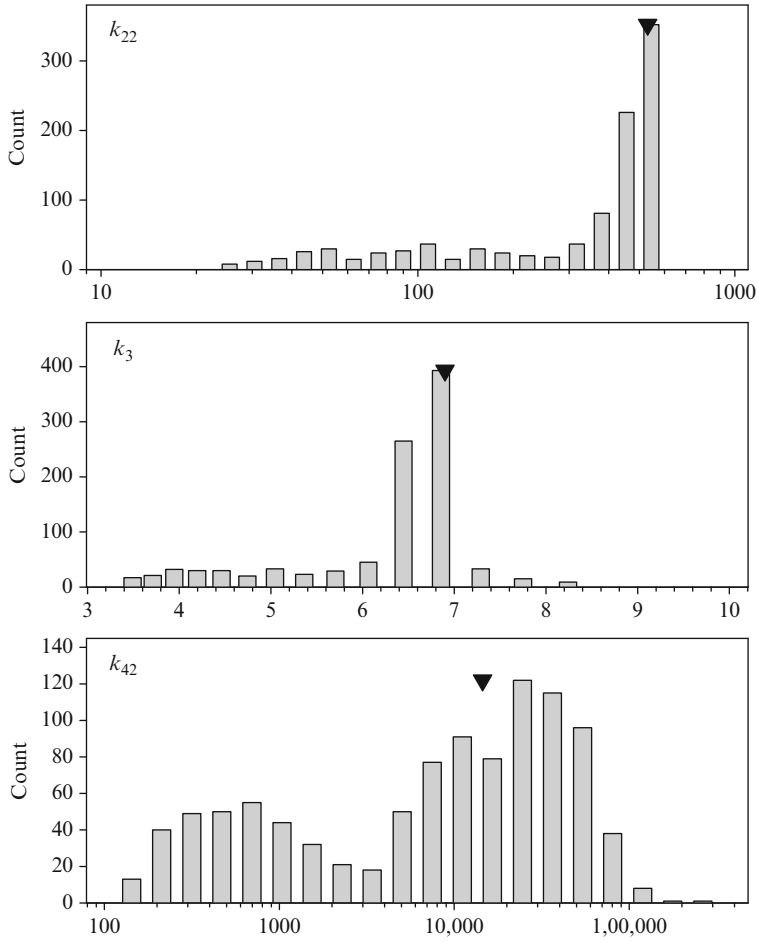


Figure 10.7 Monte-Carlo confidence intervals for model parameters: Distribution histograms for rate constants k_{22}, k_3, and k_{42} from least-squares fit of HIV protease inhibition data shown in Fig. 10.5.

larger than the formal standard errors would suggest, but perhaps also larger than would appear "publishable."

However, it must be strongly emphasized that the formal standard errors for nonlinear parameters reported by DynaFit should never be given much credence. The program reports them mostly for compatibility with other software package typically used by biochemists. In order to obtain a more realistic interpretation of the experimental data, DynaFit users are encouraged to go beyond formal standard errors, and utilize both the previously available *profile-t* method (Brooks *et al.*, 1994), and now also the modified Monte-Carlo method (Straume and Johnson, 1992).

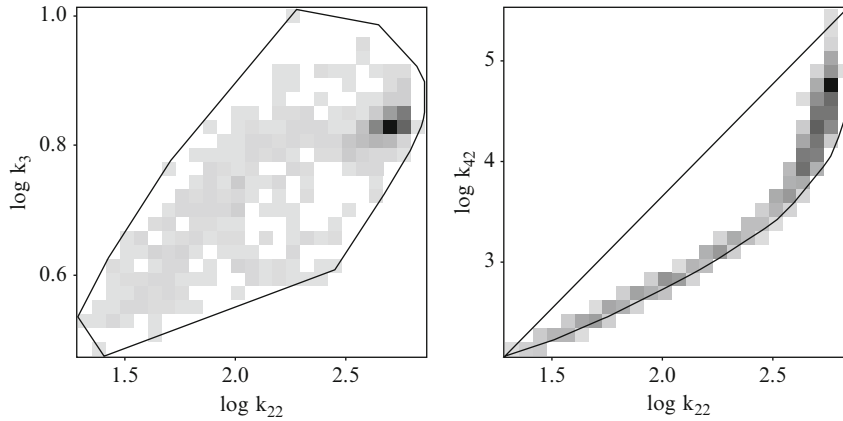


Figure 10.8 Monte-Carlo confidence intervals for model parameters: Two-dimensional correlation histograms for rate constants k_{22} versus k_3 (*left*; uncorrelated) and k_{22} versus k_{42} (*right*; strong correlation) from least-squares fit of HIV protease inhibition data shown in Fig. 10.5.

5.3. Model-discrimination analysis

The problem of selecting the most plausible theoretical model among several candidates (e.g., deciding whether a given enzyme inhibitor is competitive, noncompetitive, or mixed-type) represents one of the most challenging tasks facing the data analyst. Myung and Pitt (2004) and Myung *et al.* (2009) reviewed recent developments in earlier volumes of this series. This section contains only a very brief summary of the model-discrimination features available in DynaFit (Kuzmič, 1996). The reader is referred to the full program documentation available online (http://www.biokin.com/dynafit/).

DynaFit (Kuzmič, 1996) currently offers two distinct methods for statistical model discrimination. First, for nested fitting models, the updated version of DynaFit continues to offer the *F*-statistic method previously discussed by Mannervik (1981, 1982) and many others. Secondly, for any group of alternate models, whether nested or nonnested, DynaFit uses the second-order AIC_c (Burnham and Anderson, 2002) to perform model discrimination.

Briefly, the AIC_c criterion is defined by Eq. (10.1), where S is the residual sum of squares; n_P is the number of adjustable model parameters; and n_D is the number of experimental data points. For each candidate model in a collection of alternate models, DynaFit computes ΔAIC_c as the difference between AIC_c for the particular model, and the AIC_c for the best model (with the lowest value of AIC_c). Thus, the best model is by definition assigned $\Delta AIC_c = 0$. The *Akaike weight*, w_i, for the ith model in a collection of m alternatives, is defined by Eq. (10.2):

$$\mathrm{AIC_c} = -\log S + 2n_\mathrm{P} + \frac{2n_\mathrm{P}(n_\mathrm{P}+1)}{n_\mathrm{D} - n_\mathrm{P} - 1} \tag{10.1}$$

$$w_i = \frac{\exp\left(\frac{1}{2}\Delta\mathrm{AIC_c}^{(i)}\right)}{\sum_{i=1}^{m}\exp\left(\frac{1}{2}\Delta\mathrm{AIC_c}^{(i)}\right)} \tag{10.2}$$

Burnham and Anderson (2002) formulated a series of empirical rules for interpreting the observed $\Delta\mathrm{AIC_c}$ values for each alternate fitting model, stating that $\Delta\mathrm{AIC_c} > 10$ might be considered a sufficiently strong evidence against the given model.

Practical experience with the Burnham and Anderson rule suggests that it is applicable only when the number of experimental data points is a reasonably small multiple of the number of adjustable model parameters (e.g., $n_\mathrm{D} < 20 \times n_\mathrm{P}$). In some cases, the number of data points is very much larger. For example, in certain continuous assays or stopped-flow measurements, it is not unusual to collect thousands of experimental data points in order to determine two or three kinetic constants. In such cases, the $\Delta\mathrm{AIC_c} > 10$ rule has been found unreliable. In general, a candidate model should probably be rejected only if its Akaike weight, w_i, is smaller than approximately 0.001.

The DynaFit notation needed to compare a series of alternate models, and to select the most plausible model if a selection is possible, is illustrated on the following input file fragment. Please note the use of question marks after each (arbitrarily chosen) model name. This notation instructs DynaFit to evaluate the plausibility of the given model, in comparison with other models that are marked identically.

```
[task]
    model = Competitive ?
[mechanism]
    E + S <===> E.S : Ks  dissoc
      E.S ---> E + P : kcat
    E + I <===> E.I : Ki  dissoc
...
[task]
    model = Uncompetitive ?
[mechanism]
        E + S <===> E.S : Ks  dissoc
          E.S ---> E + P : kcat
    E.S + I <===> E.S.I : Kis  dissoc
...
[task]
    model = Mixed-type noncompetitive ?
```

```
[mechanism]
    E + S <===> E.S      : Ks  dissoc
    E.S ---> E + P       : kcat
    E + I <===> E.I      : Ki  dissoc
    E.S + I <===> E.S.I  : Kis  dissoc
...
[task]
    model = Partial mixed-type ?
[mechanism]
    E + S <===> E.S      : Ks  dissoc
    E.S ---> E + P       : kcat
    E + I <===> E.I      : Ki  dissoc
    E.S + I <===> E.S.I  : Kis  dissoc
    E.S.I ---> E.I + P   : kcat'
...
```

When DynaFit is presented with a series of alternate models in a similar way, it will fit the available experimental data to each postulated model in turn. After the last model in the series is fit to the data, the program presents to the user a summary table listing the values of ΔAIC_c. The AIC-based model discrimination feature available in DynaFit has been utilized in a number of reports (Błachut-Okrasinska *et al.*, 2007; Collom *et al.*, 2008; Gasa *et al.*, 2009; Jamakhandi *et al.*, 2007; Kuzmič *et al.*, 2006).

6. Concluding Remarks

DynaFit (Kuzmič, 1996) has proved quite useful in a number of projects, as is evidenced by the number of journal publications that cite the program. It is hoped that the software will continue to enable innovative research. This section offers a few closing comments on DynaFit enhancements currently in development.

6.1. Model discrimination analysis

The AIC criterion is based solely on the *number* of optimized parameters and the corresponding sum of squares. The degree of *uncertainty* associated with each particular set of model parameters is completely ignored. However, if two candidate models with exactly identical number of adjustable parameters hypothetically produced exactly identical sums of squares, but one of these models was associated with significantly narrower confidence regions, then this model should be preferred (Myung and Pitt, 2004). The minimum description length (MDL) also known as stochastic complexity (SC) measure (Myung and Pitt, 2004) would clearly be a more appropriate

model-discrimination criterion. Unfortunately, for technical reasons, the MDL criterion is extremely difficult to compute (Myung *et al.*, 2009). Investigations are currently ongoing into at least an approximate computation of the MDL/SC test.

6.2. Optimal design of experiments

Most biochemists—probably like most experimentalists—prefer to do the experiment first, then proceed to data analysis, and finally to publication. However, to paraphrase the eminent statistician G. E. P. Box (Box *et al.*, 1978), no amount of the most ingenious data analysis can salvage a poorly designed experiment. When examining the extant enzymological literature, one often wonders exactly how the concentrations were chosen. Why was an exponential series (1, 2, 4, 8, 16) used for substrate concentrations, instead of a linear series (3, 6, 9, 12, 15) (Kuzmič *et al.*, 2006)? Was it by design, or was it because "that's how we always did it"? Similar choices profoundly affect how much—if anything—can be learned from any given experiment. A well-established statistical theory of optimal experiment design (Atkinson and Donev, 1992; Fedorov, 1972) has been used by biochemical researchers in the past (Duggleby, 1981; Endrényi, 1981; Franco *et al.*, 1986). At the present time, DynaFit is being modified to implement these ideas, and deploy them for computer-assisted *rational design of experiments*.

ACKNOWLEDGMENTS

Klára Briknarová and Jill Bouchard (University of Montana) are gratefully acknowledged for sharing their as yet unpublished NMR titration data. Jan Antosiewicz (Warsaw University) provided stimulating discussions and procured the PNP inhibition data for testing the statistical-factors feature in DynaFit; the raw experimental data were made available by Beata Wielgus-Kutrowska, Agnieszka Bzowska, and Katarzyna Breer (Warsaw University). Stephen Bornemann and his colleagues (John Innes Center, Norwich) graciously invited me to peek into the mysteries of their unique SPR on-chip kinetic system, and inspired the development of the invariant concentration algorithm. Liz Hedstrom (Brandeis University) made helpful comments and suggestions. I am grateful to Andrei Ruckenstein (formerly of the BioMaPS Institute for Quantitative Biology, Rutgers University; currently at Boston University) for illuminating discussions regarding thermodynamic boxes in biochemical mechanisms. Sarah McCord (Massachusetts College of Pharmacy and Health Sciences) provided expert assistance in editing this manuscript.

REFERENCES

Atkinson, A., and Donev, A. (1992). Optimum Experimental Designs. Oxford University Press, Oxford.

Bates, D. M., and Watts, D. G. (1988). Nonlinear Regression Analysis and its Applications. Wiley, New York.

Beechem, J. M. (1992). Global analysis of biochemical and biophysical data. *Methods Enzymol.* **210,** 37–54.

Benkovic, S. J., Fierke, C. A., and Naylor, A. M. (1988). Insights into enzyme function from studies on mutants of dihydrofolate reductase. *Science* **239,** 1105–1110.

Błachut-Okrasinska, E., Bojarska, E., Stepiński, J., and Antosiewicz, J. (2007). Kinetics of binding the mRNA cap analogues to the translation initiation factor eIF4E under second-order reaction conditions. *Biophys. Chem.* **129,** 289–297.

Bosco, G., Baxa, M., and Sosnick, T. (2009). Metal binding kinetics of Bi-Histidine sites used in ψ analysis: Evidence of high-energy protein folding intermediates. *Biochemistry* **48,** 2950–2959.

Box, G. E. P., Hunter, W. G., Hunter, J. S., and Hunter, W. G. (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. John Wiley, New York.

Briknarová, K., Zhou, X., Satterthwait, A., Hoyt, D., Ely, K., and Huang, S. (2008). Structural studies of the SET domain from RIZ1 tumor suppressor. *Biochem. Biophys. Res. Commun.* **366,** 807–813.

Brooks, I., Watts, D., Soneson, K., and Hensley, P. (1994). Determining confidence intervals for parameters derived from analysis of equilibrium analytical ultracentrifugation data. *Methods Enzymol.* **240,** 459–478.

Burnham, K. B., and Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer-Verlag, New York.

Bzowska, A. (2002). Calf spleen purine nucleoside phosphorylase: Complex kinetic mechanism, hydrolysis of 7-methylguanosine, and oligomeric state in solution. *Bioch. Biophys. Acta* **1596,** 293–317.

Bzowska, A., Koellner, G., Stroh, B. W.-K. A., Raszewski, G., Holý, A., Steiner, T., and Frank, J. (2004). Crystal structure of calf spleen purine nucleoside phosphorylase with two full trimers in the asymmetric unit: Important implications for the mechanism of catalysis. *J. Mol. Biol.* **342,** 1015–1032.

Chakraborty, U. K. (2008). Advances in Differential Evolution. Springer-Verlag, New York.

Clé, C., Gunning, A. P., Syson, K., Bowater, L., Field, R. A., and Bornemann, S. (2008). Detection of transglucosidase-catalyzed polysaccharide synthesis on a surface in real-time using surface plasmon resonance spectroscopy. *J. Am. Chem. Soc.* **130,** 15234–15235.

Clé, C., Martin, C., Field, R. A., Kuzmič, P., and Bornemann, S. (2010). Detection of enzyme-catalyzed polysaccharide synthesis on surfaces. *Biocatal. Biotransform.* in press.

Collom, S. L., Laddusaw, R. M., Burch, A. M., Kuzmič, P., Grover, P. Miller, and Andand, M. D. P. (2008). CYP2E1 substrate inhibition: Mechanistic interpretation through an effector site for monocyclic compounds. *J. Biol. Chem.* **383,** 3487–3496.

Corana, A., Marchesi, M., Martini, C., and Ridella, S. (1987). Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Trans. Math. Softw.* **13,** 262–280.

Deng, Q., and Huang, S. (2004). PRDM5 is silenced in human cancers and has growth suppressive activities. *Oncogene* **17,** 4903–4910.

Digits, J. A., and Hedstrom, L. (1999). Kinetic mechanism of tritrichomonas foetus inosine 5′-monophosphate dehydrogenase. *Biochemistry* **38,** 2295–2306.

Duggleby, R. (1981). Experimental designs for the distribution free analysis of enzyme kinetic data. *In* "Kinetic Data Analysis" (L. Endrényi, ed.), pp. 169–181. Plenum Press, New York.

Dwass, M. (1957). Modified randomization tests for nonparametric hypotheses. *Ann. Math. Stat.* **28,** 181–187.

Endrényi, L. (1981). Design of experiments for estimating enzyme and pharmacokinetic parameters. *In* "Kinetic Data Analysis" (L. Endrényi, ed.), pp. 137–169. Plenum Press, New York.

Fedorov, V. (1972). Theory of Optimal Experiments. Academic Press, New York.
Feoktistov, V. (2008). Differential Evolution: In Search of Solutions. Springer-Verlag, New York.
Fierke, C. A., Johnson, K. A., and Benkovic, S. J. (1987). Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from *Escherichia coli*. *Biochemistry* **26,** 4085–4092.
Franco, R., Gavalda, M. T., and Canela, E. I. (1986). A computer program for enzyme kinetics that combines model discrimination, parameter refinement and sequential experimental design. *Biochem. J.* **238,** 855–862.
Gasa, T., Spruell, J., Dichtel, W., Srensen, T., Philp, D., Stoddart, J., and Kuzmič, P. (2009). Complexation between methyl viologen (paraquat) bis(hexafluorophosphate) and dibenzo[24]crown-8 revisited. *Chem. Eur. J.* **15,** 106–116.
Gilbert, H. F. (1999). Basic Concepts in Biochemistry. McGraw-Hill, New York.
Guldberg, C. M., and Waage, P. (1879). Über die chemische Affinität. *J. Prakt. Chem.* **127,** 69–114.
Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. (2006). COPASI—A COmplex PAthway SImulator. *Bioinformatics* **22,** 3067–3074.
Jamakhandi, A. P., Kuzmič, P., Sanders, D. E., and Miller, G. P. (2007). Global analysis of protein–protein interactions reveals multiple cytochrome P450 2E1reductase complexes. *Biochemistry* **46,** 10192–10201.
Johnson, M. L. (1992). Why, when, and how biochemists should use least squares. *Anal. Biochem.* **206,** 215–225.
Johnson, M. L. (1994). Use of least-squares techniques in biochemistry. *Methods Enzymol.* **240,** 1–22.
Johnson, M. L., and Frasier, S. G. (1985). Nonlinear least-squares analysis. *Methods Enzymol.* **117,** 301–342.
Johnson, K. A., Simpson, Z. B., and Blom, T. (2009). Global Kinetic Explorer: A new computer program for dynamic simulation and fitting of kinetic data. *Anal. Biochem.* **387,** 20–29.
King, E. L., and Altman, C. (1956). A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *J. Phys. Chem.* **60,** 1375–1378.
Kirkpatrick, S., Gelatt, C., and Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* **220,** 671–680.
Kuzmič, P. (1996). Program DYNAFIT for the analysis of enzyme kinetic data: Application to HIV proteinase. *Anal. Biochem.* **237,** 260–273.
Kuzmič, P. (2006). A generalized numerical approach to rapid-equilibrium enzyme kinetics: Application to 17β-HSD. *Mol. Cell. Endocrinol.* **248,** 172–181.
Kuzmič, P. (2009a). A generalized numerical approach to steady-state enzyme kinetics: Applications to protein kinase inhibition. *Biochim. Biophys. Acta—Prot. Proteom.* in press, doi:10.1016/j.bbapap.2009.07.028.
Kuzmič, P. (2009b). Application of the Van Slyke–Cullen irreversible mechanism in the analysis of enzymatic progress curves. *Anal. Biochem.* **394,** 287–289.
Kuzmič, P., Peranteau, A. G., García-Echeverría, C., and Rich, D. H. (1996). Mechanical effects on the kinetics of the HIV proteinase deactivations. *Biochem. Biophys. Res. Commun.* **221,** 313–317.
Kuzmič, P., Cregar, L., Millis, S. Z., and Goldman, M. (2006). Mixed-type noncompetitive inhibition of anthrax lethal factor protease by aminoglycosides. *FEBS J.* **273,** 3054–3062.
Kuzmič, P., Lorenz, T., and Reinstein, J. (2009). Analysis of residuals from enzyme kinetic and protein folding experiments in the presence of correlated experimental noise. *Anal. Biochem.* **395,** 1–7.

Le Clainche, L., and Vita, C. (2006). Selective binding of uranyl cation by a novel calmodulin peptide. *Environ. Chem. Lett.* **4,** 45–49.

Leskovar, A., Wegele, H., Werbeck, N., Buchner, J., and Reinstein, J. (2008). The ATPase cycle of the mitochondrial Hsp90 analog trap1. *J. Biol. Chem.* **283,** 11677–11688.

Mannervik, B. (1981). Design and analysis of kinetic experiments for discrimination between rival models. *In* "Kinetic Data Analysis" (L. Endrényi, ed.), pp. 235–270. Plenum Press, New York.

Mannervik, B. (1982). Regression analysis, experimental error, and statistical criteria in the design and analysis of experiments for discrimination between rival kinetic models. *Methods Enzymol.* **87,** 370–390.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **11,** 431–441.

Mendes, P., and Kell, D. (1998). Non-linear optimization of biochemical pathways: Applications to metabolic engineering and parameter estimation. *Bioinformatics* **14,** 869–883.

Morrison, J. F., and Walsh, C. T. (1988). The behavior and significance of slow-binding enzyme inhibitors. *Adv. Enzymol. Relat. Areas Mol. Biol.* **61,** 201–301.

Myung, J. I., and Pitt, M. A. (2004). Model comparison methods. *Methods Enzymol.* **383,** 351–366.

Myung, J. I., Tang, Y., and Pitt, M. A. (2009). Evaluation and comparison of computational models. *Methods Enzymol.* **454,** 287–304.

Nichols, T. E., and Holmes, A. P. (2001). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Map.* **15,** 1–25.

Niedzwiecka, A., Stepiński, J., Antosiewicz, J., Darzynkiewicz, E., and Stolarski, R. (2007). Biophysical approach to studies of Cap-eIF4E interaction by synthetic Cap analogs. *Methods Enzymol.* **430,** 209–245.

Onwubolu, G. C., and Davendra, D. (2009). Differential Evolution: A Handbook for Global Permutation-Based Combinatorial Optimization. Springer-Verlag, New York.

Penheiter, A. R., Bajzer, Ž., Filoteo, A. G., Thorogate, R., Török, K., and Caride, A. J. (2003). A model for the activation of plasma membrane calcium pump isoform 4b by Calmodulin. *Biochemistry* **42,** 12115–12124.

Peranteau, A. G., Kuzmič, P., Angell, Y., García-Echeverría, C., and Rich, D. H. (1995). Increase in fluorescence upon the hydrolysis of tyrosine peptides: Application to protease assays. *Anal. Biochem.* **227,** 242–245.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). Numerical Recipes in C. Cambridge University Press, Cambridge.

Price, K. V., Storn, R. M., and Lampinen, J. A. (2005). Differential Evolution: A Practical Approach to Global Optimization. Springer-Verlag, New York.

Reich, J. G. (1992). Curve Fitting and Modelling for Scientists and Engineers. McGraw-Hill, New York.

Schlippe, Y. V. G., Riera, T. V., Seyedsayamdost, M. R., and Hedstrom, L. (2004). Substitution of the conserved Arg-Tyr dyad selectively disrupts the hydrolysis phase of the IMP dehydrogenase reaction. *Biochemistry* **43,** 4511–4521.

Segel, I. H. (1975). Enzyme Kinetics. Wiley, New York.

Slyke, D. D. V., and Cullen, G. E. (1914). The mode of action of urease and of enzymes in general. *J. Biol. Chem.* **19,** 141–180.

Storme, T., Deroussent, A., Mercier, L., Prost, E., Re, M., Munier, F., Martens, T., Bourget, P., Vassal, G., Royer, J., and Paci, A. (2009). New ifosfamide analogs designed for lower associated neurotoxicity and nephrotoxicity with modified alkylating kinetics leading to enhanced *in vitro* anticancer activity. *J. Pharmacol. Exp. Ther.* **328,** 598–609.

Straume, M., and Johnson, M. L. (1992). Monte Carlo method for determining complete confidence probability distributions of estimated model parameters. *Methods Enzymol.* **210,** 117–129.

Szedlacsek, S., and Duggleby, R. G. (1995). Kinetics of slow and tight-binding inhibitors. *Methods Enzymol.* **249,** 144–180.

Van Boekel, M. (2000). Kinetic modelling in food science: A case study on chlorophyll degradation in olives. *J. Sci. Food Agric.* **80,** 3–9.

Von Weymarn, N., Kiviharju, K., and Leisola, M. (2002). High-level production of D-mannitol with membrane cell-recycle bioreactor. *J. Ind. Microbiol. Biotechnol.* **29,** 44–49.

Watts, D. G. (1994). Parameter estimates from nonlinear models. *Methods Enzymol.* **240,** 23–36.

Wielgus-Kutrowska, B., and Bzowska, A. (2006). Probing the mechanism of purine nucleoside phosphorylase by steady-state kinetic studies and ligand binding characterization determined by fluorimetric titrations. *Biochim. Biophys. Acta* **1764,** 887–902.

Wielgus-Kutrowska, B., Bzowska, A., Tebbe, J., Koellner, G., and Shugar, D. (2002). Purine nucleoside phosphorylase from *cellulomonas* sp.: Physicochemical properties and binding of substrates determined by ligand-dependent enhancement of enzyme intrinsic fluorescence, and by protective effects of ligands on thermal inactivation of the enzyme. *Biochem. Biophys. Acta* **1597,** 320–334.

Wielgus-Kutrowska, B., Antosiewicz, J., Dlugosz, M., Holý, A., and Bzowska, A. (2007). Towards the mechanism of trimeric purine nucleoside phosphorylases: Stopped-flow studies of binding of multisubstrate analogue inhibitor—2-amino-9-[2-(phosphonomethoxy)ethyl]-6-sulfanylpurine. *Biophys. Chem.* **125,** 260–268.

Williams, J. W., and Morrison, J. F. (1979). The kinetics of reversible tight-binding inhibition. *Methods Enzymol.* **63,** 437–467.

Williams, C. R., Snyder, A. K., Kuzmič, P., O'Donnell, M., and Bloom, L. B. (2004). Mechanism of loading the *Escherichia coli* DNA polymerase III sliding clamp. I. Two distinct activities for individual ATP sites in the γ complex. *J. Biol. Chem.* **279,** 4376–4385.