MAXIMUM ENTROPY ANALYSIS OF KINETIC DATA

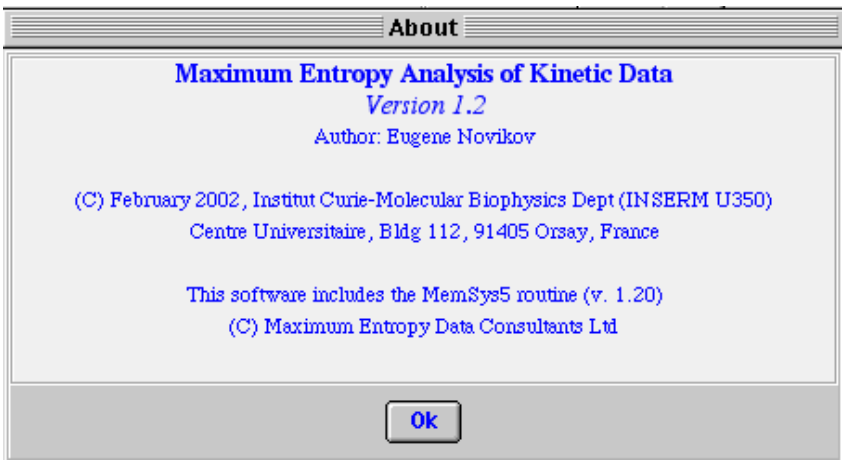
This software is part of the MEM Analysis System developed in my laboratory by Dr. E. Novikov.

Intended to increase our potential to analyse complex kinetics, this highly specialized software system was especially designed for laser photolysis studies of very heterochromatic protein–ligand rebinding reactions over wide ranges of time, temperature and viscosity. The appended review article presents typical results that could only be obtained by using the Maximum Entropy Method.

The present software is a simplified, but general purpose, module. It was written in Java in order to run on different platforms.

It is offered freely to the research community in order to promote and popularise this powerful and rigorous approach.

Daniel Lavalette, November 2005



Contents

BACKGROUND.

Many experimental kinetic data are expressed by a sum of exponentials :

$$N(t) = \sum_{1}^{N} \alpha_i e^{-k_i \ t}$$

where all $\alpha_i > 0$ and are independent of time. Examples are : chemical kinetics, flash photolysis, fluorescence decay, survival in a heterogeneous cell population, etc…

At t=0, the initial value is $: N(0) = \sum_{1}^{N} \alpha_i$. The sum being independent of time, the α_i are representing the permanent fraction of "species" *i*, or , but for an unessential normalization factor, the probability for finding the rate parameter k_i in the kinetics, N(t). The kinetician must solve the difficult problem of determining the α_i from the mere knowledge of N(t).

The distribution { α_i…} of the amplitudes (or probability, if properly normalized) is the **"Rate spectrum"** corresponding to the Kinetics N(t).

Information theory and Entropy (Shannon, 1949).

The entropy **S** of a probability distribution {p_1,p_2,….p_n) was defined by Claude Shannon as

$$S = -\sum p_i \log p_i$$

The minus sign warrants a positive S, since all p<0 yield negative logs. Note that S is "simply" the ensemble average of log p_i :

$$S = - \langle \log p_i \rangle$$

Entropy presents unique properties.

S is maximum for a uniform distribution, i.e.when all p_i are equal (there is no reason to expect one value more than any other).

*Consequently any constraint forcing the probabilities to depart from uniformity will lower the entropy of a distribution. The more "constrained" the distribution, the smaller **S** will be.*

One is generally interested in finding the particular distribution that is just necessary to account for some precise constraints, and not more. For example, Shannon proved that the maximum entropy distribution submitted to the only constraint that a variance exists is the Gaussian and that the maximum entropy distribution constrained to positive *x* and having an average value is the Exponential.

The fact that two of the most important distributions found in Science can be derived from the Maximum Entropy Principle highlights its deep significance and its power.

Use of maximum entropy principle to solve kinetic problems.

The kinetic problem can be reformulated in terms of probability distributions.

Before performing the experiment, we have no information about which rate parameters will be present in the kinetics. Therefore all of them must be assumed to be equally probable (the starting distribution will be the uniform one).
After recording N(t) the problem is to find the distribution of the α_i that will fit the data best at all time i.e. with $\chi^2=1$ **as first constraint** while decreasing its entropy as little as possible (otherwise we would admit subrepticely additional constraints). This is the essence of the Maximum Entropy Method (MEM)..

What are "data" ?

Any data set should actually include our knowledge about the confidence to be assigned to each data point. Think of the width of error bars in a graph : if the "best" fit is defined as the one that is constrained to remain within the limits of each error bar, it is clear that different fits will result from varying their width. This indicates that **noise is a second constraint**.

Thus the answer can be quite different if unrealistic noise values are introduced because the question will be each time a different one ! The real scientific approach is to seek the most likely answer to a question **given all our present knowledge and information**. In an experiment, our knowledge includes measured data together with their associated uncertainty. There is no way out.

In other words the **noise** must be quantified (knowledge of σ), and not only because we need calculate the χ^2.

Consider two oppositely tempered scientists :
1-*The optimist*. He (she) thinks that his (her) data are excellent and therefore he (she) uses too low a value for the noise input. Consequences : maximum entropy will assume that the goal is to run closely though all details. The rate spectrum will generally consist of a large number of lines, but most of them will not fit data, but the noise. Most of the structure of the spectrum is illusory.
2-*The pessimist.* He (she) thinks that his (her) data are very poor and therefore he (she) uses too high a value for the noise input. Consequences : maximum entropy will assume that the goal is to run rather loosely around the data points. The spectrum will contain little structure. Important information may be obscured or even lost.

Some people think (but they are wrong!) that this is an additional complication that does not arise when they perform least-square fits. They are just forgetting that least square fits assume that the function to be fitted is the correct one. Lest square fitting can only estimate the parameter values of a known function. For the kinetic problem, this amounts to assuming that one already knows the answer to the question one is asking….Actually, it is impossible without strongly biasing the problem, to be sure whether 1,2, or… 27 exponentials are present in N(t).

The maximum entropy approach is entirely different. There is no a priori restriction to the number of exponentials in the solution. Even continuous distributions are allowed (see appended article).

The MEM solution yields the distribution of α_i that is both necessary ($\chi^2=1$)and sufficient (S is maximum) to account for the [data+noise] system.

Note : MEM is not limited to exponentials of course, but the present software has been designed to treat that particular and very common form of function.

What this software does.

As a general purpose MEM software it will accept any type of (positive) data that can be expressed in the form of equation 1. The data must be provided as tab-separated text files : time->signal->noise σ

The software also provides a simulation module. Using simple and common build-in distributions (spectra), the corresponding kinetics can be simulated. Noise can be added at will. Once saved, the simulation files can be used as data files and further analysed by the MEM module, giving to the user the opportunity of testing and experimenting with the adjustment of the fitting parameters.

What this software cannot do.

N(t) must be uniformly decreasing, i.e. **N(t) is not allowed to contain negative** α_i**.**
Special kinetics like energy transfer displaying rising parts cannot be adequately treated with the present version.

This software does **not provide a deconvolution for the** "trigger function". For instance, in fluorescence as well as in flash photolysis, the flash is assumed to be a delta function. If this not the case, only data recorded after the trigger function has died out can be correctly analysed.

Sorry.

Copying the software on your hard disk.

Create a " **MEM Folder"** on your Desktop and copy the following 6 items to this folder.
firstMEM_1.2.jar : a self-suppporting Java Archive that will launch and operate the program when installed on computers running under MacOS X or Windows and equiped with an updated Java Machine (usually present in the various OS's).
firstMEM_1.2 : a highly zipped form of the program needing a special installation procedure when operated with MacOS 9.x.
Init.txt : a text file for parameter initializing before running MEM
defaults.txt : the same file, to be used as such after copying into the Preferences Folder under MAcOS 9.x
DemoKin.txt : text file of real kinetic data for experimenting with MEM analysis.

MEM_UsersManual.pdf

IMPORTANT NOTICE

This software incorporates the MemSys5 routine (Maximum Entropy Data Consultants Ltd) that offers many input facilities for specialized applications only.

To avoid any troubles, always Load the Init.txt defaults settings file immediately after launching and before performing any calculation or simulation as described on page 10.

SOME TECHNICAL PRINCIPLES

The kinetics N(t) are the result of an arbitrary number of species "decaying" independently (the α_I are independent of time)

$$N(t) = \sum_i \alpha_i e^{-k_i t} \qquad \text{sum over species}$$

alternatively

$$N(t) = \sum_k P(k) e^{-kt} \qquad \text{sum over rates}$$

P(k) or $\{ \alpha_i \ldots \}$ are the amplitude (or probability, if properly normalized) of the distribution of rates k. Their generic denomination will be the **"Rate spectrum"** corresponding to the Kinetics N(t).
For continuous functions or distributions, such as observed with protein statistical substates, the sums must be replaced by integrals.

$$N(t) = \int P(k) e^{-kt} dk$$

Mathematically N(t) is the Laplace transform of P(k). For computational purposes, the integral is always considered as a sum over a large number of rates ("discretization").

Use of log scales.

Virtually all recording set-up are linear in time. When the kinetics is very inhomogeneous (i.e. covering several orders of magnitude in time), this will invariably lead to a situation where long times are statisticallt oversampled (they gather many more points). Any type of fitting will be strongly biased in favour of long times, even though they correspond to the end of the kinetics that is generally less accurate because of noise. In addition, the total number of points will be enormous (sometimes hundred of thousands).

To obviate these problems, we use logarithmic time and rate scales. A log (time) representation is the only one permitting to display inhomogeneous kinetics. For instance, the DemoKin.txt data file provided with this software covered 6 time decades as shown in figure 1.

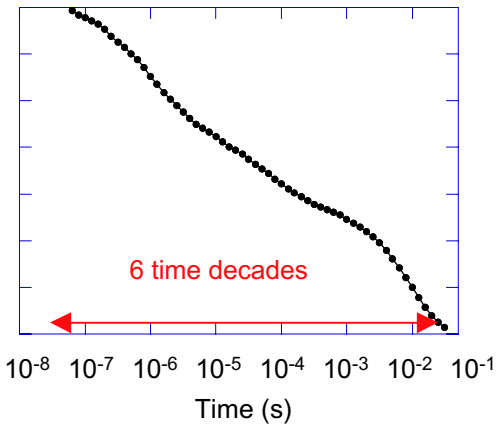


Fig. 1 absorbance changes recorded during geminate rebinding of Myoglobin with Carbon Monoxide in glycerol/water at 250 K

In such a case, the use of a **data pre-treatment** is recommended. In the figure, the data were **logarithmically averaged** (see APPENDIX) to get 10 equally spaced data points per time decade with ≈ constant signal/noise ratio. Experience indicates that this is usually sufficient. Each user must find the correct way of pre-treating his particular data.

The rate spectrum calculated by MEM from these data is shown in fig.2. below.

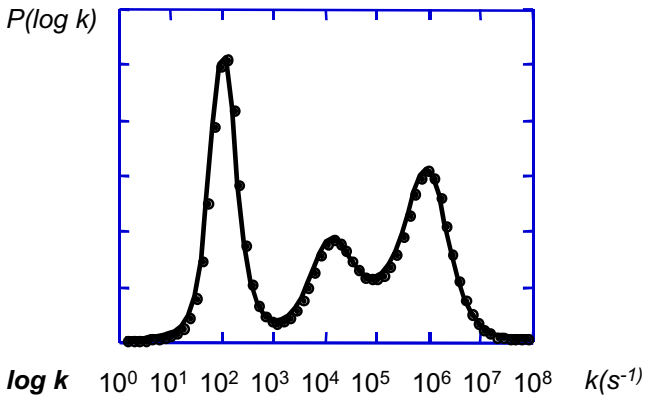


Figure 2. Rate spectrum corresponding to the kinetics of figure 1.

Again, because the rates extend over many orders of magnitude, a log distribution is recommended. The rate spectrum is displayed usually as P(log k)

$$N(t) = \sum_k P(\log k) e^{-kt}$$

Note : be careful when normalizing plotted results to integrate the area of the spectrum according to the independent variable used (k or log k).

In FLUORESCENCE SPECTROSCOPY decay rates are discrete and usually contained within one or at most two decades only. In that case, one may dispense with logarithmic pre-treatment and run MEM directly with the linear time scale of the recording device. However, MEM will always calculate the spectrum as P(log k). *Be careful however that the number of long time data points do not outweighs short time data points too much (the fit of the beginning of the kinetics could be difficult).*

How the software is working.

First, look at the time range of the kinetics in order to define the minimum and maximum rate parameters: k_{min} and k_{max} susceptible to contribute.

The program :
-internally generates the required number of log-equidistant rates between k_{min} and k_{max}
-pre-calculates all these exponentials exp(-kt) for the number of time-points specified in the input window;
-starts calculating the Laplace transform by assuming all the α_i equal to the "flat amplitude" indicated by the user and at the same times calculates the entropy;
-iteratively tries to improve the fit to the data by changing the α_i until $\chi^2 => 1$ while keeping the entropy of the α_I maximum.(Internal Memsys5 routine).
Convergence is usually achieved within 100-200 iterations at most when calculating with 100 datapoints.

INSTALLING MEM

Installing and Running MEM has become very easy with the MacOSX and present Windows OS or Linux Operating Systems.

If working with these OS, use the **firstMEM_1.2.jar** file. Double clicking will launch Java and MEM

Just make sure that your computer's OS has an up-to-date **Java machine**. It usually has. If not, download an update from the internet site for Mac, Windows, Sun, etc…

The installation described below is for Mac users with MacOS9.x.

1-Use the Program file named : **firstMEM_1.2**, . It is a highly zipped text file written in Java. **Useless** to try to open it.
2- Place immediately a copy of **defaults.txt** in the **Preference Folder** (do not rename the file, the program would not recognize it!).

Three **Java** items are required for installing and running MEM.

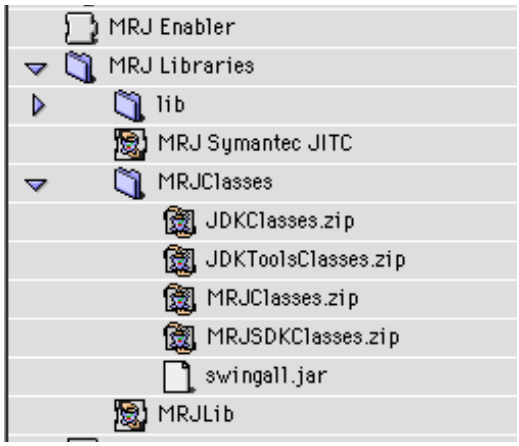
a- A Computer's internal **Java Machine** (usually already present in the OS of the computer)
For Mac users, it is **MRJ** contained in the system or updated from the kit **MRJ SDK** (Mac Runtime for Java, Standard Developer Kit) downloaded from <developer.apple.com/java/>.

•b- Jbindery also in the kit MRJ SDK to create an **Applet** that will launch the program (after automatically unzipping the program and establishing the required links between the program's modules and the Java machine).

•c A Java archive, called **swingall.jar,** obtained from the Application called **Swing** downloaded from <java.sun.com/>

Once these elements have been copied on your disk, proceed as follows.

1-The Computer's Java Machine has a Libraries Folder in the Extensions Folder, containing a JavaClasses Folder. Copy **swingall.jar** (or equivalent) to this Folder.



Close.

2- **Launch Jbindery.**
A dialog box appears

In the field **ClassName,** type : **MainApplication** (exactly as written here!)
In the field **Classpath** : click on **<Add.zip.file>** , then browse to select **firstMEM_1.2.**
Click on **<Save Settings> after selecting <as an application>.**

This creates the Applet (that can be renamed at will) to be used for running MEM.
Put this Applet in the MEM folder.

Installation is now finished.

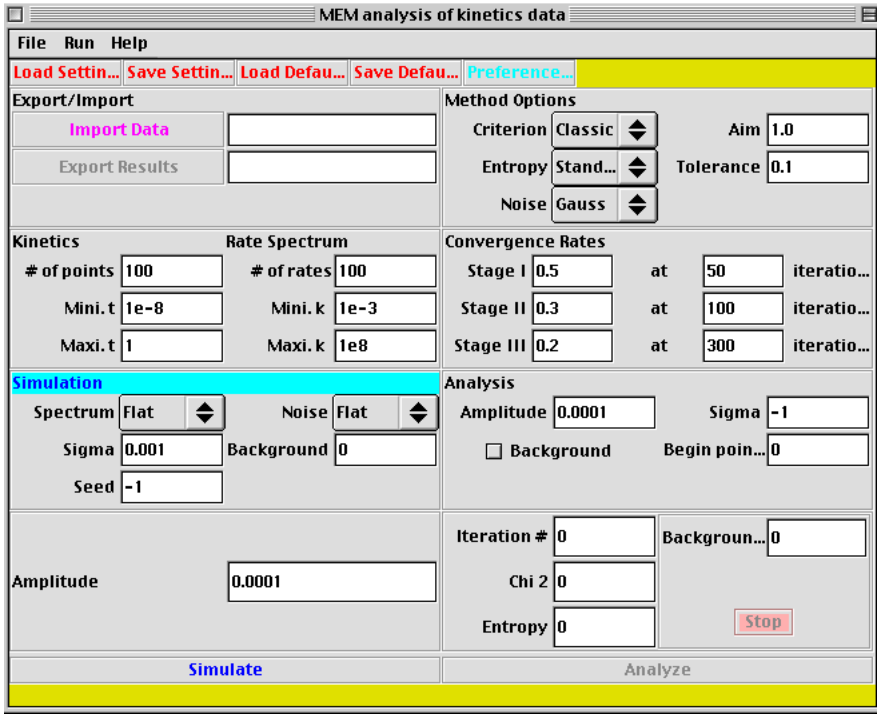
Unless you move the MEM Folder to another place, you will not have to re-use the java softwares. If you move the MEM folder, you will have only to run Jbinderry again.

To run MEM, just double click on the Applet created with Jbinderry.

_____

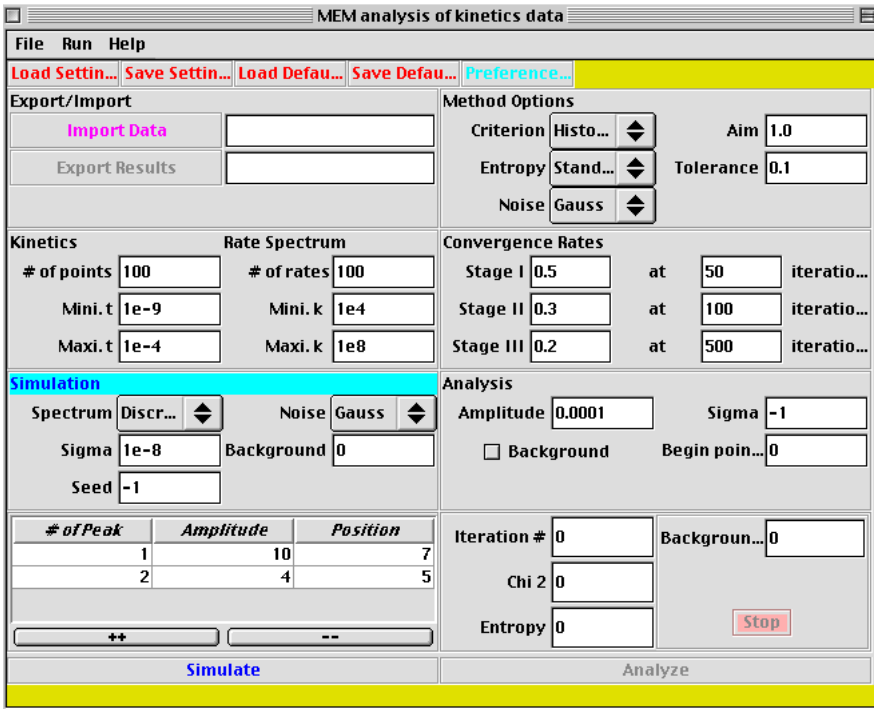
GETTING STARTED

Launch MEM.
This window appears.



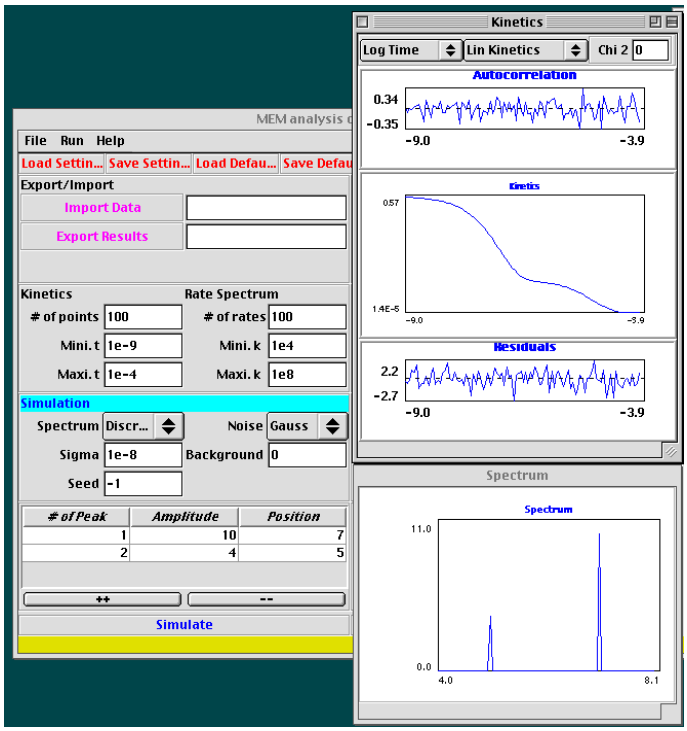
The fields are filled with internal default values. These defaults values are **NOT** appropriate for most uses.

• Select **<Load Settings>** and browse to select the file "**Init.txt**". This is a good general purpose default setting. The window changes to :



The simulator module.

The Simulator Module (bottom left) is very useful to get acquainted with the basic inputs of the program.
• Click on the <Simulate> button.
Two new windows appear .They are re-sizable and you can move them around. For the moment, they hide the right part of the main window that we do not need yet.



By clicking on <Simulate> with the present default values you have already generated:

- a rate spectrum consisting of two discrete lines : one at k=1e7 (s^{-1})with amplitude 10 and a second one at k=1e4 (s^{-1}) with amplitude 4.
- The corresponding kinetics (the Laplace transform of the spectrum) to which some minimum amount (σ = 1e-8) of Gaussian Noise has been added.

The "Kinetics" Window provides **linear/log display facilities** independently for the amplitude and the time axes.

Kinetics & Rate Spectrum Panel.

These 6 fields are common to the Simulator and for Analysing real data.
To perform any calculation MEM needs to know (cf. Technical Principles)

How many different rates will be used for calculation (100 is generally quite enough)
The minimum rate (s^{-1})
The maximum rate (s^{-1})
The program internally reserves the value of 100 (or more) **logarithmically equidistant rates within the indicated range.**
And symmetrically:
How many kinetics data point the user wishes for the kinetics (here also 100 is a typical value).
From which minimum time (s)
Up to which maximum time (s)
The program internally calculates 100 (or more) **logarithmically equidistant times within the indicated range. The kinetics will be simulated for these time-points.**

Note *: Although you may enter any values in the 6 fields, times and rates are generally chosen to be reciprocal values ±1 (it is useless to calculate the kinetics before* t_{min} *= 1/*k_{max} *and after* t_{max} *≈ 5 to 6 x 1/*k_{min}*).*

Simulation Panel

Now, you can play around simulating various spectra.

1-To increase the number of peaks, select <+++>. To decrease the number of peaks, select <--->

2-To change the parameters of the peaks :enter the value of the **amplitude** (arbitrary) and of the **decimal log of the rates**

Note : a field is not validated as long as you do not click in another field.

3-Adding noise.

Noise Menu: **Flat** (white) noise or **Gaussian** noise
Sigma : enter here the amplitude of the noise expressed in terms of **Standard deviation** (square root of Variance)
Seed=**-1**. Seed of the noise generator. If ≠ -1, the same noise will be reproduced again and again. It is better to keep it to –1 (new noise series at each trial).

4-Background
You may wish to add a constant value to the kinetics (baseline ≠0).Enter the baseline here.

Note *: while simulating a kinetics it may happen that a negative value is obtained at the end of the kinetics when the noise becomes of the same order of magnitude. A Warning message appears. You must either decrease the noise sigma or add some positive background.*

5-Type of spectra provided.

Flat : a uniform distribution of k. All k values have the same amplitude.
Discrete : for simulating the sum of a few exponentials (the present example).
Gaussian : a Gaussian distribution of rates : $P(\log k) = \frac{1}{\sigma_n \sqrt{2\pi}} e^{-\frac{(\log k - \log k_0)^2}{2\ \sigma_n^2}}$ Enter peak value (decimal log of mean rate) and std. Deviation. This distribution is symmetrical.
Gamma :a Gamma distribution of rates according to : $P(\log k) = \frac{1}{\Gamma(n+1)} \cdot \frac{1}{\Delta} \left(\frac{\log k - \log k_0}{\Delta} \right)^n e^{-\frac{\log k - \log k_0}{\Delta}}$
This distribution may be quite asymmetrical depending on the parameters.

6- Save / Load settings.
You may save the whole configuration of the MEM window for later use. Browse for choosing the directory (Folder) where you wish the(text) file to be saved after naming it.

7- Save / Load defaults.
If there is a "Preference" Folder, as in a Mac, launching MEM will automatically load your default settings (the file must be named "defaults.txt").
Otherwise, use <Load settings> to load your preferred configuration, ad done above with the Init.txt file. You may have several different settings according to you own choices.

8-Export results.

The results of the simulation are saved as a tab-separated text file.

Note *: depending on the OS, the results file may not be recognizable by the Applications present. With MacOSX or Windows, just add the .txt extension. For Mac users with MacOS9.x the freeware "Drop•Attribute", provides a drag-and-drop applet that immediately give the correct signature. The file can then be opened with any Application able to import Text files (even graphic softwares such as Kaleidagraph).*

When opened in Excel, the result file looks as below.
Equidistant log k values-Corresponding amplitude (as entered)-equidistant log time values-Here, MEMfit is simply the calculated kinetics-Residuals : here simply reproduce the noise that was added.

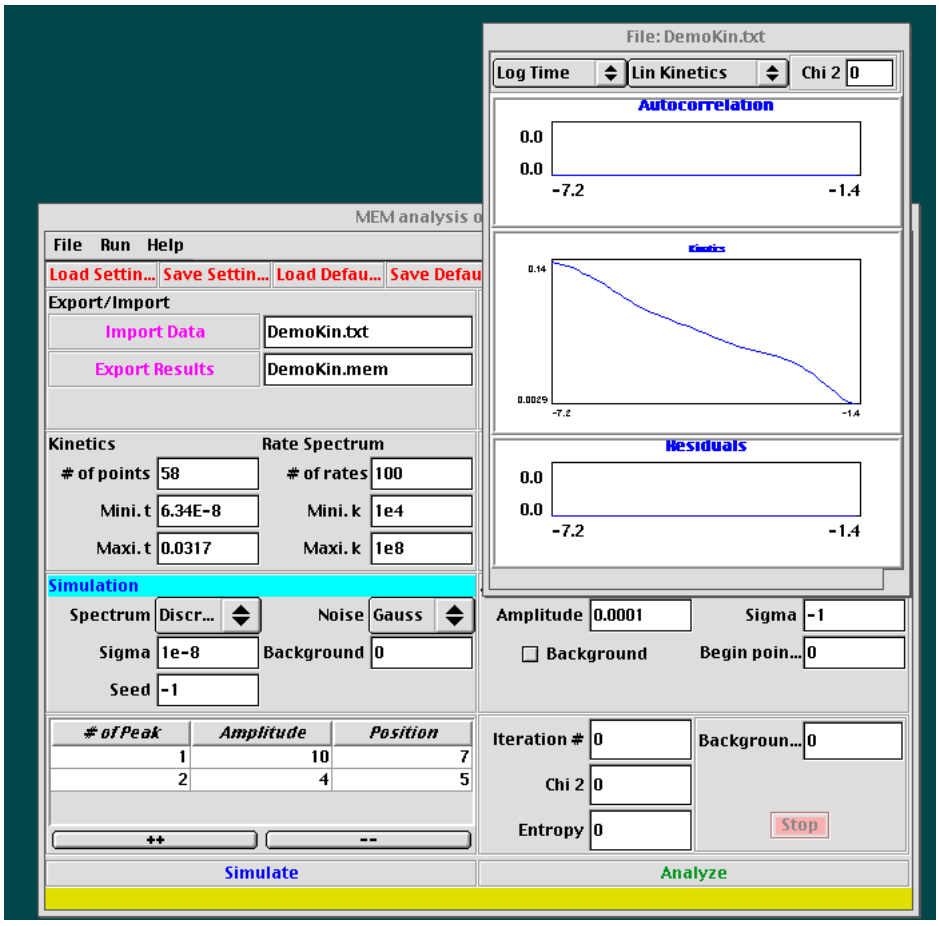
log k	P(log k)	log t	D exp	MEM fit	Residuals
4	0	-9	0.56171319	0.5617132	-2.3682105
4.04040404	0	-8.9494949	0.56122955	0.56122954	1.98619362
4.08080808	0	-8.8989899	0.56068693	0.56068693	-0.0250037
4.12121212	0	-8.8484848	0.56007827	0.56007827	0.59312713
4.16161616	0	-8.7979798	0.55939565	0.55939565	0.1142843
4.2020202	0	-8.7474747	0.55863022	0.55863022	-0.9117783
4.24242424	0	-8.6969697	0.55777214	0.55777215	-1.39924
4.28282828	0	-8.6464646	0.55681045	0.55681045	-0.4438257
4.32323232	0	-8.5959596	0.55573291	0.55573291	0.47938877
4.36363636	0	-8.5454545	0.55452595	0.55452596	-0.4951795
4.4040404	0	-8.4949495	0.55317453	0.55317453	-0.9904832
4.44444444	0	-8.4444444	0.55166195	0.55166195	0.91253084
4.48484849	0	-8.3939394	0.54996973	0.54996973	-0.1926186
.....					
4.84848485	0	-7.9393939	0.5225529	0.52255291	-0.6666156
4.88888889	0	-7.8888889	0.5175635	0.51756349	1.0801757
4.92929293	0	-7.8383838	0.51204023	0.51204023	0.13620256
4.96969697	4	-7.7878788	0.50593703	0.50593703	-0.53645
5.01010101	0	-7.7373737	0.49920665	0.49920665	-0.4603271
5.05050505	0	-7.6868687	0.49180151	0.49180151	0.7116114
5.09090909	0	-7.6363636	0.48367477	0.48367476	1.46527766
5.13131313	0	-7.5858586	0.47478165	0.47478165	-0.2086885
.....					
6.70707071	0	-5.6161616	0.12896257	0.12896257	1.00333938
6.74747475	0	-5.5656566	0.12542246	0.12542246	0.00183466
6.78787879	0	-5.5151515	0.12156153	0.12156153	1.21601185
6.82828283	0	-5.4646465	0.11736602	0.11736602	-0.5552926
6.86868687	0	-5.4141414	0.11282555	0.11282555	0.31490421
6.90909091	0	-5.3636364	0.10793427	0.10793427	-0.9957231
6.94949495	0	-5.3131313	0.10269221	0.10269221	0.44798246
6.98989899	10	-5.2626263	0.0971067	0.0971067	-0.2639959
7.03030303	0	-5.2121212	0.09119384	0.09119384	0.53230746
7.07070707	0	-5.1616162	0.08498008	0.08498008	1.33215456
7.11111111	0	-5.1111111	0.0785035	0.0785035	-0.1111116
.....					
7.7979798	0	-4.2525253	8.79E-04	8.79E-04	-0.1701803
7.83838384	0	-4.2020202	4.62E-04	4.62E-04	0.65524842
7.87878788	0	-4.1515152	2.24E-04	2.24E-04	-0.5318456
7.91919192	0	-4.1010101	9.97E-05	9.97E-05	0.99238497
7.95959596	0	-4.0505051	4.01E-05	4.01E-05	-0.2141467
8	0	-4	1.44E-05	1.44E-05	1.83278275
0	0				
0	0				

Note 1 *: the last two lines always contain 0 values that may cause problems in graphic softwares or if you want to use these results for simulating later an analysis by MEM. It is a good idea to remove them immediately.*

ANALYSING DATA

After finishing simulation exercises, we proceed to analyse real kinetic data with MEM.
From now on, the simulator serves no purpose anymore.

Select <Import Data>
Browse the MEM folder and select the file "DemoKin.txt"
(This file contains kinetic data about geminate rebinding of CO with Myoglobin measured by absorbance changes as signal, but it could have been anything else).
Click on the Kinetic window to bring it in front. You should see a graphical display of the imported kinetics.



Compared to the preceding screen, note the following changes.

1- Name of the imported Data file appears in the Import Data field.
2- Default name for the result file to be exported after performing the MEM calculation (you may modify this name at will).
3-The 3 fields "Kinetics" have been updated to correspond to the actual values of the imported file.
- the file contains 58 time-data points
- minimum time in the file
-maximum time in the file.

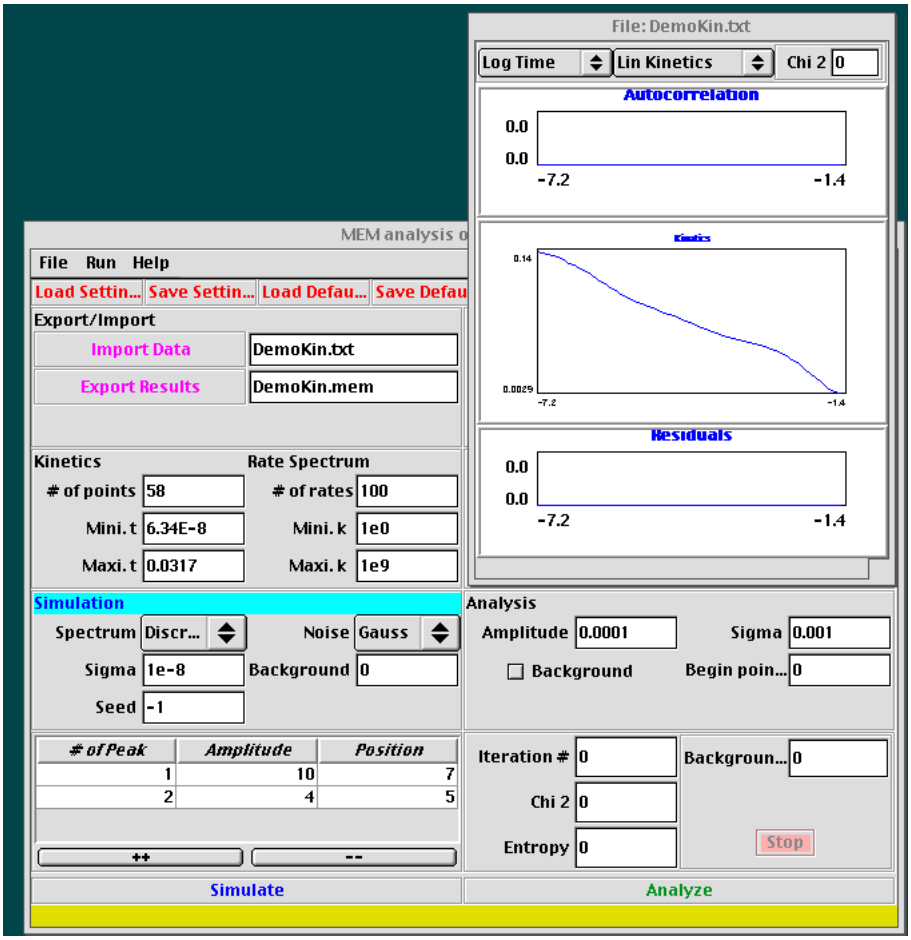
All other fields are unchanged.

Setting the analysis parameters for this demo.

a) First, the range of the desired rate spectrum must be entered.

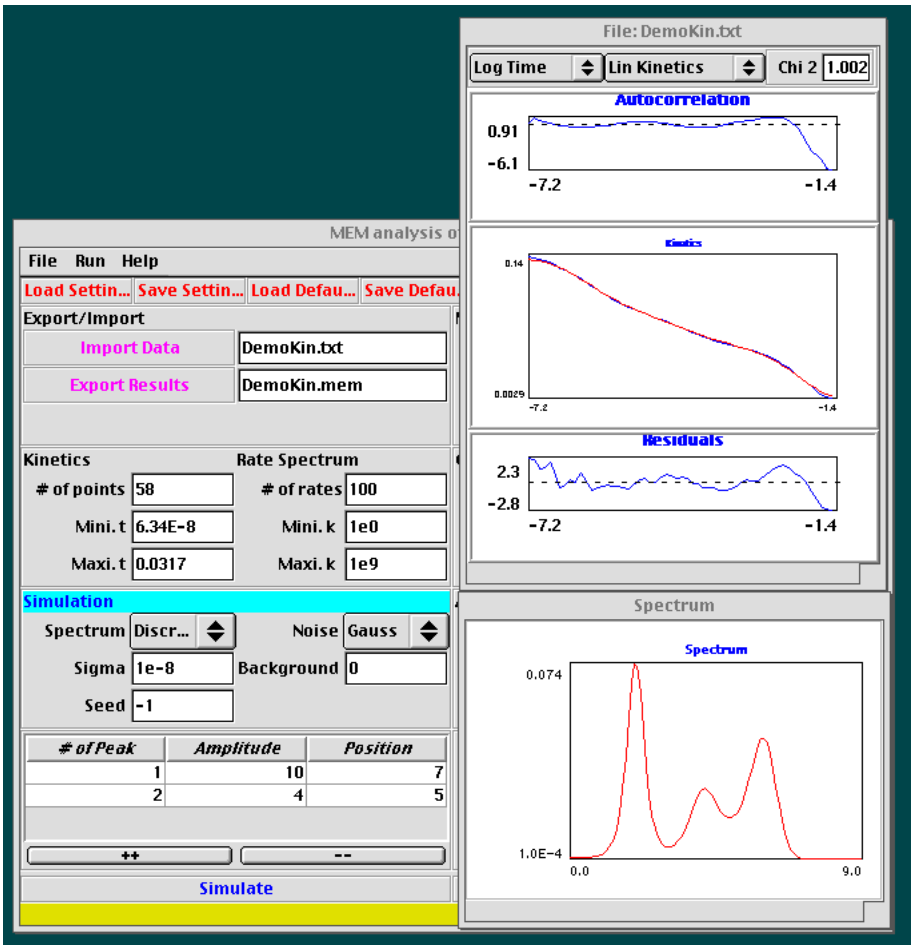
Leave 100 in the field : the spectrum will consist of 100 k values.
Enter 1e0 as mini k
Enter 1e9 as maxi k

b) In the "Analysis " panel, right, :
Enter : 0.001 (or 1e-3) in the Sigma box.



Everything is ready. Just click the button **"<Analyze>** and see what happens.

Within seconds, the spectrum builds up while the fit progressively comes closer to the data.



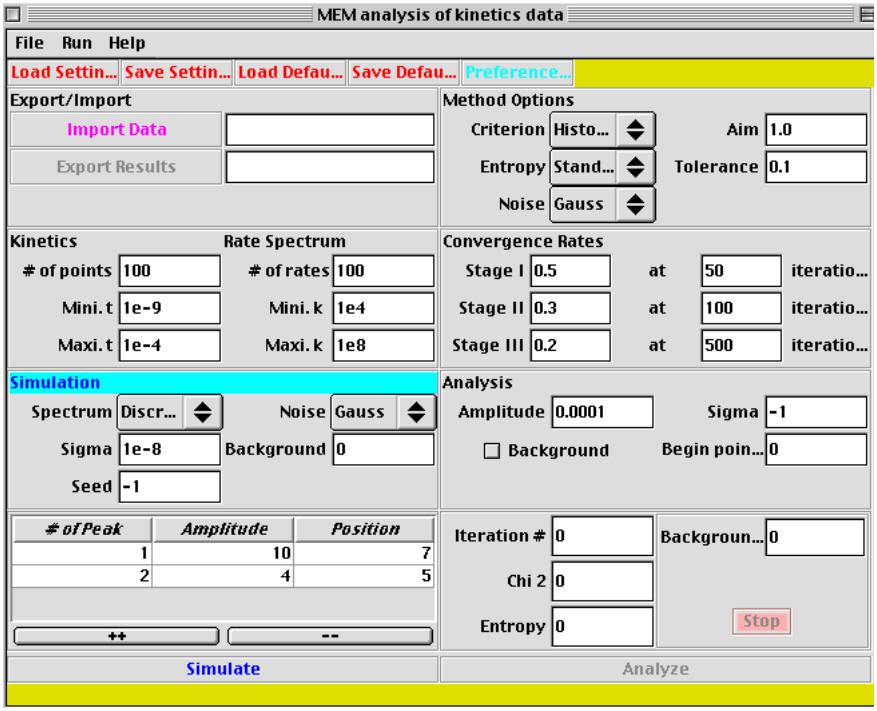
The program stops automatically after 97 iterations because Chi 2 = 1 has been attained.

Performing a MEM analysis is rarely (?) more complicated than that.

You may now export the results.

Note : It may be clever to change the extension to .txt before saving. You will then be able to open the text file easily.

Analysis Module Settings



Method Option Panel : except for the Noise option, you will never have to change the settings of this panel.

Criterion
Is a choice of options within the Memsys5 routine. Use only **HISTORIC**.

Entropy
Is a choice of options within the Memsys5 routine. Use only **STANDARD**

Noise
Indicate the type of noise affecting your data : **GAUSS** or **POISSON**
NB: Poisson noise is the one to use preferentially with photon-counting data.

Aim: is the target Chi2. Should remain at **1**

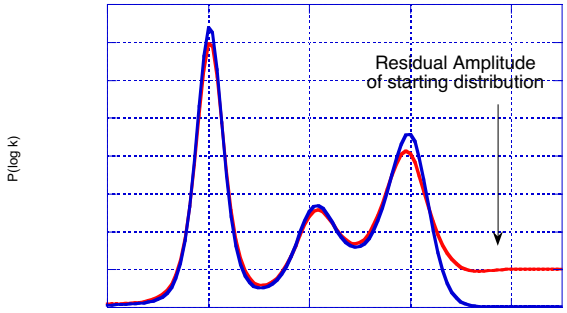
Convergence Rates Panel

The calculation stops as soon as convergence is achieved, otherwise continues to the largest iteration number entered in the lower right box.
Three successive decreasing convergence rates can be applied to the iteration numbers indicated on the right in order to get a smooth convergence.
The default values are adapted to most situations. Do not try experimenting with these inputs unless you are already familiar with MEM.

Analysis Panel

This is the only panel that requires your inputs at each particular calculation.

Amplitude:
The calculation must be initialised by indicating some (arbitrary) amplitude for the uniform starting rate distribution.
Enter a small value, such as 1e-4
The first idea is to enter as amplitude the reciprocal of the number of rates specified in the Kinetic and Rate Spectrum Panel, e.g. with 100 *a priori* equiprobable rates to enter 1/100. This is correct, but not optimal. Remember that MEM will **not** change the amplitude of the rates for which there is no information in the data. Therefore, these "flat amplitudes" will appear in the spectrum. This not only inaesthetic, but also makes the calculation of the band area (when necessary) difficult.



By entering a very small " flat amplitude to start" you will not spoil your spectrum since MEM will let grow the amplitudes only where this is required to fit the data.

Background
Select this option in case you have reason to suspect that there is a non zero baseline in your data. MEM will also calculate this background amplitude for you.

Begin Point.
Analysis will consider only data after the begin point entered. This is very convenient for suppressing a number of initial points without having to create independent data files in case the beginning of your kinetics is spoiled (initial time constant, contribution of the flash function ,…)

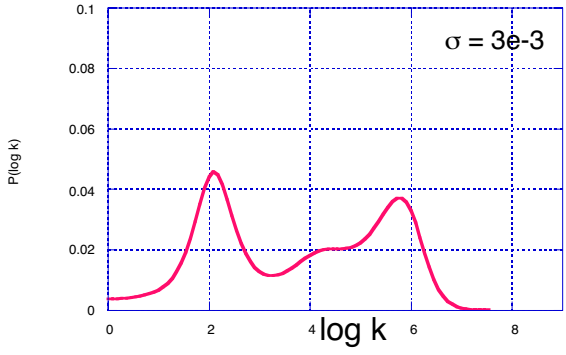
Sigma
Certainly the most important parameter to enter.
If you enter –1 : the program will use exactly the sigma associated with each point in the data file.
I you enter any other value, the program will apply that noise value to all data points (that is, it does not consider the individual sigmas of the data file).

This is often the best solution, for several reasons:
First, noise is part of the data, although certainly the most difficult part to estimate correctly. Any over or under estimation will seriously affect the spectrum.
Second :
- noise statistics coming from your set up are usually too optimistic.
- by entering successively different sigmas you will be able to estimate the stability of your results with respect to your noise estimation (remember that the analysis takes only a few seconds, so you can repeat successively with different sigmas). An example is provided in the following section.

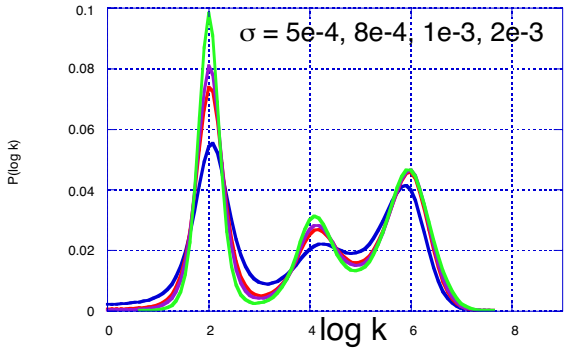
Lower Panel : self explanatory indications given by the program.

Experimenting with noise.

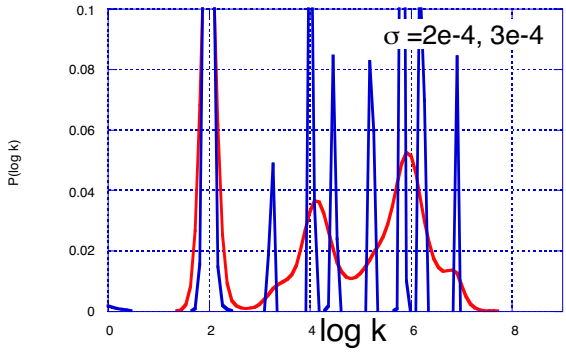
The figures show the P(log k) plots obtained after analysing the DemoKin file using various sigma inputs around 1e-3.



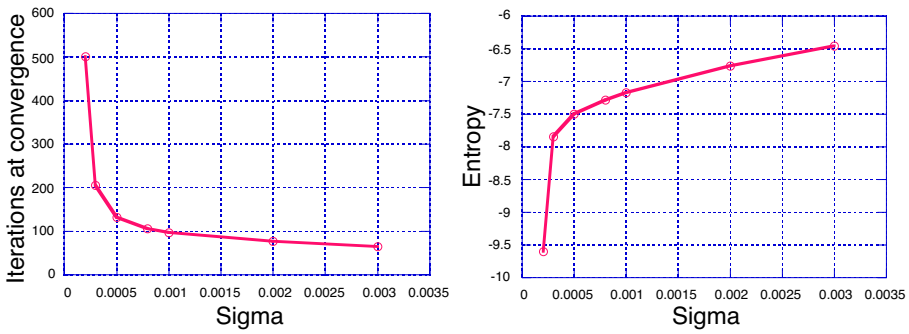
Three bands are visible, but not very well resolved. (The fit is a little bit deficient at the beginning of the kinetics).



The results are stable even with a fourfold change of sigma. The fits are all OK.



Noise is underestimated. Additional structure appears first, then a "comb" spectrum is obtained. The fits are still (obviously) excellent…because noise is fitted altogether.



The sudden changes in the iteration number for convergence and in the entropy for sigma < 0.0005 indicate some unreasonably optimistic noise estimation !

The simplicity with which data can be re-analysed by entering different sigma estimations is a great advantage. In our example, it is clear that **the features of the spectrum (3 bands) are warranted because the solution remains stable upon a 4-fold change of the noise**.

Users should always perform this check with important data.

When starting to use MEM for analysing a particular type of data obtained with a particular set-up, it is advisable to perform such "noise" experiments before jumping to conclusions. You will learn a lot, not only about MEM, but, more important, about your data (they are rarely as good as you believe)…modesty is in order, and this will encourage you improving your set-up.

The Chi2 criterion.

There is no absolute and universal criterion to tell when the calculation should stop. Here, as in many other situation, the Chi2 is used for its convenience and simplicity.

$$\chi^2 = \left\langle \frac{(N_i^{\text{calc}} - N_i^{\text{exp}})^2}{\sigma_i^2} \right\rangle$$

One should always keep in mind that $\chi^2 = 1$ **does not mean that a fit is perfect, but that the fit is, on the average, as good as the precision of data points permits.**

Because of this statistical nature, $\chi^2 = 1$ may be compatible with caricatural situations in which simple visual inspection reveals that the fit is very poor in some parts of the kinetics ! This happens when unrealistic pessimistic sigmas are entered. Therefore, it is always recommended to look also at the residuals…

Experimenting with MEM without any experimental data.

A good start for getting the "MEM touch" without needing experimental data is to analyse some kinetics created using the simulator with different amount of noise.

The great advantage is that you know already the answer you are looking for … which is of course usually not the case with actual data.

To avoid confusion, the result file format of the simulator is different from the data file format (see below), and you **cannot** import it immediately as data.

If you wish to analyze a simulated noisy kinetics, use a text processor or Excel to convert the simulation results to the format of a genuine data file.

a) Remove the last two lines (zero values would freeze the program).
b) Calculate the real times from their log;
c) Copy the Dexp column;
d) …and add the sigma column using the sigma value of the simulation.

DATA FILE FORMAT

	TIME	SIGNAL	SIGMA
2	6.34E-08	1.38E-01	5.74E-04
3	8.08E-08	1.37E-01	5.32E-04
4	1.01E-07	1.35E-01	4.59E-04
5	1.27E-07	1.34E-01	4.09E-04
6	1.60E-07	1.33E-01	3.72E-04
7	2.01E-07	1.30E-01	3.30E-04
8	2.52E-07	1.27E-01	2.92E-04
9	3.19E-07	1.25E-01	2.30E-04
10	3.99E-07	1.23E-01	1.46E-04
11	5.04E-07	1.20E-01	9.43E-05
12	6.33E-07	1.18E-01	8.32E-05
13	7.96E-07	1.14E-01	7.42E-05
14	1.00E-06	1.10E-01	6.61E-05
15	1.26E-06	1.07E-01	5.88E-05
16	1.59E-06	1.04E-01	5.43E-05
17	2.00E-06	1.01E-01	6.52E-05
18	2.52E-06	9.79E-02	6.74E-05
19	3.17E-06	9.52E-02	6.01E-05
20	3.99E-06	9.24E-02	5.32E-05
21	5.02E-06	9.00E-02	4.74E-05
22	6.33E-06	8.82E-02	4.22E-05
23	7.96E-06	8.65E-02	3.76E-05
24	1.01E-05	8.45E-02	7.13E-05
25	1.26E-05	8.23E-02	8.74E-05
26	1.60E-05	8.02E-02	8.20E-05
27	2.00E-05	7.88E-02	7.30E-05
28	2.55E-05	7.70E-02	7.13E-05
29	3.19E-05	7.50E-02	9.73E-05
30	4.05E-05	7.26E-02	1.05E-04
31	5.07E-05	7.08E-02	9.39E-05
32	6.36E-05	6.87E-02	8.43E-05
33	7.99E-05	6.63E-02	7.45E-05
34	1.00E-04	6.44E-02	6.63E-05
35	1.27E-04	6.21E-02	5.99E-05
36	1.59E-04	6.04E-02	6.66E-05
37	2.03E-04	5.87E-02	6.66E-05
38	2.52E-04	5.71E-02	5.90E-05
39	3.18E-04	5.56E-02	5.35E-05
40	4.01E-04	5.43E-02	4.71E-05
41	5.02E-04	5.32E-02	4.20E-05
42	6.33E-04	5.22E-02	3.76E-05
43	7.96E-04	5.10E-02	3.33E-05
44	1.01E-03	4.91E-02	4.61E-05
45	1.26E-03	4.74E-02	6.10E-05
46	1.60E-03	4.57E-02	6.12E-05
47	2.00E-03	4.38E-02	5.45E-05
48	2.53E-03	4.16E-02	4.85E-05
49	3.17E-03	3.90E-02	4.31E-05
50	4.02E-03	3.59E-02	5.04E-05
51	5.04E-03	3.23E-02	5.70E-05
52	6.36E-03	2.82E-02	5.47E-05
53	7.99E-03	2.43E-02	4.83E-05
54	1.00E-02	2.02E-02	4.30E-05
55	1.26E-02	1.57E-02	3.82E-05
56	1.59E-02	1.15E-02	3.38E-05
57	2.00E-02	7.87E-03	3.02E-05
58	2.52E-02	5.04E-03	2.69E-05
59	3.17E-02	2.98E-03	2.39E-05
60	# Fin Data		
61	# Nbre de points : 58		
62	# nbre de zon	8	
63	# 1	9	
64	# 2	16	
65	# 3	24	
66	# 4	28	
67	# 5	35	
68	# 6	44	
69	# 7	50	
70	# 8	58	
71			

The data file DemoKin.txt (separators are tabs) look like the one on the left.

Time : in seconds.
Here, our own data pre-treatment yielded logarithmically spaced times, but this was just for convenience.

Signal : arbitrary units. Depends on user and experiment but must be **positive** values.

Sigma: the standard deviation estimate of the data point. Depends on the application. Must be provided by the user.

Lines 60 and following contain experiment-dependent information generated by our computer interface. Any line beginning with # will not be taken into account by the program. This may be useful for keeping records.

Important notice about noise

In some applications, it may be quite difficult to obtain the sigma value for each point. Moreover, experiment indicates that such estimates are often much too optimistic. As a consequence, MEM will try to fit many more details than reasonable.

In flash photolysis, a many experiments and simulations convinced us that the actual noise was approximately constant and that sigma ≈ 1e-3, was a much more realistic estimate than the low theoretical values listed in column 3. (This value is not universal, but is of course dependent on the experimental set-up).The reason is that noise is estimated by performing statistics "around" each data point. But for long kinetics, (from 50 ns to 1 sec !) this cannot take "low frequency noise" into account such as a slight drift in light intensity, amplifier gains, sample temperature and/or transparency, etc….Finally, this "low frequency" noise which is not properly accounted for by the statistics, is the dominant one.

See indications about the Sigma box in order to know how to "experiment" with noise.

RESULTS FILE FORMAT

log k	P(log k)	log t	Dexp	MEM fit	Residuals
0	6.11E-04	-7.19791	0.138	0.135757	2.239403
0.090909	6.33E-04	-7.092589	0.137	0.134844	2.151827
0.181818	6.62E-04	-6.99567	0.135	0.133833	1.162839
0.272727	6.98E-04	-6.896196	0.134	0.132599	1.396163
0.363636	7.46E-04	-6.79588	0.133	0.131131	1.864161
0.454545	8.08E-04	-6.696804	0.13	0.12944	0.555751
0.545455	8.90E-04	-6.598599	0.127	0.127511	-0.515748
0.636364	1.00E-03	-6.496209	0.125	0.125227	-0.231291
0.727273	0.00115	-6.399027	0.123	0.122805	0.190793
0.818182	0.001359	-6.297569	0.12	0.120033	-0.035921
0.909091	0.00165	-6.198596	0.118	0.117121	0.876583
1	0.002091	-6.099087	0.114	0.114035	-0.03241
1.090909	0.002739	-6	0.11	0.110865	-0.86644
1.181818	0.003713	-5.899629	0.107	0.107622	-0.623221
1.272727	0.005277	-5.798603	0.104	0.104395	-0.39001
1.363636	0.007723	-5.6989	0.101	0.101313	-0.314969
1.454545	0.011598	-5.598599	0.0979	0.098356	-0.458993
1.545455	0.017623	-5.498941	0.0952	0.095598	-0.40197
1.636364	0.02656	-5.399027	0.0924	0.093021	-0.62636
1.727273	0.038707	-5.299296	0.09	0.090629	-0.635651
1.818182	0.052995	-5.198596	0.0882	0.088374	-0.181361
1.909091	0.066241	-5.099087	0.0865	0.086272	0.220801
2	0.073844	-4.99567	0.0845	0.084181	0.311649
2.090909	0.072492	-4.899629	0.0823	0.082292	0.001905
2.181818	0.062776	-4.79588	0.0802	0.080275	-0.080206
2.272727	0.048759	-4.6989	0.0788	0.078391	0.40488
2.363636	0.034958	-4.59946	0.077	0.076324	0.67378
2.454545	0.023966	-4.496209	0.075	0.074398	0.601598
2.545455	0.016292	-4.392545	0.0726	0.072326	0.274995
2.636364	0.011346	-4.294992	0.0708	0.07037	0.432095
2.727273	0.008309	-4.196543	0.0687	0.068407	0.29533
2.818182	0.006518	-4.097453	0.0663	0.066465	-0.162521
2.909091	0.005533	-4	0.0644	0.06461	-0.208149
3	0.005097	-3.896196	0.0621	0.062718	-0.616372
3.090909	0.005076	-3.798603	0.0604	0.061032	-0.631678
3.181818	0.005415	-3.692504	0.0587	0.059309	-0.610445
3.272727	0.006118	-3.598599	0.0571	0.057875	-0.77051
3.363636	0.00722	-3.497573	0.0556	0.05641	-0.815216
3.454545	0.00878	-3.396856	0.0543	0.055002	-0.708482
3.545455	0.01085	-3.299296	0.0532	0.05365	-0.457992
3.636364	0.01344	-3.198596	0.0522	0.052223	-0.031756
3.727273	0.016464	-3.099087	0.051	0.050733	0.257923
3.818182	0.01969	-2.99567	0.0491	0.049048	0.043003
3.909091	0.022741	-2.899629	0.0474	0.047313	0.078691
4	0.025168	-2.79588	0.0457	0.04521	0.483315
4.090909	0.026595	-2.6989	0.0438	0.042995	0.799499
4.181818	0.02685	-2.59968	0.0416	0.040375	1.221976
4.272727	0.026022	-2.498941	0.039	0.037573	1.426052
4.363636	0.024406	-2.395774	0.0359	0.034324	1.576333
4.454545	0.02239	-2.297569	0.0323	0.030981	1.320914
4.545455	0.020338	-2.196543	0.0282	0.027348	0.855988
4.636364	0.018523	-2.097453	0.0243	0.023678	0.626184
4.727273	0.017115	-2	0.0202	0.020073	0.130811
4.818182	0.016201	-1.899629	0.0157	0.01649	-0.78617
4.909091	0.01582	-1.798603	0.0115	0.013143	-1.64108
5	0.015993	-1.6989	0.00787	0.01021	-2.339061
5.090909	0.016745	-1.598599	0.00504	0.007699	-2.660171
5.181818	0.018114	-1.498941	0.00298	0.005679	-2.701732
5.272727	0.020144				
5.363636	0.022879				
5.454545	0.026322				
5.545455	0.030397				

Note : all logs are decimal.

log k:
Calculated values (equidistant in log) for the number of rates specified for the spectrum range.

P(log k) : rate spectrum
The amplitude for each rate

log t
Log of the time data points of the original data file

Dexp
Signal value of the data file for these points

MEM fit
Best fit of the data by MEM

Residuals
Kinetic Listing Stops at line 58 because only 58 time data points were introduced. But the k listing proceeds to the end.

End of file :
Note again the two lines with 0's to be removed for using the file in most graphic softwares.

90	8	1.00E-04
91	8.090909	1.00E-04
92	8.181818	1.00E-04
93	8.272727	1.00E-04
94	8.363636	1.00E-04
95	8.454545	1.00E-04
96	8.545455	1.00E-04
97	8.636364	1.00E-04
98	8.727273	1.00E-04
99	8.818182	1.00E-04
100	8.909091	1.00E-04
101	9	1.00E-04
102	0	0
103	0	0
104		

APPENDIX
About logarithmic averaging

From Tetreau et al., Biochemistry, 36,10262-10275,(1997).

Since low temperature kinetics extend over several orders of magnitude in time, the use of a logarithmic clock is recommended (Austin et al., 1976). To this end we process the linear time base with our software. Data from several runs at different scan rates are first pieced together after performing an appropriate renormalization based on a numerical integration of areas of two successive and temporally overlapping scans. This procedure minimizes the difficulties brought about by noisy signals. Noise estimates are also renormalized accordingly. We have devised a "logarithmic smoothing" subroutine to convert the data to a logarithmic time base, as required by MEM (see below) while taking advantage of the large number of data points to improve the signal to noise ratio. We arbitrarily define a set of logarithmically spaced times:

$$t_n = t_0 . \exp(\lambda n) \qquad [1]$$

where λ^{-1} is the number of points per log unit. It is easy to show (unpublished results) that the integral average of exp(-kt) between t_{n-1} and t_{n+1} is virtually equal to

$$< e^{-kt} > = e^{-kt_n} . \frac{\sinh(\lambda k t_n)}{\lambda k t_n} \qquad [2]$$

The limit of the second factor is unity within a few thousands if $\lambda^{-1} \leq 10$. In practice, the kinetics are averaged by integration between the data points closest to t_{n-1} and t_{n+1}, and the result is assigned to the (geometric average) intermediate time $t_n = (t_{n+1} . t_{n-1})^{1/2}$. Simulations showed that with 10 to 20 points per decade no detectable error occurred. As long as the kinetics consist of a sum of exponentials, the procedure will yield a limited number of points of great accuracy which reproduce the whole time course without distortion.

oOo

Available online at www.sciencedirect.com

SCIENCE DIRECT

Biochimica et Biophysica Acta 1724 (2005) 411 – 424

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

http://www.elsevier.com/locate/bba

Minireview

Dominant features of protein reaction dynamics: Conformational relaxation and ligand migration

Catherine Tetreau, Daniel Lavalette*

Institut Curie-Recherche, Bâtiment 112, Centre Universitaire, 91405 ORSAY, France

Received 7 March 2005; received in revised form 8 April 2005; accepted 11 April 2005
Available online 13 May 2005

Abstract

Here, we review the dominant aspects of protein dynamics as revealed by studying hemoproteins using the combination of laser flash photolysis, kinetic spectroscopy and low temperature. The first breakthrough was the finding that geminate ligand rebinding with myoglobin is highly non-exponential at temperature $T<200$ K, providing evidence for the trapping of a large number of protein statistical substates. Another major advance was the introduction of a "model free" approach to analyze polychromatic kinetics in terms of their rate spectrum rather than to fit the data to some arbitrarily predefined kinetic scheme. Kinetic processes are identified and quantified directly from the rate spectrum without a priori assumptions. In recent years, further progresses were achieved by using xenon gas as a soft external perturbing agent that competes with ligand rebinding pathways by occupying hydrophobic protein cavities. The first part of this paper introduces several basic principles that are spread throughout a vast literature. The second part describes the main conclusions regarding conformational relaxation and ligand migration in hemoproteins obtained by combining these approaches.

Keywords: Laser photolysis; Ligand migration; Protein relaxation; Kinetics of rebinding; Hemoprotein; Protein dynamics

1. Introduction

Proteins are responsible for most biological functions such as recognition and signaling, transport, chemical transformations and metabolism. To perform its task, a protein molecule usually needs an input that very often is just the binding of a small molecule. Depending on the context, the small molecule is called a ligand, an effector or a substrate. Its binding triggers protein motions that are at the origin of all subsequent events. Binding or releasing a ligand are thus elementary processes giving access to protein dynamics.

Hemoproteins accomplish a variety of functions by binding a ligand. Oxygen transport and storage proteins such as myoglobin (Mb) and hemoglobin (Hb) simply bind and release O_2. Other hemoproteins are involved in a large number of redox and electron transfer reactions or catalyze the oxygenation of substrates such as in various cytochromes P450 that are involved in the detoxification of xenobiotics or in the biosynthesis of hormones. As hemoproteins with enzymatic activity, cytochromes P450 are more complex than oxygen transporters because they also bind a substrate in the heme pocket.

Such a variety of behavior is possible because the active site, the Fe-heme, is a very versatile prosthetic group that may considerably change its chemical reactivity according to the oxidation state of the metal and depending on the number or nature of coordinated ligands.

In this review, we shall consider myoglobin and cytochromes P450 (Fig. 1). Although O_2 is their physiological ligand, it is often difficult to avoid irreversible oxidation or side reactions while studying these very reactive species in vitro. In model studies, carbon monoxide (CO) is often preferred to oxygen as a ligand. The use of CO warrants that one is actually looking at the consequences of binding only because no subsequent chemical reaction of

* Corresponding author. Fax: +33 1 69 07 53 27.
E-mail address: Daniel.Lavalette@curie.u-psud.fr (D. Lavalette).

0304-4165/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbagen.2005.04.024

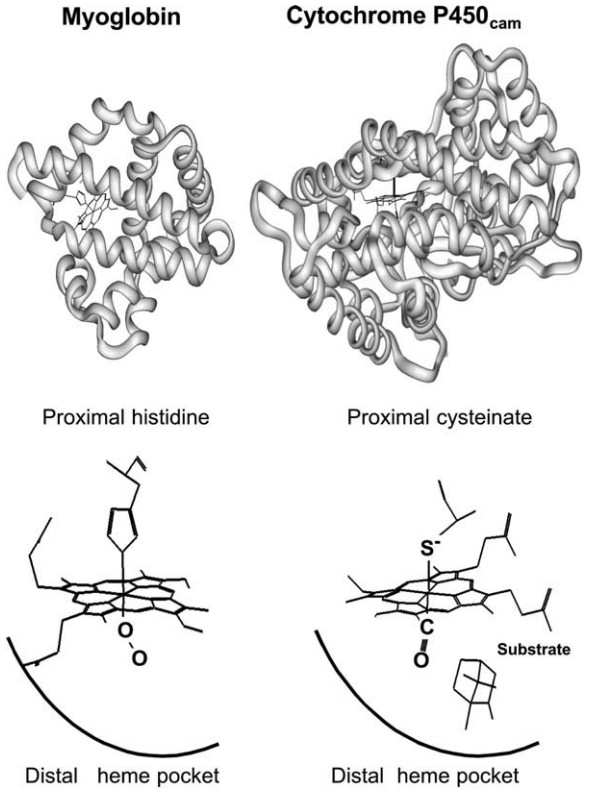


Fig. 1. Views of the active center of myoglobin (left) and cytochrome $P450_{cam}$ (right) obtained using the Insight II (MSI, San Diego) software and the pdb entries 1a6M (Mb) and 3CPP ($P450_{cam}$). Both proteins bind a distal ligand such as O_2 or CO when the central iron atom of the heme is in its Fe^{+2} oxidation state. The protein residue providing proximal coordination is a histidine in oxygen transporters and a cysteinate in cytochromes P450. The enzymatic function of $P450_{cam}$ is to hydroxylate a camphor molecule (substrate).

the ligand can take place, contrary to oxygen. CO generally binds even more strongly than O_2 to the heme and its complexes are extremely stable on the long term (this is the reason for its toxicity because CO quickly replaces oxygen in blood hemoglobin). Cytochromes P450 derive their name from the appearance of a very intense absorbance band (Soret) near 450 nm upon binding CO.

This fact underlines the first useful property of hemoproteins: the heme undergoes strong spectral changes upon binding or releasing a ligand. It is a very convenient spectroscopic marker for monitoring the course of the addition reaction by using fast absorption spectroscopy.

The second, especially useful property of hemoproteins is that heme-ligand complexes are easily dissociated by light. The excess energy provided by one single visible photon is sufficient to cleave the Fe-ligand bond and to release the ligand without any other photochemical damage. This explains why flash photolysis has become the ideal tool to trigger the reaction, especially since short laser pulses have become available in the seventies.

The reaction partners, the free heme protein and the free ligand, are generated in situ by a laser pulse that dissociates the ligand–protein complex on a subpicosecond time scale. The rebinding reaction is monitored spectroscopically by using the absorbance changes of the heme.

As we shall see, the interest is of course not in the addition reaction but rather in the protein dynamics that may affect its rate. One actually seeks the kinetic simplicity of a true first-order or pseudo first-order reaction because it provides, at each temperature, an internal time standard to which protein motions can be compared.

2. The typical laser flash photolysis experiment

Following photodissociation of the protein–ligand complex by the laser pulse, the sudden absorbance change observed at $t=0$ quickly recovers as ligands rebind (nanosecond scale). This phase ends within a few hundreds of nanoseconds at most. By that time, absorbance recovery is usually not complete. The residual absorbance ultimately recovers completely, but very slowly. This phase must be recorded at a higher sensitivity and scanned at a slower speed. After full rebinding, the experiment can be repeated and always yields identical results: the system is reversible and the protein–ligand complex remains active (Fig. 2).

The fast initial phase is due to geminate recombinations, i.e., when the dissociated ligand rebinds without leaving its partner protein. Being a purely internal, first-order process, its rate parameter k_{gem} is independent of ligand concentration.

On the contrary, the rate of the slower phase is proportional to the concentration of free ligand in the surrounding environment, indicating that a *bimolecular* reaction between free ligands and deliganded proteins is taking place. These two processes can be combined into one single expression for the survival fraction $N(t)$ of proteins that have not recovered their ligand:

$$N(t)=(1-N_{esc})\exp(-k_{gem}t)+N_{esc}\exp(-k_{bim}t). \quad (1)$$

In Eq. (1), N_{esc} is the fraction of ligands that have escaped their protein before geminate rebinding occurred. Once its ligand has escaped into the solvent, the protein molecule must wait for an encounter with a new ligand

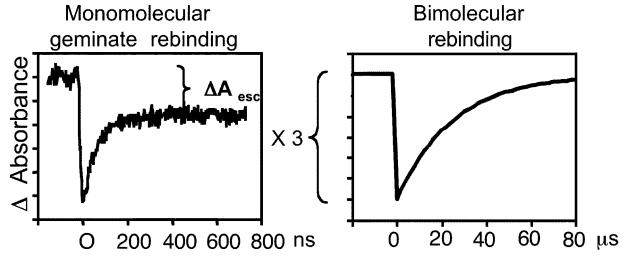


Fig. 2. A typical laser photolysis experiment. Photodissociation of the ligand induces an instantaneous absorbance change. Geminate rebinding (left) is a fast first-order process, but the absorbance recovery is not complete due to a fraction N_{esc} of ligands that have escaped into the solvent. The subsequent final recovery is slow because it involves a bimolecular reaction between a ligand-free protein and a free ligand molecule (right).

molecule. This gives rise to the slow bimolecular phase. In the experiments, the protein concentration is typically about 20 μM; the free CO gas solubility is of the order of 1 mM at 293 K and 1 atmosphere. Because of the large excess of ligand, the rate of the bimolecular phase is pseudo first-order $k_{bim}=k\ [L]$.

Interestingly, N_{esc} rapidly decreases upon increasing the solvent viscosity at constant temperature [1,2]. This is a strong indication that motions at the protein surface are involved in the escape process. These motions are strongly damped in a viscous surrounding. When they become too slow for allowing escape, all ligands rebind via the geminate path.

3. Protein kinetics are complex because proteins are degenerate systems

As a polymer with N atoms, a protein molecule has 3N-6 degrees of freedom. Even with small proteins, this is a huge number and many vibrational modes are densely close in energy. Proteins are thus highly degenerate systems and this degeneracy introduces a remarkable complexity. Because a protein molecule is not isolated in space but is bathed in a solvent that collides at high frequency, perturbations induce permanent transitions between all low-lying energy states located within the range of ≈kT. In the thermodynamical view, one may say that the system has a shallow free energy well with a rough structure involving many small relative minima separated by small barriers that are easily overcome ($\Delta G<$ or ≈kT). The system then appears to jump randomly from one minimum to another.

Fluctuations are much faster than any laboratory recording device. In particular, they are considerably faster than the ligand binding reaction we wish to use for studying the protein dynamics. As a consequence, the reactions rates obtained in the experiment shown in Fig. 2 are actually ensemble averages. Because the ensemble is very large, the average is very accurate, but information about the individual behavior of protein molecules is lost. This can be illustrated with a simple example.

Suppose a molecule M fluctuating between two substates A and B that are not spectroscopically distinguishable but rebind at different rates according to $d[A]/dt=-k_A\ [A]$ and $d[B]/dt=-k_B[B]$ respectively. Two important limiting situations must be considered.

Assume first that fluctuations between A and B are much slower than either decay rate. Each substrate will then decay independently of the other and the overall rebinding kinetics of M will be the sum of two exponential terms:

$$[M(t)]=[A_0]\exp(-k_A t)+[B_0]\exp(-k_B t) \quad (2)$$

from which the rate parameters of A and B can be easily extracted.

In the opposite, assume that fluctuations are much faster than either decay rate so that A and B permanently equilibrate and therefore remain at any time a constant fraction, respectively α and β, of M. The kinetic equation for the overall rebinding of M is now: $d[M]/dt=-\alpha k_A[M]-\beta k_B[M]$. After an obvious factorization, M is found to rebind according to a single exponential:

$$[M(t)]=[M_0]\exp(-[\alpha k_A+\beta k_B]t) \quad (3)$$

the rate parameter of which is the weighted average of the rebinding rates of the individual substates. The rate parameters of A and B cannot be extracted from the data anymore.

Eq. (3) describes the process called kinetic averaging and explains why the rebinding kinetics of Fig. 2 are single exponentials corresponding to an ensemble average in spite of the fact that the protein fluctuates between many conformational substates. In a fluid solvent near room temperature, fluctuations are much faster than rebinding. Eq. (2) indicates that information about the rebinding rate of individual substates may be regained provided we make fluctuations slower than rebinding or even, whenever possible, if we are able to stop them completely.

The most straightforward solution is to cool the protein at a very low temperature. In the 60s, Pierre Douzou et al. in Paris [3] had already laid down the principles of cryobiochemistry. Proteins remain native and active in carefully selected hydro-organic mixtures that act as cryoprotectants. Some of these mixed solvents form even clear solid glasses below their glass transition temperature, T_g, thus permitting to perform spectroscopic measurements continuously from room temperature down to liquid helium.

The viscosity of cryoprotective solvents increases dramatically upon cooling and N_{esc} quickly vanishes. Rebinding becomes purely geminate. Geminate recombinations play a central role in low temperature experiments because they are independent of external diffusion and thus remain observable even when the solvent becomes a solid glass.

Even at temperature above the glass transition, internal protein fluctuations become strongly damped and come to a stop. The protein ensemble may be compared to the Castle of the Sleeping Beauty of the fairy tale: each individual molecule remains frozen in a particular attitude (conformational substrate). Because reaction rates are coupled to structure, geminate recombinations take place at a different rate in different substates. The ensemble geminate kinetics become a superposition of an infinite number of exponentials:

$$N(t)=\sum_j a_j\exp(-k_j t) \quad (4)$$

summed over all substates j reacting with rate k_j.

Summing over k (a measurable quantity) rather than over the substates (that are unknown), the kinetics are better described by:

$$N(t)=\int_0^\infty P(k)e^{-kt}dk \quad (5)$$

in which $P(k)\,dk$ is the fraction (probability) of substates reacting with rate k.

By freezing all motions, we have replaced the ensemble averaged reaction rate by an ensemble of reaction rates. The example displayed in Fig. 3 shows that the ensemble kinetics are inhomogeneous or polychromatic. Note the extraordinary time range of the kinetics covering 6 to 7 decades. It means that among otherwise identical protein molecules, some rebind their ligand within nanoseconds, others need almost a second to do so depending on the particular conformation they were frozen in.

The group of Hans Frauenfelder at Urbana Champaign first performed such low temperature photolysis experiments with myoglobin and laid down the concept of Conformational substates (CS) in 1975 [4]. (Nowadays, they are rather called statistical substates). Observation of polychromatic kinetics with other, non-heme, proteins further supported the generality of the concept [5,6].

4. Thermodynamic parameters of Conformational substates

Before trying to analyze the consequences of Eq. (5) one would like to answer the question of the origin of the distribution $P(k)$ of reaction rates among substates. To a first approximation, a reaction rate is given by the Arrhenius relation:

$$k = A\exp(-H/RT) \tag{6}$$

where H is the activation enthalpy of the reaction ($\approx$ the potential energy barrier) whereas the preexponential term A is related to the activation entropy. Both may be affected by changes of the protein structure in the vicinity of the heme. If only A was distributed, the Arrhenius plots for the CS would remain parallel at all temperature; therefore the width of the rate distribution $P(k)$ would remain constant. If only H was distributed, the Arrhenius plots of the CS would diverge as temperature is lowered and the $P(k)$ distribution would strongly broaden (Fig. 4). Experiments unambiguously show that $P(k)$ widens considerably at lower temperature. Even if both A and H are distributed (and there is no reason why A should not be), the temperature effect due to H would dominate. Simulations have confirmed that even if A was also distributed, no experiment would be accurate enough to detect this fact [7].

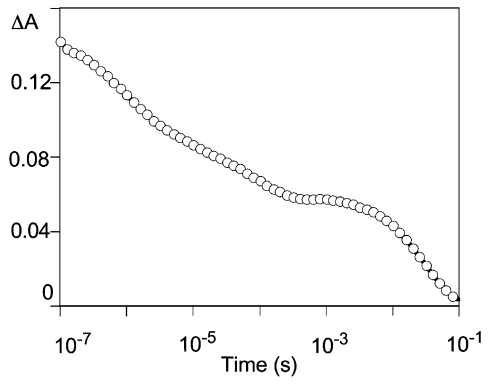


Fig. 3. Example of the polychromatic kinetic trace for geminate CO rebinding with Mb in a glycerol/water glass at 250 K. Note the logarithmic time scale. Dots are logarithmically averaged data points.

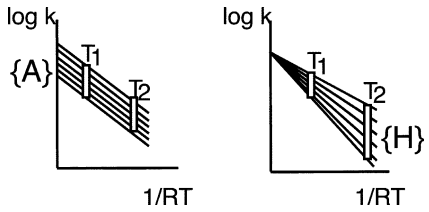


Fig. 4. Arrhenius plots of the distributed rebinding rate parameter k as expected if only the pre-exponential factor A (left) or only the activation enthalpy H (right) were distributed in the ensemble of statistical substates. For clarity, only a few statistical substates are considered. Vertical bars visualize the total width of the rate parameter distribution at two arbitrary temperatures T_1 and T_2.

5. Enthalpy distribution

For several years, there was no really adequate method for analyzing such complex kinetics extending over many time decades. It was reasoned that once the protein ensemble is completely frozen, the distribution of statistical substates, and hence that of enthalpy, do not change upon further lowering temperature. Therefore, at temperature lower than T_g, Eq. (5) can be specialized to:

$$N(t) = \int_0^\infty P\left(H; T_g\right) e^{-kt} dH \quad \left(T < T_g\right). \tag{7}$$

Clearly, at $T < T_g$, the protein ensemble is a non-equilibrium ensemble. Only reaction rates still continue to become slower due to their temperature dependence. Using the Arrhenius relation (Eq. (6)) and postulating some trial functions for $P(H;\ T_g)$ a multi parameter least square procedure could be performed to fit the real kinetics. In this way, enthalpy distributions could be estimated.

6. A strategy

$P(H;\ T_g)$ yields an image of the distribution of CS only for T_g, the temperature at which all protein motions were frozen. But our aim is to find out those protein motions that accompany, modulate or interfere with the binding reaction. To this purpose, they must be reactivated by raising temperature in discrete steps and performing repeated rebinding measurements. It is hoped that relevant motions will reappear more or less sequentially. Indeed, the polychromatic kinetics undergo intriguing changes revealing that protein motions are interfering with rebinding. The most striking example is the counterintuitive "inverse temperature dependence" displayed in Fig. 5 [8,9]. The

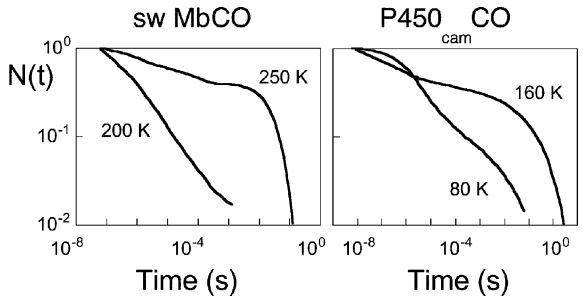


Fig. 5. Examples of the apparent "inverse temperature effect". The overall rebinding seems to become slower upon raising the temperature because the faster processes are progressively replaced by new, slower ones at higher temperature. Note that the temperature range in which the effect is observed differs with the protein.

overall rebinding kinetics seem to slow down when temperature is raised!

At $T > T_g$, however, the simplifying assumption of a constant enthalpy distribution does not hold, as thermal equilibrium of the protein ensemble is restored. $P(H, T)$ becomes temperature dependent and additional (i.e. non-statistical) motions may interfere in a unpredictable way.

To cope with this new complexity, more powerful tools were needed.

7. The kinetic rate spectrum $P(k)$ and the Maximum Entropy algorithm

At any temperature, the $P(k)$ distribution contains all information that can be objectively extracted from the kinetic traces without making any a priori assumption. All details about rebinding species and processes in the protein ensemble are contained in this unknown distribution of the rate parameters. In the late 80s, this problem was solved by the Maximum Entropy Method, a mathematical algorithm based on information theory [8,10–13]. Fig. 6 shows a polychromatic kinetic trace and its associated $P(k)$ distribution. We call $P(k)$ the "kinetic rate spectrum".

First, note the log scales. As previously shown in Fig. 3, polychromatic kinetics cannot be visualized on a linear scale. Symmetrically, the rate spectrum is also shown on a log scale. This amounts replacing Eq. (5) by its equivalent:

$$N(t) = \int_0^\infty P(\log k) e^{-kt} d(\log k). \quad (8)$$

Reading the kinetic rate spectrum is considerably more informative than just looking at the kinetic trace, although the information content is exactly the same. The rate spectrum is comparable to the Fourier frequency spectrum of a sound or any electric signal, except that we are not dealing with frequencies of elementary periodic functions but with reaction rate parameters of exponentials. For instance, there is no difficulty for identifying in the spectrum of Fig. 6 three main kinetic processes, with well distinct reaction rates (the peak values) and a certain bandwidth showing that each is being still distributed among CS. If necessary, the spectrum can be analyzed in terms of Gaussian (or rather log-normal) band shapes to evaluate the amplitude and distribution width of each process. If kinetics are recorded at different temperature, the temperature dependence of each process can be followed individually. This is a wealth of information obtained without any a priori model or assumption, a fact that distinguishes this new approach from the former parametric fitting, that always require some a priori kinetic model.

8. A gallery of kinetic processes

Fig. 7 presents in the form of rate spectra all the kinetic knowledge about myoglobin [8,13–16] and cytochrome $P450_{cam}$ [9,17,18] gathered over many years. It clearly shows that several processes appear and disappear or replace each other between ≈100 K and 300 K. Although these changes occur at different temperature with different proteins and/or ligands, some features are common. We

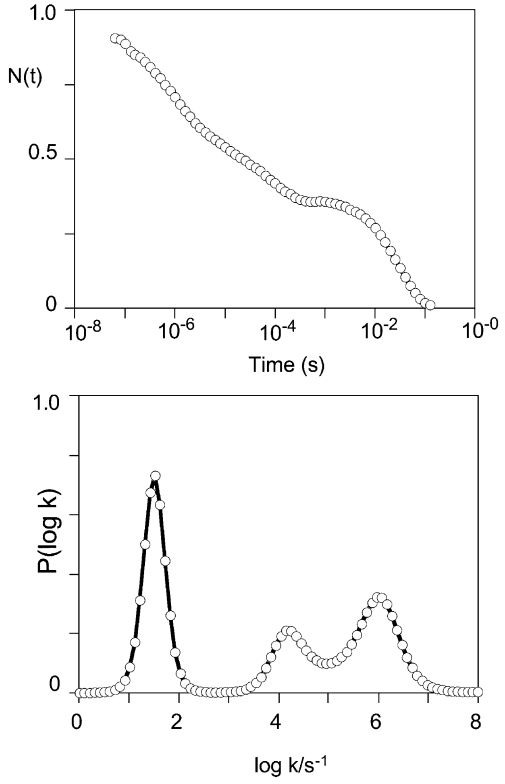


Fig. 6. Top: the polychromatic kinetic trace for CO geminate rebinding with Mb at 250 K in a glycerol/water glass. Bottom: the corresponding rate spectrum $P(\log k)$, obtained by inversion of the Laplace transform (Eq. (8)) using the maximum entropy method. The spectrum displays very clearly the contribution of three kinetic processes peaking at log k values of about 1.5, 4 and 6, respectively. In particular, the two latter peaks resolve what appears only as a slight undulation in the fast decreasing part of the kinetic trace. The rate spectrum was developed on a basis set of 100 logarithmically spaced rate parameters.

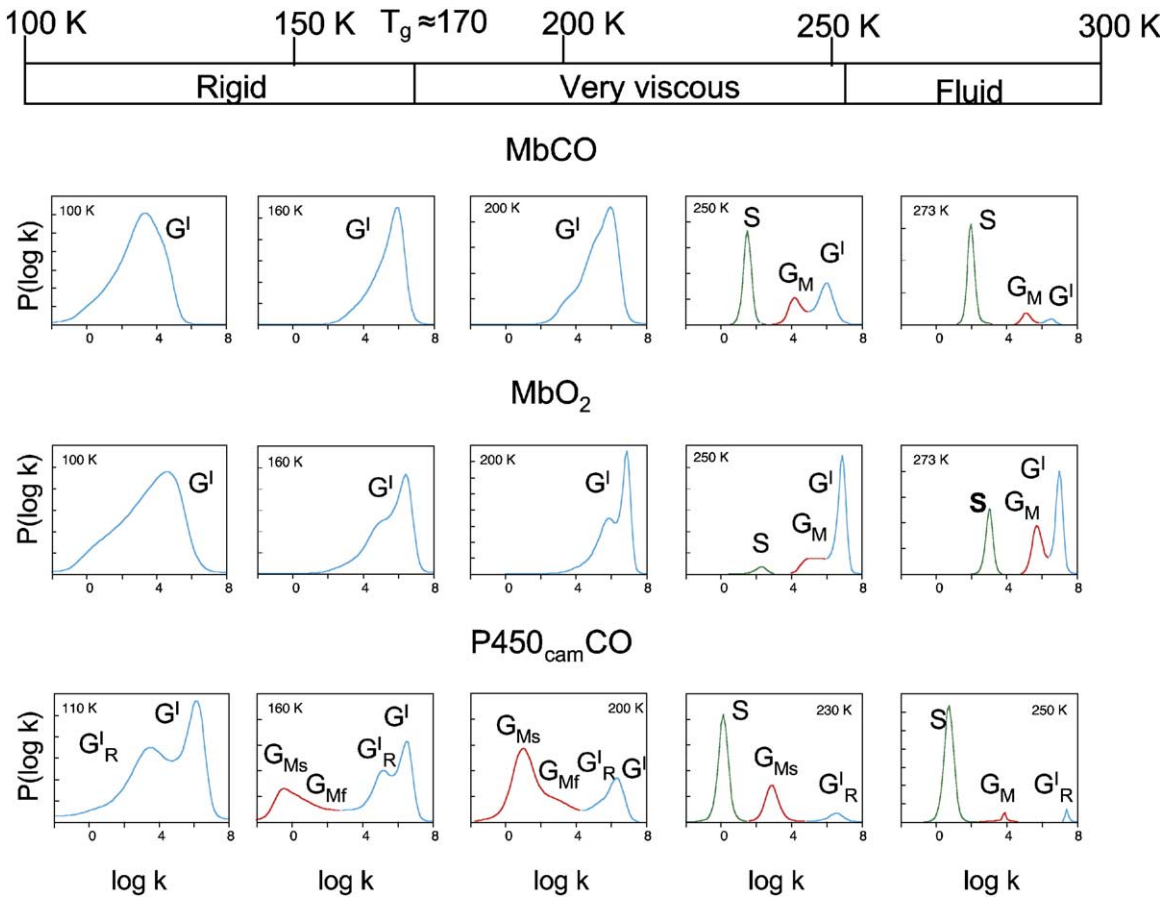


Fig. 7. Overview of the rebinding kinetics of MbCO, MbO_2 and $P450_{cam}$ (CO) in the temperature range 100 K–273 K. Colors and symbols permit to follow the various kinetic processes discussed in the text. Blue : geminate rebinding from the primary ligand docking site, before (G^I) and after (G^I_R) heme pocket relaxation. Red : geminate rebinding from secondary docking site(s) after ligand migration (G_M, G_{MF}, G_{MS}). Green : bimolecular rebinding subsequent to initial escape of the ligand into the solvent (S). Note that primary rebinding, already complex in Mb, is visibly bimodal in $P450_{cam}$ (CO). In Mb, this is due to the superposition of the distributions of taxonomic substates. In $P450_{cam}$ (CO) the slower band (G^I_R) corresponds to rebinding after the heme pocket has relaxed. The spectral narrowing observed as temperature increases has two causes : the Arrhenius dependence of the rate parameters (see Fig. 4) and, above 200–250 K, kinetic averaging (Eqs. (2) and (3)). Ligand migration is observed from about 140 K onwards in $P450_{cam}$ (CO), but only near 240 K in Mb. CO escape appears near 200 K in MbCO and $P450_{cam}$ (CO). In MbO_2., geminate oxygen rebinding with hemes being intrinsically about tenfold faster than CO rebinding, remains too fast at 200 K for escape to compete. All rate spectra are normalized, i.e., their total area is equal to unity. The importance of individual processes is given by their band area. They are therefore directly comparable among proteins and over the whole temperature range.

shall first state their assignments before describing the experimental evidence in detail in the later sections.

At very low temperature ($T<150$ K), all protein motions are suppressed. Upon photodissociation the ligand moves to a (presumably) unique primary docking site located somewhere between heme and protein and from which no other alternative exists but direct rebinding with the heme. From 77 K upwards, the primary rebinding process (blue curves) can be followed continuously in the rate spectra. As expected, it shifts to higher rates and its width narrows as temperature increases simply by virtue of the Arrhenius law. Surprisingly, even at very low temperature, the rate spectrum of primary rebinding displays a "fine structure" that is particularly marked in cytochrome $P450_{cam}$. This "fine structure" has a different origin in each protein. In Mb, the band shape is due to a superposition of rebinding spectra corresponding to protein taxonomic substates. In cytochrome $P450_{cam}$, a dynamic conformational relaxation of the heme pocket is responsible for the bimodal spectrum.

From ≈140 K onwards in $P450_{cam}$, but only at ≈230 K in Mb, a new, slower kinetic process appears at the expense of primary rebinding (red curves). In both circumstances, kinetic competition of the ligand with xenon indicates that these processes correspond to "delayed ligand rebinding" in which the ligand migrates within protein cavities before rebinding.

Finally, above ≈230 K, the solvent becomes sufficiently fluid to restore surface fluctuations of the protein and to allow for ligand escape and bimolecular rebinding (green curves). At room temperature, this becomes the dominant process and geminate rebinding is difficult to observe because it is either practically absent or very fast.

These few kinetic rate spectra thus exhibit a series of kinetic processes for which we now have to provide explanations. Before going into details, let us first consider what

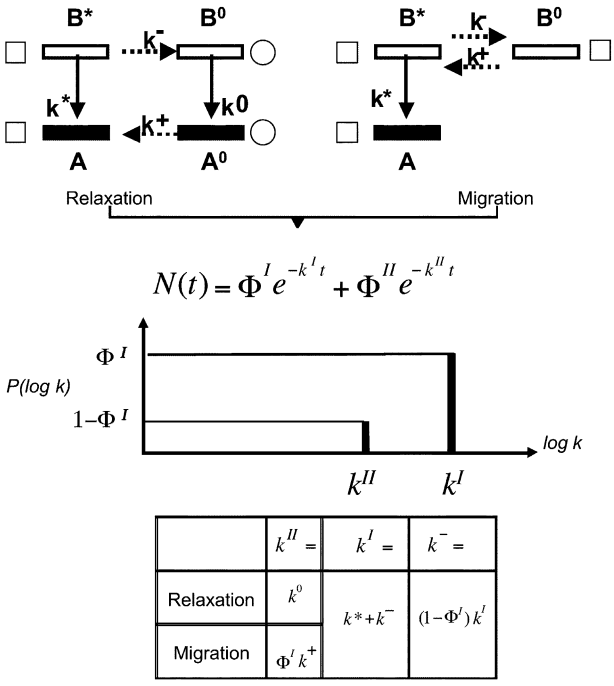
may happen to a protein after binding a ligand. First, the uv-vis spectrum of the heme changes as already mentioned. But, in addition, the presence of the ligand may possibly perturb more or less the local structure in the surrounding of the active site. Broadly speaking, one may assume that there is a ligand-free structure and another, ligand-bound structure and that binding necessarily induces a local conformational change. At low temperature, the active site retains the "bound-" structure immediately after dissociation. Whenever some internal motions become activated, the active site is expected to "relax" to its "ligand-free" structure. So, depending on the relative rates, ligand rebinding may take place either with the initial "bound-" structure or with the relaxed "free-" structure. This process, called conformational relaxation, operates in cytochrome $P450_{cam}$ [9,18].

Another feature of low temperature kinetics is that in the frozen protein the dissociated ligand is initially blocked in a particular (primary) docking site not far removed from the active center. But there may be other possible docking sites in the protein. Above a certain temperature, the ligand may begin to migrate from one docking site to another one before it ultimately rebinds. This migration process competes with direct rebinding. It gives rise to "delayed rebinding". It has been first found in Mb [16,19,20] but occurs in other proteins as well.

How can one experimentally distinguish conformational relaxation and ligand migration? Fig. 8 displays the simplest possible kinetic schemes in the ideal case of non-distributed reaction rates. Although they are topologically very different (parallel for relaxation and sequential for migration) both have a similar analytical solution predicting two "lines" in the rate spectrum: a fast and a slow one. Unfortunately, although the physical meaning of the elementary rates differ, the formal solutions look quite alike and do not permit a distinction based on kinetic data only. Additional information is required. This missing information has been provided by introducing xenon gas as a new partner into the kinetic game.

9. Fine structure of the rate spectrum of CO primary rebinding

Crystallographic experiments with trapped intermediates [21–23] and time-resolved X-ray crystallographic studies of MbCO [24–27] have allowed to localize the CO primary docking site above pyrrole C at 3.6 Å from the iron and parallel to the heme plane. In the absence of any information available about the CO photolysis intermediates, the primary docking site of cytochrome $P450_{cam}$ still remains unknown, but is likely to be near the heme, above one pyrrole group. The blue curves in the rate spectra of Fig. 7 correspond to rebinding from the primary site. Clearly for both proteins, the rate spectrum of primary rebinding displays features suggesting a underlying fine structure. As we shall now see, the reason is, however, profoundly different in Mb and in cytochrome $P450_{cam}$.



	k^{II} =	k^{I} =	k^{-} =
Relaxation	k^{0}	$k^{*}+k^{-}$	$(1-\Phi^{I})k^{I}$
Migration	$\Phi^{I}k^{+}$		

Fig. 8. Top : the basic kinetic schemes for protein relaxation (left) and for ligand migration (right). The ligand is bound to the heme in the black states and occupies other docking sites in the white states. Squares and circles indicate respectively that the heme pocket is in the liganded or deliganded structure. Middle : despite the fact that relaxation and migration correspond respectively to parallel and sequential schemes, the analytical expression obtained for the rebinding (survival) function is the same, as well as the predicted rate spectrum. Bottom : only the meaning of the overall rate parameters in terms of elementary transitions differs according to the expressions given in the table.

9.1. The taxonomic substates of myoglobin

Mb has been the object of considerable work implying many laboratories during the last 30 years. Although the rate spectrum of primary rebinding looks more or less unimodal, its shape is sufficiently modified with temperature to suspect that it may be composite.

Myoglobin has been shown to exist in three "taxonomic substates" (CS 0) A_0, A_1 and A_3 (Fig 9, top) [28]. These are in fact local isomers differing slightly in the interactions of the ligand with the distal histidine [29,30]. They could be characterized in MbCO by IR spectroscopy because they display different CO stretching bands. They were even found to rebind at different rates when rebinding was monitored in these IR bands [31]. Though very likely to exist, taxonomic substates of MbO_2 were unknown because of the unfavorable spectral region and intrinsic weakness of O_2 stretching bands [32].

The Soret band (≈400 to 450 nm) of the heme where rebinding is generally monitored does not distinguish among CS^0 because the heme uv-vis spectrum is essentially sensitive to the state of coordination of the Fe(II) atom. However, all CS^0 contribute to the Soret and if their rebinding rate differs, one should be able to observe some kinetic consequences. Because the taxonomic substates do

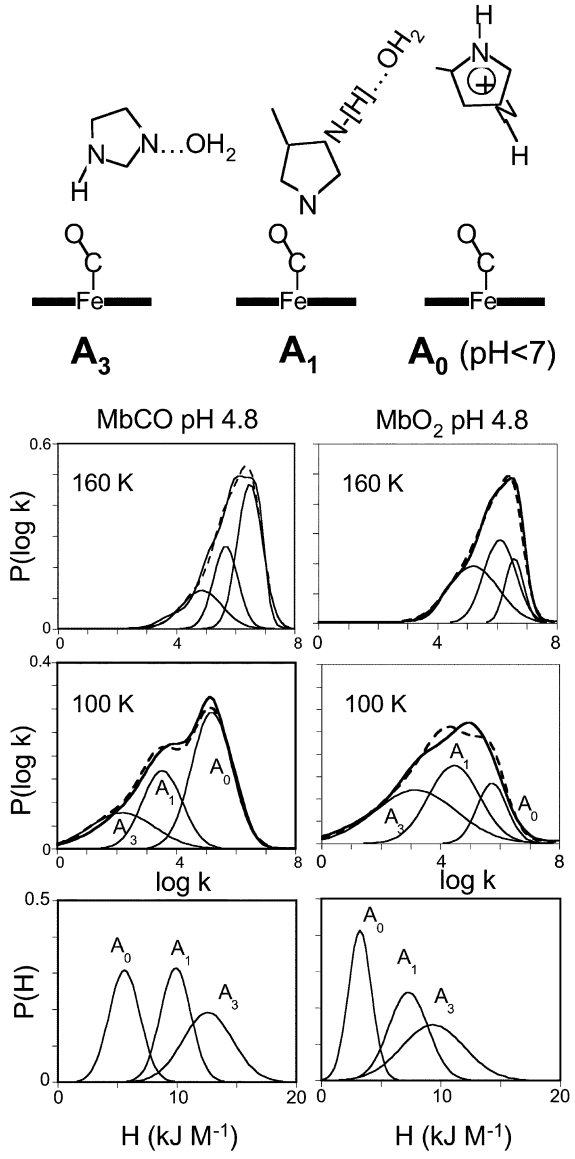


Fig. 9. Analyzing the rate spectra of the primary geminate rebinding in MbCO and MbO_2 at 160 K and 100 K at acidic pH. At both temperatures, the spectra are nicely fitted with three 'log normal' distributions assumed to originate from three temperature-independent enthalpy distributions (bottom). At pH 7 only distributions A_1 and A_3 are needed (not shown). The absence of A_0 at pH 7 as well as the enthalpy distributions are in agreement with their attribution to taxonomic substates of MbCO (left) [31]. The data also revealed for the first time the taxonomic substates of MbO_2 (right). All kinetics were recorded in the visible (Soret) region.

not interconvert at low temperature, each CS^0 remains characterized by one unique enthalpy spectrum $P(H; Tg)$ at $T<Tg$. To verify this point, we developed a software allowing to perform a global MEM analysis of P(H) fitting simultaneously all P(k) recorded at different temperature [33]. For both CO and O_2, the best fit was obtained with two components at pH>7, while three components were required at pH<7 (Fig. 9). This is consistent with the known pH dependence of A_0 in MbCO that is populated only when the distal histidine is doubly protonated [31,34]. In addition to proving that CS^0 can be observed in the visible part of the spectrum, the results provided for the first time an experimental evidence for the existence of CS^0 in MbO_2 too (Fig. 9, right).

Above ≈160 K taxonomic substates begin to interconvert. The rate spectrum becomes complex and cannot be readily analyzed until ≈200 K. Here, a sudden simplification occurs: the rate of CS^0 interconversion and the rate of statistical substates fluctuations becomes faster than rebinding. The consequence is a band narrowing due to kinetic averaging: all CS^0 behave kinetically as one single average species.

9.2. Conformational relaxation in cytochrome $P450_{cam}$

Let us now consider cytochrome $P450_{cam}$ at very low temperature. Here, primary rebinding clearly consists of two widely separated bands. But surprisingly, their amplitude varies even at temperature as low as 80 and 110 K (Fig. 10). These bands cannot correspond to taxonomic substrates because the reported IR CO stretch spectrum is absolutely temperature independent below Tg in $P450_{cam}$ [35,36].

Because of the low temperature at which both bands equilibrate, the passage from one to the other can only involve the motion of small parts of the heme pocket. Similar observations were reported not only for cytochrome $P450_{cam}$ but also for cytochromes $P450_{SCC}$, $P450_{LM2}$ and for NO synthase [17,37]. All these hemoproteins performing catalytic functions bind a substrate in the heme pocket in addition to the O_2 ligand. This led us to suspect that the substrate might be at the origin of the bimodality, and the hypothesis of substrate involvement was further reinforced by an examination of the X-ray structure reported for different forms of cytochrome $P450_{cam}$ [38,39].

To clarify this issue, we have energy-minimized the Fe(II) structure in presence and in absence of ligand [18]. As anticipated, the camphor substrate moves like a pendulum toward the heme normal when the ligand is released (Fig. 11). As a consequence subsequent rebinding will meet with a greater steric hindrance and will be slower. This is a genuine conformational relaxation that competes with rebinding.

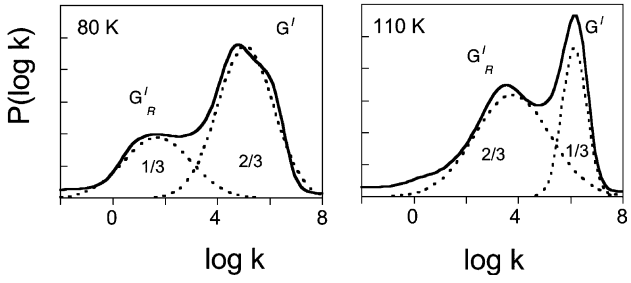


Fig. 10. Normalized rate spectra for primary rebinding in $P450_{cam}$(CO) at the lowest temperatures. The Gaussian (log normal) decomposition shows that two processes are present and that their relative weights, measured from their band area, are reversed between 80 K and 110 K. Process G^I corresponds to ligand rebinding from its primary site before heme pocket relaxation occurs. Process G^I_R corresponds to rebinding with the relaxed heme pocket.

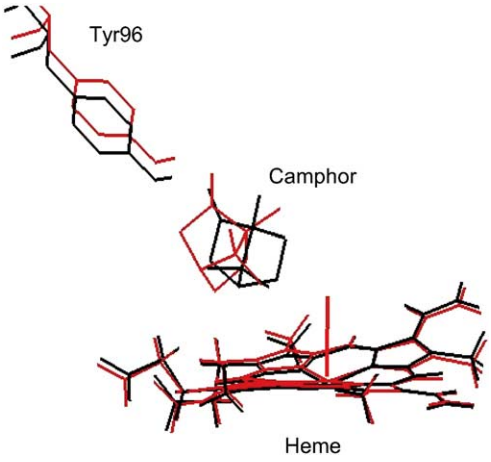


Fig. 11. Heme pocket conformational relaxation in $P450_{cam}$. These energy-minimized structures of Fe(II)-$P450_{cam}$ (black) and Fe(II)-$P450_{cam}$(CO) (red) show that upon CO deligandation, the substrate is displaced by about 1 Å closer to the CO-binding axis.

Working out the rate parameters of relaxation in presence of substrates analogs (Section 11) confirms the attribution. In cytochrome $P450_{cam}$, camphor is maintained in the heme pocket only by one hydrogen bond with Tyr96 and a few non-bonding interactions. Its motion is a consequence of minimizing steric interactions with the heme propionates, with or without the ligand present. The energy requirements are small. In addition, camphor is buried in the heme pocket and therefore not exposed to the solvent rigidity. These factors explain why conformational relaxation can take place already at very low temperature.

In conclusion, the fine structure of CO primary rebinding with cytochrome $P450_{cam}$ may be assigned to rebinding before (G^I) and after (G^I_R) substrate relaxation toward a more crowded position.

10. Using xenon as a probe of ligand migration in proteins

Fig. 12 (top) summarizes what was known a few years ago about the docking sites of CO in Mb. These data refer to X-ray crystallography of trapped intermediates in Mb crystals after prolonged steady state illumination and to time-resolved studies of the photolyzed MbCO species. Depending on temperature, the ligand was found at essentially two locations: in what is called its primary docking site near and above the heme plane (B) and also, in one location on the proximal side (E) [21–27]. In the particular case of a Mb mutant, CO was also found (around 100 K) in a distal position (C) [40]. This was unexpected and immediately raised the question: how was this possible?

It was also previously known from crystal studies that a small number of hydrophobic cavities exist in Mb that are able to bind the rare gas xenon [41]. Three of them are located in the vicinity of the heme: Xe1, Xe2 and Xe4. As shown in Fig. 12, the locations where the ligand has been found coincide with some of these xenon sites.

The presence of multiple ligand docking sites corresponding to protein cavities has been further supported by kinetic competition experiments performed in presence of xenon [16,19,20,42]. Several approaches including kinetic studies of mutants [19,20,43–46], molecular dynamics calculations of ligand trajectories [42,43,45–50], FTIR and TDS spectroscopies [51,52] have converged to a consistent description of ligand migration in Mb along preferred pathways toward well-defined cavities.

We considered that the xenon/ligand kinetic competition approach was worth to be generalized. To this purpose, the protein–CO complex was pre-equilibrated several hours at 5 °C with various xenon gas pressures. Rapid cooling ensured that xenon concentration remained constant in the viscous/solid glass. Rebinding kinetics were then recorded at different temperature. As shown in Fig. 13, xenon pressure induces a smooth change in the rate spectrum of Mb. Band G_{Ms} progressively disappears while a new process (G_{Mf}) appears at higher rates. In Mb cavity Xe1 has by far the highest affinity for xenon and is the only one to be occupied at the relatively low pressure used. So, it was concluded that upon blocking the proximal Xe1

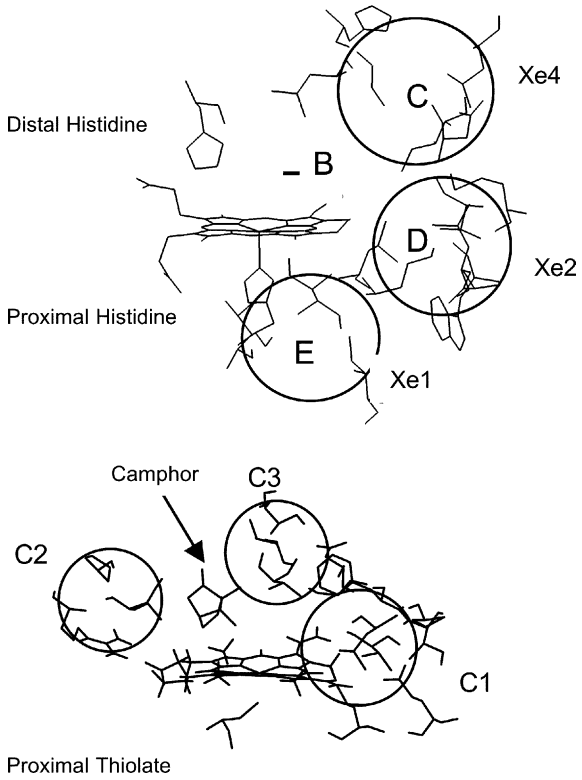


Fig. 12. Protein cavities surrounding the heme pocket region in Myoglobin (top) and cytochrome $P450_{cam}$ (bottom). In Myoglobin (Pdb entry:Mb:1dws), three cavities are known to correspond to sites that may be occupied by Xenon. Cavities C_2 and C_3 calculated from the energy-minimized structure of $P450_{cam}$ are probably too small to accommodate a Xenon atom and cavity C_1 is occupied by three water molecules.

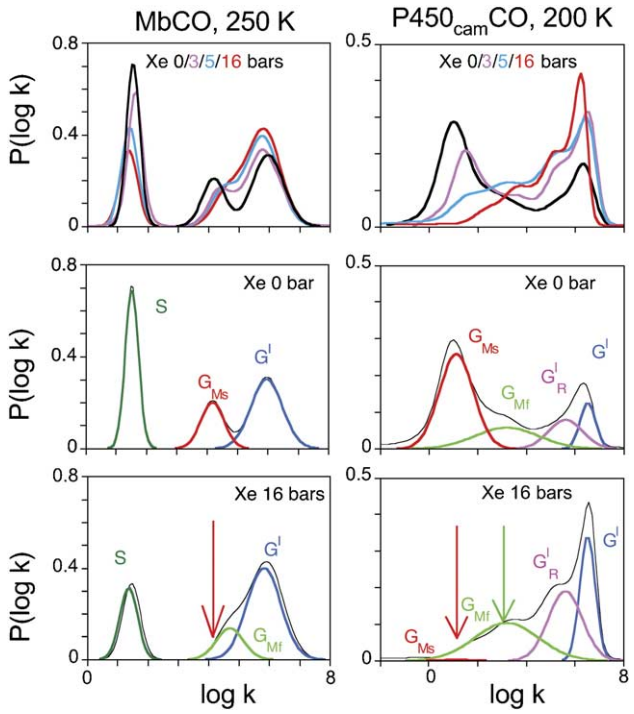


Fig. 13. Top : rate spectra for CO rebinding under a variable xenon pressure. The Gaussian decomposition of the rate spectra in the absence of xenon (middle) and at the maximum pressure of 16 bar (bottom) explicits the changes brought about by xenon on the various kinetic processes.

cavity, GMs disappears in favor of the new and faster GMf process.

Let us now consider ligand escape. Xenon was found to have no effect on MbO_2 near room temperature, in agreement with a conclusion reported by Scott and Gibson some years ago [19,20]. But unexpectedly, a very significant xenon effect was observed in MbCO (see Fig. 13) [16]. It turns out that N_{esc} (measured from the band area of the bimolecular rebinding) is composed of a Xe-dependent and a Xe-independent part that nicely titrates with Xe pressure. It was even possible to calculate the fraction of escape that is xenon dependent, in other words, that occurring in the proximal Xe1 cavity and the fraction escaping already in the primary site via the distal histidine gate pathway. It was found that CO escape via Xe1 is negligible at room temperature, increases to a maximum at 260–250 K and then decreases again to zero because viscosity has become too high. At 260–250 K, half of the ligand escapes from the Xe1 cavity.

These findings are consistent with what may be expected for a ligand migration process across the three xenon cavities in series. Immediately after photodissociation, the ligand is docked in the primary site (B) from which three events may occur: (i) the ligand may directly rebind: this gives the fast G^I geminate rebinding; (ii) the ligand may also escape into the outside, presumably by what is known as the histidine gate pathway; (iii) the ligand may also start migrating, occupying only transiently Xe4 and Xe2 to finally accumulate in Xe1. The return is of course a slow process that gives rise to the geminate delayed rebinding band G_{Ms}. Because the residence time of CO is long in the Xe1 cavity, about half of the ligand may escape the protein in this site even though the escape rate is low. When Xe1 cavity is occupied by a xenon atom, migration cannot proceed beyond Xe2. The ligand now rebinds about 10 times faster (G_Mf) because its migration path is shorter and because one rate limiting step (Xe1->Xe2) has been removed. In addition, no ligand can escape from Xe1 anymore and the yield of the bimolecular rebinding is therefore decreased. All these data therefore provided a clear kinetic evidence that in Mb internal ligand migration takes place as discrete jumps from one cavity to the other.

The situation is significantly more complex in cytochrome $P450_{cam}$ where four processes are observed simultaneously at 200 K. This temperature is a convenient one because ligand escape has not yet set in and all processes are geminate (Fig. 13). G^I and G^I_R are primary rebinding, respectively, before and after substrate relaxation as described above. In addition, two bands G_{Ms} and G_{Mf} occupy the low rate region. GMs is clearly sensitive to xenon pressure. However, as it disappears, we do not observe the appearance of a new process as in Mb, but a replenishment of primary rebinding. Band GMf is hardly affected up to 16 bars, but was found to start decreasing at higher pressure (not shown).

We must conclude to the presence of two docking cavities: one with high affinity for Xe is responsible for the slow red band G_{Ms}; another one, filled only at high pressure, must be responsible for the fast red band, G_{Mf}. If we try to dispose these cavities graphically in a kinetic scheme, then only two situations are compatible with the data. The cavities may be either parallel and independent, or sequential only if the high affinity cavity is at the end of the sequence. But what do we now about cavities in $P450_{cam}$? The answer is: nothing. It remains to crystallography to find out where xenon is binding to cytochrome $P450_{cam}$. A search for empty space that could be filled with hydrogen atoms in the energy-minimized structure revealed the presence of three cavities close enough to the heme to interfere with CO rebinding, but the single one with a size sufficient to bind xenon is occupied by three water molecules (C1 cavity of Fig. 12). It is tempting to conclude that the cavities actually involved are not permanent ones in cytochrome $P450_{cam}$ or that their access is rate limited by small internal fluctuations. Computations are currently in progress in order to identify the strategic residue(s) implied and to determine ligand trajectories following photodissociation.

11. Migration and relaxation rates

According to the previous section, the final kinetic scheme for $P450_{cam}$ combines both relaxation and migration (Fig. 14). This scheme, comprising at least 5 states and 6 rate parameters, cannot be solved in general. It should be

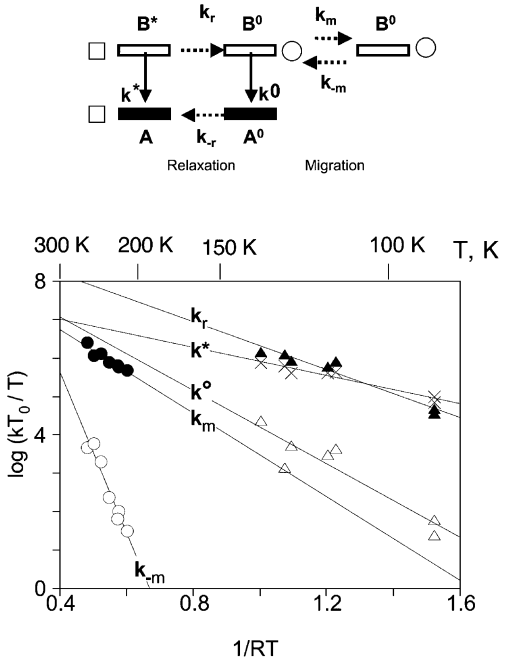


Fig. 14. Top : The global kinetic scheme for CO rebinding with $P450_{cam}$ comprises relaxation and migration from the relaxed state of the heme. Bottom : Arrhenius–Eyring plot of the rate parameters according to Eq. (9).

realized that adding one more state in any kinetic scheme results in at least two additional rate parameters. Very rapidly the number of unknowns exceeds the number of observables. The latter are limited to two per kinetic process, namely the peak rate and the amplitude or band area. The situation becomes even worse when bimolecular rebinding is observed in addition to geminate processes. It is generally impossible to know beforehand from which state escape has been taking place and the Mb case illustrates that escape may not be restricted to one particular ligand docking site. Yet, only one unique bimolecular rate is measurable experimentally, which again reduces the number of observables. Such a kinetic complexity can only be overcome by carefully scanning the temperature at about 10 K intervals in order to find a temperature range in which only a smaller number of processes are simultaneously present. If sufficiently sampled, the temperature dependence permits to compare rate parameters that were determined in different temperature ranges. To this purpose, Eq. (9) has come into current use:

$$k = A\frac{T}{T_0}\exp(-H/RT) \qquad (9)$$

in place of the conventional Arrhenius plot (Eq. (6)). The reference temperature T_0 is arbitrarily set to 100 K [53]. This form respects better Eyring's theory of reaction rates. The linearity of the plots is not affected in any significant way, but the values of A and H differ (though moderately) in Eqs. (6) and (9).

Further complexity is introduced by the statistical substates distributions. It could be shown, however, that the peak values of the bands in the rate spectra are mutually connected in the same way as non-distributed states would be [9] and that ordinary kinetic equations can be used. With these restrictions in mind, we may now discuss and compare the various rates that could be determined.

In cytochrome $P450_{cam}$, primary rebinding with the relaxed substrate is considerably slower than direct rebinding. The effect was found to be strongly substrate dependent (several logs), in particular the slope, i.e., the activation enthalpy is steeper (not shown). The relaxation rate is most of the time commensurable with direct rebinding, explaining that both rebinding paths can be seen simultaneously at these temperatures. But relaxation is faster than direct rebinding at higher temperature where rebinding occurs in the relaxed form only. The migration rate is also much lower than any other rates: migration starts in the relaxed form, which justifies the kinetic scheme above. In addition, migration and return paths are independent of the substrate, so that relaxation and migration appear to be independent processes.

Finally, Fig. 15 compares Mb and $P450_{cam}$. For the sake of the discussion, the dielectric relaxation rate of the solvent cage k_s is also shown [54]. This measures the rate at which the solvent rearranges and is directly related to the solvent's viscosity. Substrate relaxation in $P450_{cam}$ is the fastest rate and remains always much faster than solvent relaxation, with no discontinuity in the region where k_s is changing so steeply. It also has the smallest activation enthalpy. This is consistent with its nature as a purely internal process involving only small structural rearrangements.

Ligand migration in $P450_{cam}$ is not only faster than in Mb but also most of the time faster than k_s, with again no discontinuity in slope near T_g. This points to an easy access to the main xenon cavity requiring only small internal fluctuations.

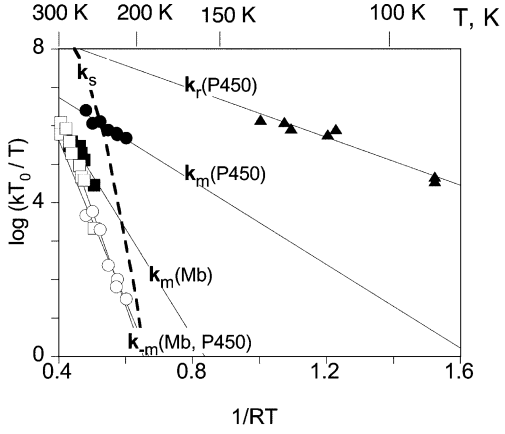


Fig. 15. Comparison of some relevant rate parameters of Cytochrome $P450_{cam}$ (from Fig. 14) with those of Myoglobin (filled and empty squares). Migration is slower in Mb than in $P450_{cam}$ though return rates (k_{-m}) are comparable in both proteins. On the average, migration is faster in $P450_{cam}$ and slower in Mb than the solvent cage relaxation rate, k_s. Arrhenius–Eyring plot of the rate parameters are according to Eq. (9).

In contrast, migration in Mb is slower than in $P450_{cam}$ and even smaller than k_s. It may be in part solvent dependent. The structure of Mb is more compact than that of $P450_{cam}$. Time-resolved crystallography of photodissociated sw MbCO [25] and of mutants [26,55] showed a motion of distal residue 64 toward the location formerly occupied by the CO ligand.

This motion of the distal histidine may be compared to that of the camphor substrate in cytochrome $P450_{cam}$ and proves that a genuine conformational relaxation occurs in myoglobin too. However, and contrary to the cytochrome $P450_{cam}$ case, the rebinding kinetics show no sign of this process between 77 K and 180 K, the temperature above which ligand migration starts [16].

In the time-resolved crystallography of the YQR-MbCO mutant [26], the authors reported that the motion of residue 64 was synchronous with the occupation of the Xe1 proximal site by CO that involves large-scale motions of the CD turn and of the E helix. Therefore, migration and relaxation in myoglobin appear to be both linked to the solvent-slaved motion of the E helix, a fact that may well account for the absence of relaxation as long as the solvent matrix remains rigid.

12. Summary of conclusions

Summarizing about 30 years of work by a large community of scientists is beyond the scope of this review. But it may be useful to quote a few simple but clear ideas that have emerged from these efforts.

Because they are polymers with a large number of degrees of freedom, proteins fluctuate permanently among a large number of energetically degenerate micro conformations. This view was first based on the evidence provided by the polychromatic kinetics of ligand rebinding in glassy solvents at cryogenic temperature. It has now become widely accepted since solution NMR spectroscopy on one hand and molecular dynamics calculations on the other hand are now able to give some images of these fluctuations (though in different time ranges).

Conformational fluctuations at the interface with the solvent are essential for accelerating recognition processes between proteins that would be otherwise unlikely, and therefore very slow. They are also generally required for permitting entry and escape of ligands by opening transient gates into the protein matrix. Surface fluctuations are sensitive to damping and therefore to the viscosity of the surrounding, a fact easily detected in kinetic experiments.

The protein matrix is usually dense, leaving little free space. Permanent internal cavities, in small number, may provide alternative and transient docking sites for ligands. But internal fluctuations are probably necessary to establish transient cavities and to open communications among permanent ones allowing ligand migration along specific pathways.

Finally, deeper into the protein core, ligand binding and release may triggers subtle conformational changes in the close vicinity of the active center. This is a genuine functional protein motion.

Surface fluctuations, opening of ligand migration pathways and conformational relaxation at the active site are general processes likely to occur in all proteins. All require the motion of some part(s) of the protein matrix and interfere with the simple ligand binding reaction. However, the works reported here show that the possibility to detect their kinetic signature depends on the nature and amplitude of the structural rearrangements involved as well as of their proximity to the solvent interface. There can be therefore no universal model for ligand binding.

It has been thought for a time that the complexity of protein behavior could be understood by considering proteins as “glass-like” objects. Clearly, there is no more need for such an analogy as far as protein reaction kinetics are concerned : their puzzling temperature dependence is explained in a straightforward way by ligand migration.

Finally, it turns out that complex protein kinetics (see for example Figs. 3 and 5) can be reduced to very elementary reaction schemes (Figs. 8 and 14) once the distribution of protein molecules among statistical substates is properly taken into account. Complexity was arising from the fact that the reaction partners are not individual molecules but rather statistical ensembles of molecules so that the underlying simplicity became blurred.

References

[1] D. Beece, L. Eisenstein, H. Frauenfelder, D. Good, M.C. Marden, L. Reinisch, A.H. Reynolds, L.B. Sorensen, K.T. Yue, Solvent viscosity and protein dynamics, Biochemistry 19 (1980) 5147–5157.
[2] S. Yedgar, C. Tetreau, B. Gavish, D. Lavalette, Viscosity dependence of O2 escape from respiratory proteins as a function of cosolvent molecular weight, Biophys. J. 68 (1995) 665–670.
[3] P. Douzou, Cryobiochemistry, Academic Press Inc., Ltd., London, 1977.
[4] R.H. Austin, K.W. Beeson, L. Eisenstein, H. Frauenfelder, I.C. Gunsalus, Dynamics of ligand binding to myoglobin, Biochemistry 14 (1975) 5355–5373.
[5] D. Lavalette, C. Tetreau, Viscosity-dependent energy barriers and equilibrium conformational fluctuations in oxygen recombination with Hemerythrin, Eur. J. Biochem. 177 (1988) 97–108.
[6] D. Ehrenstein, G.U. Nienhaus, Conformational substates in azurin, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 9681–9685.
[7] P.J. Steinbach, Two-dimensional distributions of activation enthalpy and entropy from kinetics by the maximum entropy method, Biophys. J. 70 (1996) 1521–1528.
[8] P.J. Steinbach, A. Ansari, J. Berendzen, D. Braunstein, K. Chu, B.R. Cowen, D. Ehrenstein, H. Frauenfelder, J.B. Johnson, D.C. Lamb, S. Luck, J.R. Mourant, G.U. Nienhaus, P. Ormos, R. Philipp, A. Xie, R.D. Young, Ligand binding to heme proteins: connection between dynamics and function, Biochemistry 30 (1991) 3988–4001.
[9] C. Tetreau, M. Tourbez, D. Lavalette, Conformational relaxation in hemoproteins: the cytochrome P450cam case, Biochemistry 39 (2000) 14219–14231.
[10] A.K. Livesey, M. Delaye, P. Licinio, J.-C. Brochon, Maximum

entropy analysis of dynamic parameters via the Laplace transform, Faraday Discuss. Chem. Soc. 83 (1987) 1–12.
[11] A.K. Livesey, J.-C. Brochon, Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method, Biophys. J. 52 (1987) 693–706.
[12] D. Lavalette, C. Tetreau, J.-C. Brochon, A. Livesey, Conformational fluctuations and protein reactivity, Eur. J. Biochem. 196 (1991) 591–598.
[13] P.J. Steinbach, K. Chu, H. Frauenfelder, J.B. Johnson, D.C. Lamb, G.U. Nienhaus, T.B. Sauke, R.D. Young, Determination of rate distributions from kinetic experiments, Biophys. J. 61 (1992) 235–245.
[14] D.G. Lambright, S. Balasubramanian, S.G. Boxer, Dynamics of protein relaxation in site-specific mutants of human myoglobin, Biochemistry 32 (1993) 10116–10124.
[15] J.B. Johnson, D.C. Lamb, H. Frauenfelder, J.D. Müller, B. McMahon, G.U. Nienhaus, Ligand binding to heme proteins: VI. Interconversion of taxonomic substates in carbonmonoxymyoglobin, Biophys. J. 71 (1996) 1563–1573.
[16] C. Tetreau, Y. Blouquit, E. Novikov, E. Quiniou, D. Lavalette, Competition with xenon elicits ligand migration and escape pathways in myoglobin, Biophys. J. 86 (2004) 435–447.
[17] C. Tetreau, C. Di Primo, R. Lange, H. Tourbez, D. Lavalette, Dynamics of carbon monoxide binding with cytochromes P-450, Biochemistry 36 (1997) 10262–10275.
[18] C. Tetreau, L. Mouawad, S. Murail, P. Duchambon, Y. Blouquit, D. Lavalette, Disentangling ligand migration and heme pocket relaxation in cytochrome P450cam, Biophys. J. 86 (2004) 1250–1263.
[19] E.E. Scott, Q.H. Gibson, Ligand migration in sperm whale myoglobin, Biochemistry 36 (1997) 11909–11917.
[20] E.E. Scott, Q.H. Gibson, J.S. Olson, Mapping the pathways for O2 entry into and exit from myoglobin, J. Biol. Chem. 276 (2001) 5177–5188.
[21] I. Schlichting, J. Berendzen, G.N. Phillips Jr., R.M. Sweet, Crystal structure of photolysed carbonmonoxy-myoglobin, Nature 371 (1994) 808–812.
[22] H. Hartmann, S. Zinser, P. Komninos, R.T. Schneider, G.U. Nienhaus, F. Parak, X-ray structure determination of a metastable state of carbonmonoxy myoglobin after photodissociation, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 7013–7016.
[23] K. Chu, J. Vojtechovsky, B.H. McMahon, R.M. Sweet, J. Berendzen, I. Schlichting, Structure of a ligand-binding intermediate in wild-type carbonmonoxymyoglobin, Nature 403 (2000) 921–923.
[24] V. Srajer, T.-Y. Teng, T. Ursby, C. Praverdand, Z. Ren, S.-I. Adachi, W. Schildkamp, D. Bourgeois, M. Wulff, K. Moffat, Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography, Science 274 (1996) 1726–1729.
[25] V. Srajer, Z. Ren, T.-Y. Teng, M. Schmidt, T. Ursby, D. Bourgeois, C. Pradervand, W. Schildkamp, M. Wulff, K. Moffat, Protein conformational relaxation and ligand migration in myoglobin: a nanosecond to millisecond movie from time-resolved Laue X-ray diffraction, Biochemistry 40 (2001) 13802–13815.
[26] D. Bourgeois, B. Vallone, F. Schotte, A. Arcovito, A.E. Miele, G. Sciara, M. Wulff, P. Anfinrud, M. Brunori, Complex lanscape of protein structural dynamics unveiled by nanosecond Laue Crystallography, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 8704–8709.
[27] F. Schotte, M. Lim, T.A. Jackson, A.V. Smirnov, J. Soman, J.S. Olson, G.N. Phillips Jr., M. Wulff, P.A. Anfinrud, Watching a protein as it functions with 150-ps time-resolved X-ray crystallography, Science 300 (2003) 1944–1947.
[28] J.O. Alben, D. Beece, S.F. Bowne, W. Doster, L. Eiseinstein, H. Frauenfelder, D. Good, J.D. McDonald, M.C. Marden, P.P. Moh, L. Reinisch, A. Reynolds, E. Shyamsunder, K.T. Yue, Infrared spectroscopy of photodissociated carboxymyoglobin at low temperatures, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 3744–3748.
[29] J. Vojtechovsky, K. Chu, J. Berendzen, R.M. Sweet, I. Schlichting, Crystal structures of myoglobin–ligand complexes at near-atomic resolution, Biophys. J. 77 (1999) 2153–2174.
[30] F. Yang, J.G.N. Phillips, Crystal structures of CO-, deoxy- and met-myoglobins at various pH values, J. Mol. Biol. 256 (1996) 762–774.
[31] A. Ansari, J. Berendzen, D. Braunstein, B.R. Cowen, H. Frauenfelder, M.K. Hong, I.E.T. Iben, J.B. Johnson, P. Ormos, T.B. Sauke, R. Scholl, A. Schulte, P.J. Steinbach, J. Vittitow, R.D. Young, Rebinding and relaxation in the myoglobin pocket, Biophys. Chem. 26 (1987) 337–355.
[32] W.T. Potter, M.P. Tucker, R.A. Houtchens, W.S. Caughey, Oxygen infrared spectra of oxyhemoglobins and oxymyoglobins. Evidence of two major liganded O2 structures, Biochemistry 26 (1987) 4699–4707.
[33] C. Tetreau, E. Novikov, M. Tourbez, D. Lavalette, Kinetic evidence for three photolyzable taxonomic conformational substates in oxy-myoglobin, Biophys. J. 82 (2002) 2148–2155.
[34] M.K. Hong, D. Braunstein, B.R. Cowen, H. Frauenfelder, I.E. Iben, J.R. Mourant, P. Ormos, R. Scholl, A. Schulte, P.J. Steinbach, A.-H. Xie, R.D. Young, Conformational substates and motions in myoglobin. External influences on structure and dynamics, Biophys. J. 58 (1990) 429–436.
[35] C. Jung, F. Marlow, Dynamic behavior of the active site structure in bacterial cytochrome P-450, Stud. Biophys. 120 (1987) 241–251.
[36] C. Jung, O. Ristau, H. Schulze, S.G. Sligar, The CO stretching mode infrared spectrum of substrate-free cytochrome P-450cam-CO. The effect of solvent conditions, temperature and pressure, Eur. Biochem. J. 235 (1996) 660–669.
[37] C. Tetreau, M. Tourbez, A. Gorren, B. Mayer, D. Lavalette, Dynamics of carbon monoxide binding with neuronal nitric oxide synthase, Biochemistry 38 (1999) 7210–7218.
[38] T.L. Poulos, B.C. Finzel, A.J. Howard, High-resolution crystal structure of cytochrome P450cam, J. Mol. Biol. 195 (1987) 687–700.
[39] R. Raag, T.L. Poulos, Crystal structure of the carbon-monoxide-substrate-cytochrome P-450cam ternary complex, Biochemistry 28 (1989) 7586–7592.
[40] A. Ostermann, R. Waschipky, F.G. Parak, G.U. Nienhaus, Ligand binding and conformational motions in myoglobin, Nature 404 (2000) 205–208.
[41] R.F. Tilton, I.D. Kuntz, G.A. Petsko, Cavities in proteins: structure of a metmyoglobin–xenon complex solved to 1.9 Å, Biochemistry 23 (1984) 2849–2857.
[42] M. Brunori, F. Cutruzzola, C. Savino, C. Travaglini-Allocatelli, B. Vallone, Q.H. Gibson, Structural dynamics of ligand diffusion in the protein matrix: a study on a new myoglobin mutant Y(B10) Q(E7) R(E10), Biophys. J. 76 (1999) 1259–1269.
[43] F. Draghi, A.E. Miele, C. Travaglini-Allocatelli, B. Vallone, M. Brunori, Q.H. Gibson, J.S. Olson, Controlling ligand binding in myoglobin by mutagenesis, J. Biol. Chem. 277 (2002) 7509–7519.
[44] H. Ishikawa, T. Uchida, S. Takahashi, K. Ishimori, I. Morishima, Ligand migration in human myoglobin: steric effects of isoleucine 107(G8) on O2 and CO binding, Biophys. J. 80 (2001) 1507–1517.
[45] M.L. Quillin, T. Li, J.S. Olson, G.N. Phillips Jr., Y. Dou, M. Ikeda-Saïto, R.M. Regan, M. Carlson, Q.H. Gibson, H. Li, R. Elber, Structural and functional effects of apolar mutations of the distal valine in myoglobin, J. Mol. Biol. 245 (1995) 416–436.
[46] M.L. Carlson, R. Regan, R. Elber, H. Li, G.N. Phillips, J.S. Olson, Q.H. Gibson, Nitric oxide recombination to double mutants of myoglobin: role of ligand diffusion in a fluctuating heme pocket, Biochemistry 33 (1994) 10597–10606.
[47] R. Elber, M. Karplus, Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin, J. Am. Chem. Soc. 112 (1990) 9161–9175.
[48] Q.H. Gibson, R. Regan, R. Elber, J.S. Olson, T.E. Carver, Distal pocket residues affect picosecond ligand recombination in myoglobin, J. Biol. Chem. 267 (1992) 22022–22034.
[49] H. Li, R. Elber, J.E. Straub, Molecular dynamics simulation of NO

recombination to myoglobin mutants, J. Biol. Chem. 268 (1993) 17908–17916.
[50] M.L. Carlson, R.M. Regan, Q.H. Gibson, Distal cavity fluctuations in myoglobin: protein motion and ligand diffusion, Biochemistry 35 (1996) 1125–1136.
[51] K. Nienhaus, P. Deng, J.M. Kriegl, G.U. Nienhaus, Structural dynamics of myoglobin: spectroscopic and structural characterization of ligand docking sites in myoglobin mutant L29W, Biochemistry 42 (2003) 9633–9646.
[52] K. Nienhaus, P. Deng, J.M. Kriegl, G.U. Nienhaus, Structural dynamics of myoglobin: effect of internal cavities on ligand migration and binding, Biochemistry 42 (2003) 9647–9658.
[53] D.D. Dlott, H. Frauenfelder, P. Langer, H. Roder, E.E. DiIorio, Nanosecond flash photolysis study of carbon monoxide binding to the b chain of hemoglobin Zürich [b63(E7)His→Arg], Proc. Natl. Acad. Sci. U. S. A. 80 (1983) 6239–6243.
[54] T. Kleinert, W. Doster, H. Leyser, W. Petry, V. Schwarz, M. Settles, Solvent composition and viscosity effects on the kinetics of CO binding to horse myoglobin, Biochemistry 37 (1998) 717–733.
[55] G. Hummer, F. Schotte, P.A. Anfinrud, Unveiling functional protein motions with picosecond X-ray crystallography and molecular dynamics simulations, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 15330–15334.