Day 2: Applications of Fluorescence Spectroscopy II

5. Data manipulation and data analysis

- Spectral manipulation
- Least-squares analysis
- Other approaches: Maximum entropy
- Global analysis

Fluorescence intensity depends on both excitation and emission

Generally, spectrofluorimeters scan the emission or the excitation wavelength, one a a time

For a sample containing one fluorescent species, the emission spectrum is (most of the times) independent on the excitation wavelength and the excitation spectrum is independent on the emission wavelength

The intensity at each combination of excitation-emission (2D spectrum) is proportional to the product of the corresponding 1-D spectra

$I(\lambda_{ex}, \lambda_{em})=I(\lambda_{ex})*I(\lambda_{em})$

Such 2-D spectra can be collected as a combination of emission spectra at different excitations or as excitation spectra at different emissions.

If the sample contains several fluorescent compounds, the 2-D spectrum contains the contribution of the sum of the independent species weighted by their concentration

$$I(\lambda_{ex}, \lambda_{em})=\Sigma c_i I_i(\lambda_{ex},\lambda_{em})= \Sigma c_i I_i(\lambda_{ex})*I(\lambda_{em})$$

However, the information on a single 2-D spectrum is not sufficient to uniquely determine the responses and concentrations of the independent species due to rotational ambiguity in the mathematical treatment

If a set of 2-D spectra are available, then the ambiguity can be removed. The method is called the Procrustes rotation and was originally developed for 2 data sets.

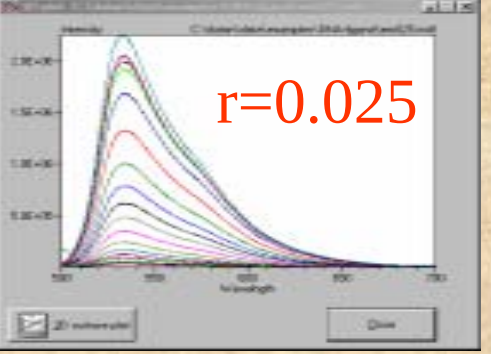
The following figures are from the commercial software DATAN, by M. Kubista

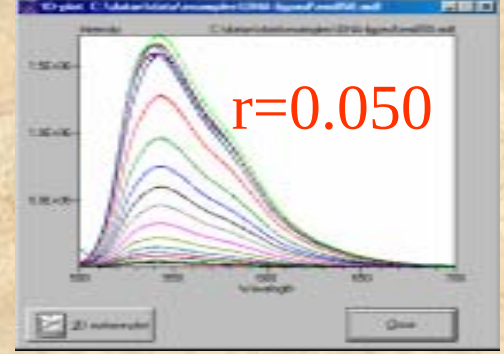
POSSIBLE VARIABLES

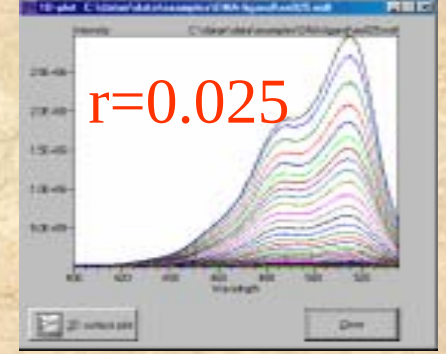
- Excitation wavelength
- Emission wavelength
- Sample composition
- Decay time
- Reaction time
- Polarization
- Coordinates

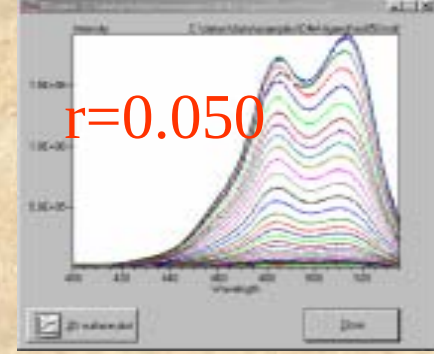
POSSIBLE APPLICATIONS

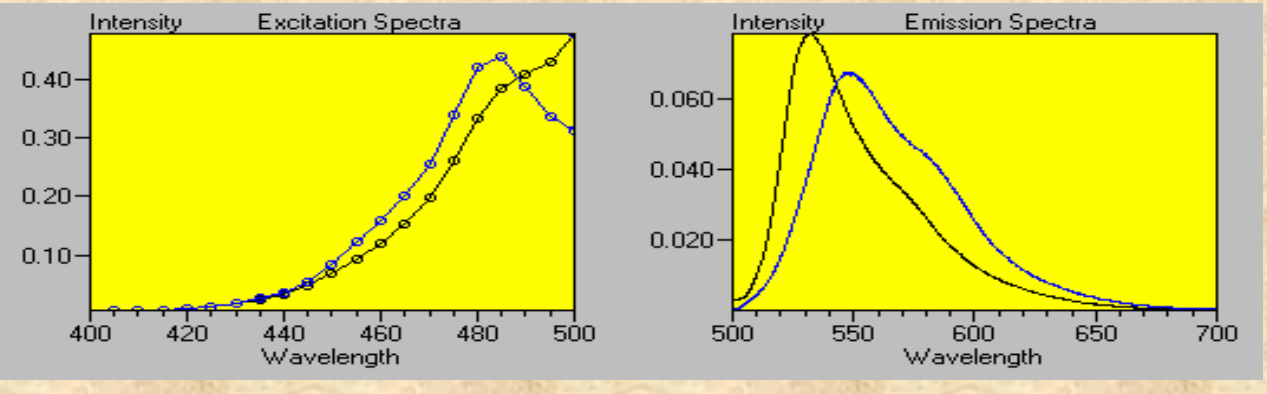
- Characterizing test samples
- Studying chemical equilibria
- Measuring reactions rates
- Studying photochemical reactions
- Microscopy

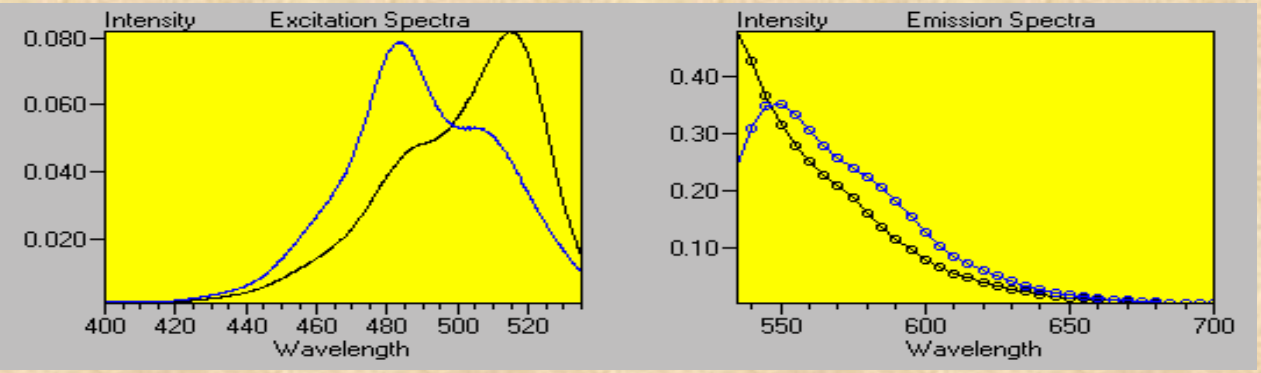




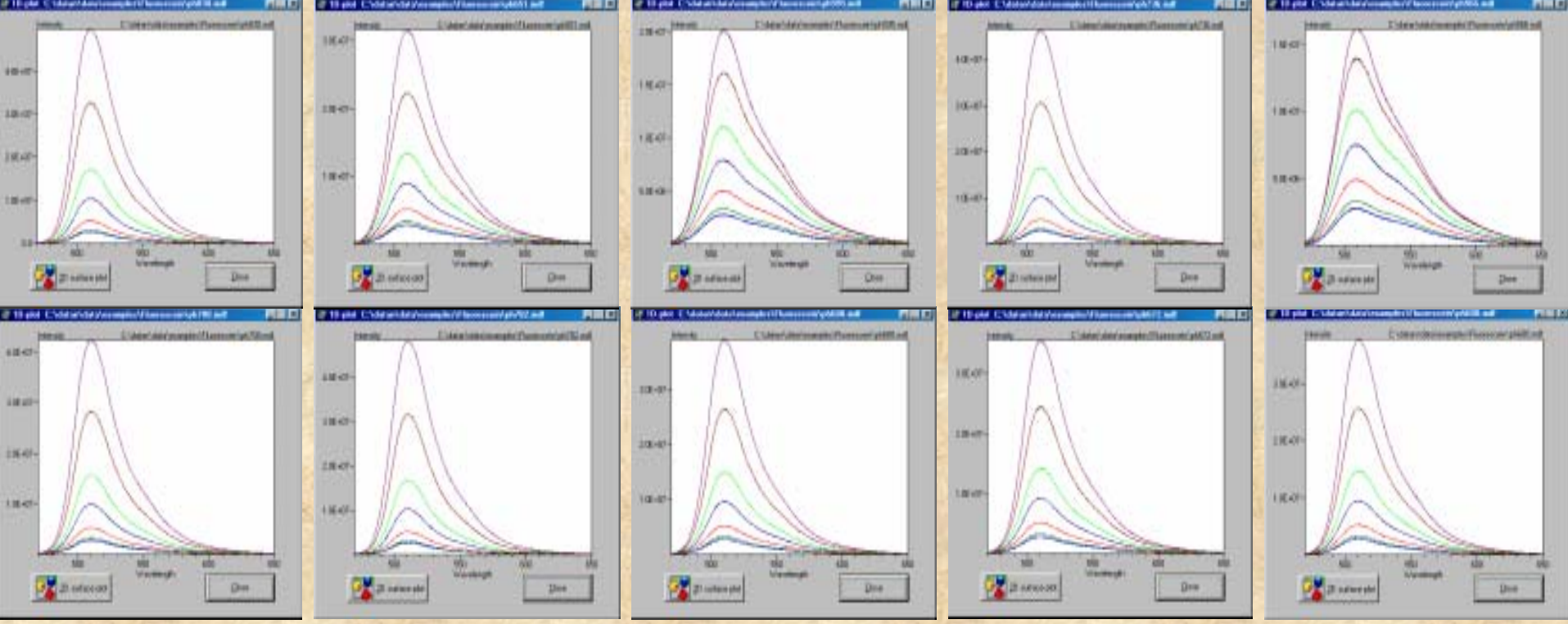


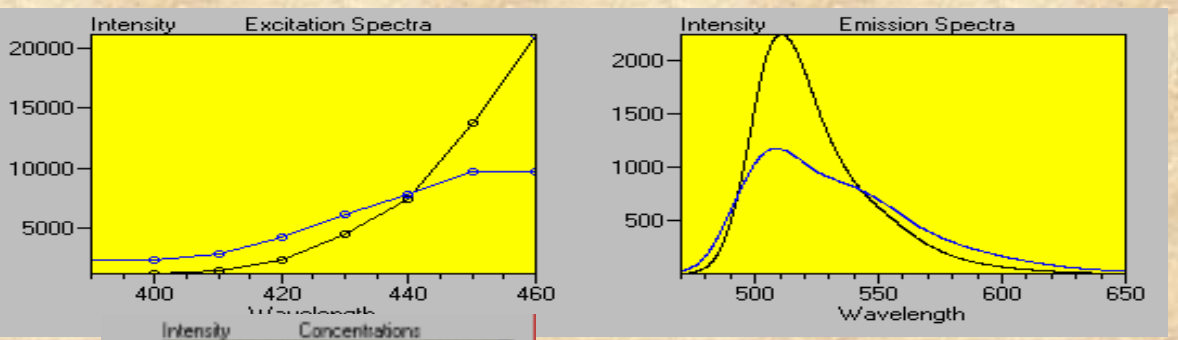


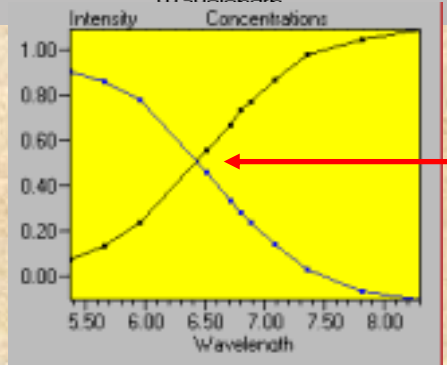




Fluorescence spectra of thiazole orange bound to DNA at binding ratio of 0.025 and 0.05 dyes per bp. Left figures show analysis with the emission wavelength scanned and right figures with the excitation wavelength scanned. Calculated spectra from the two independent characterizations clearly agree evidencing the presence of bound monomer and bound dimer.





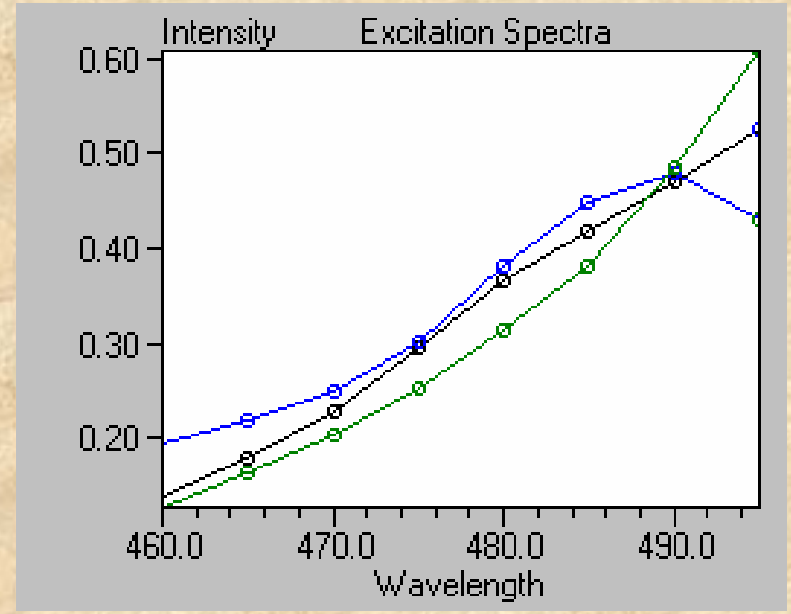


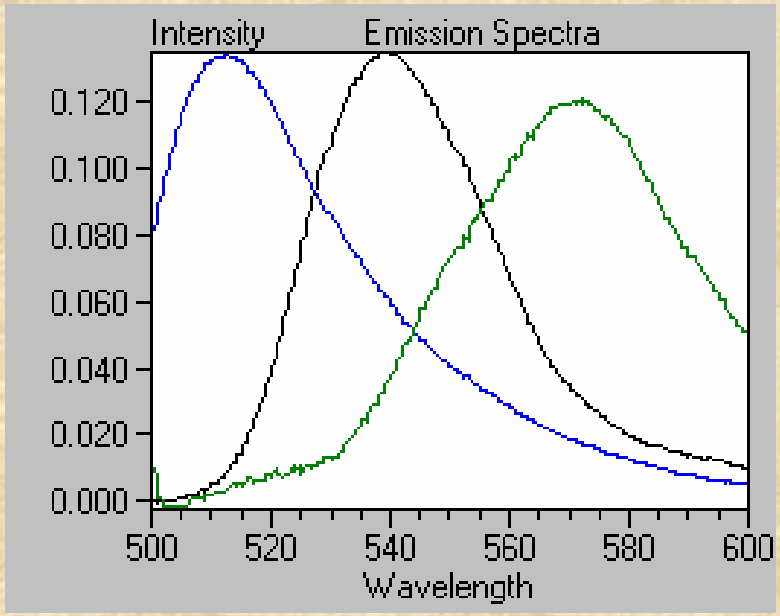
pKa = 6.41

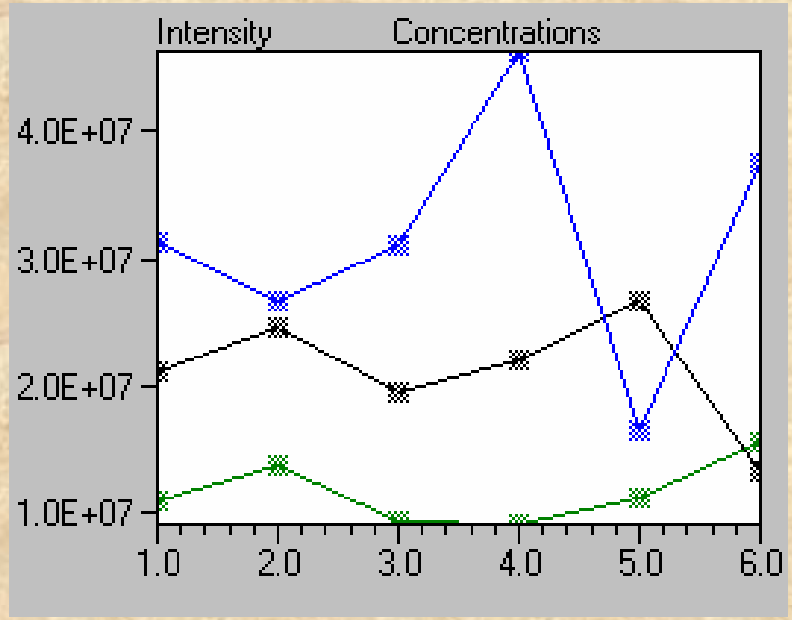
pH titration

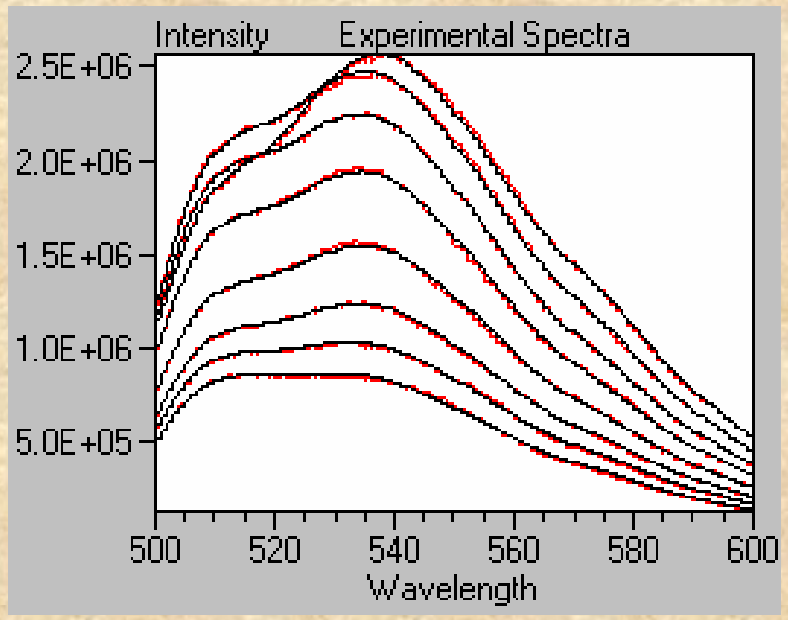
2-dimensional fluorescence spectra of fluorescein measured at different pH. Calculated spectra reveal presence of the fluorescein mono- and dianion, and from the calculated concentrations pKa is determined.

Decomposition of a mixture of 3 components

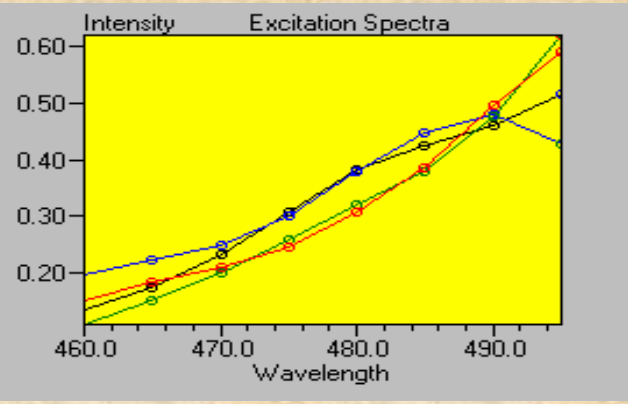


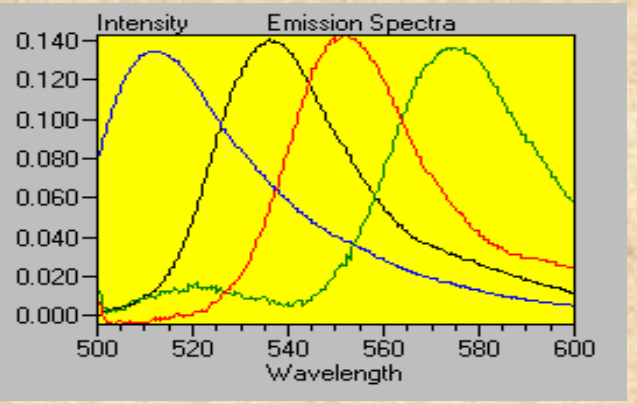


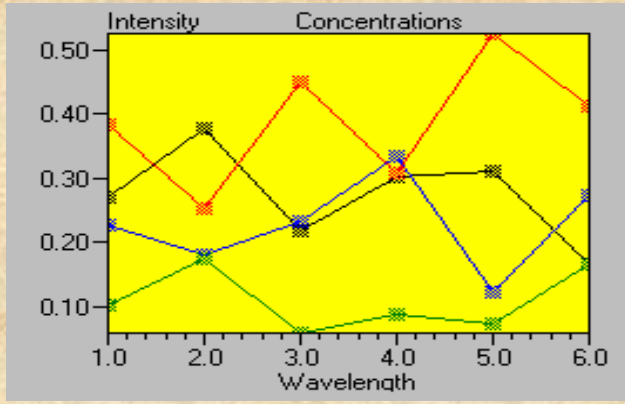


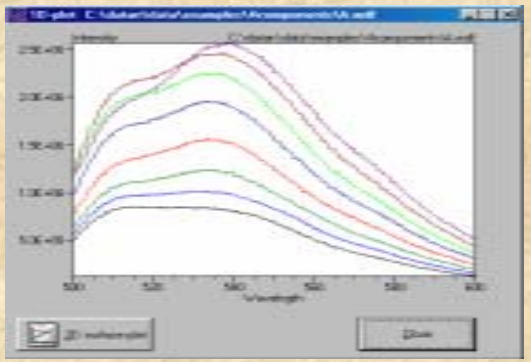


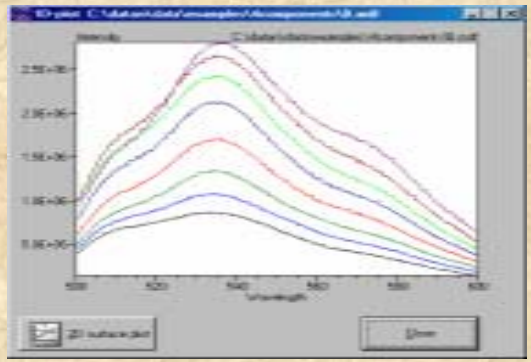
4-component samples

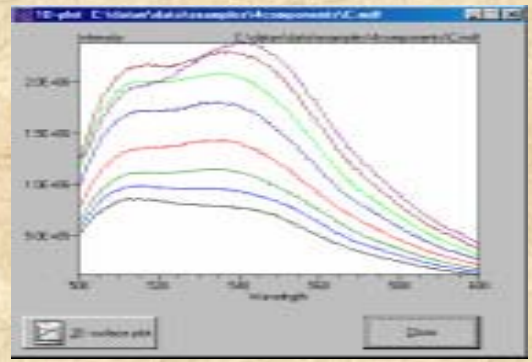


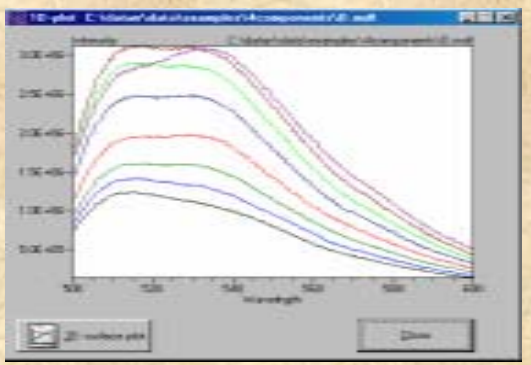


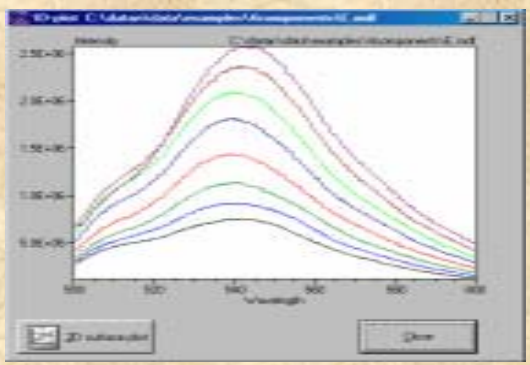


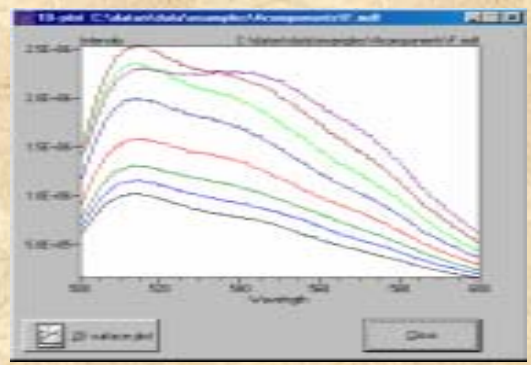






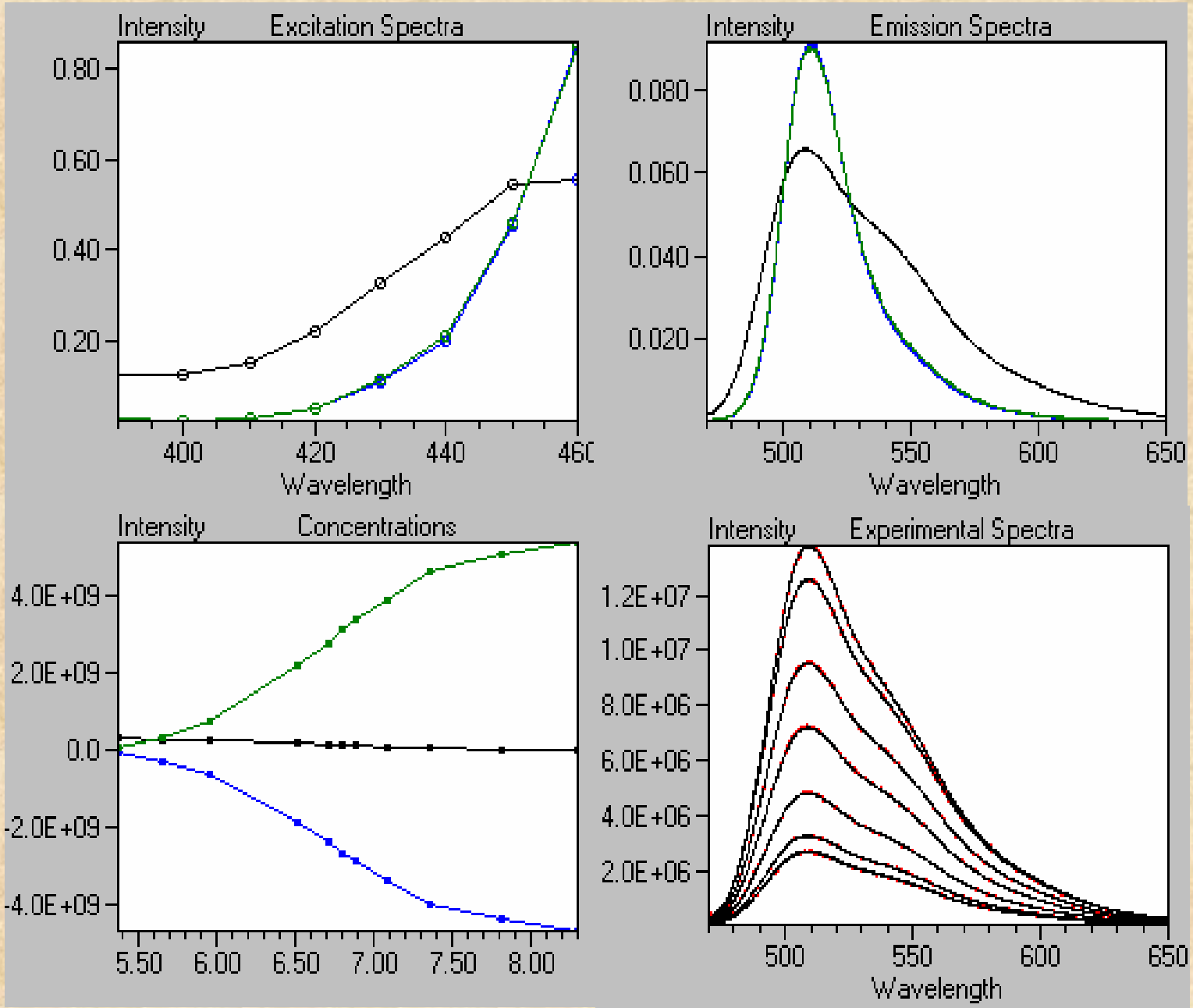






Test samples containing mixtures of fluorescein, Eosin Y, Rhodamine 6G and Rhodamine B. By trilinear decomposition excitation and emission spectra, as well as component concentrations are determined.

Procrustes rotation decomposition of fluorescein data using 3 components



Principal component analysis (PCA) is the traditional method to analyze matrix data (equivalent to singular value decomposition). The matrix is decompose into the product of a target and a projection matrix with orthogonal columns and rows

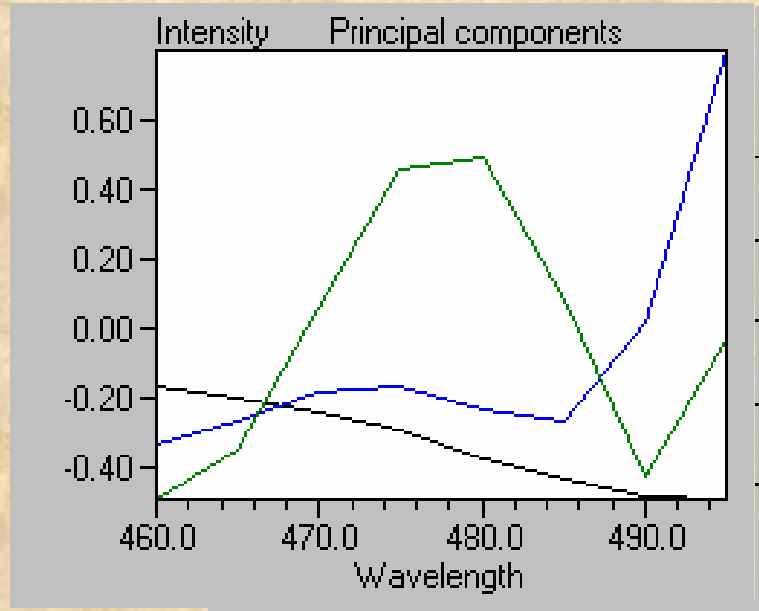
$$A = \sum_{i=1}^{q} t_i p_i$$

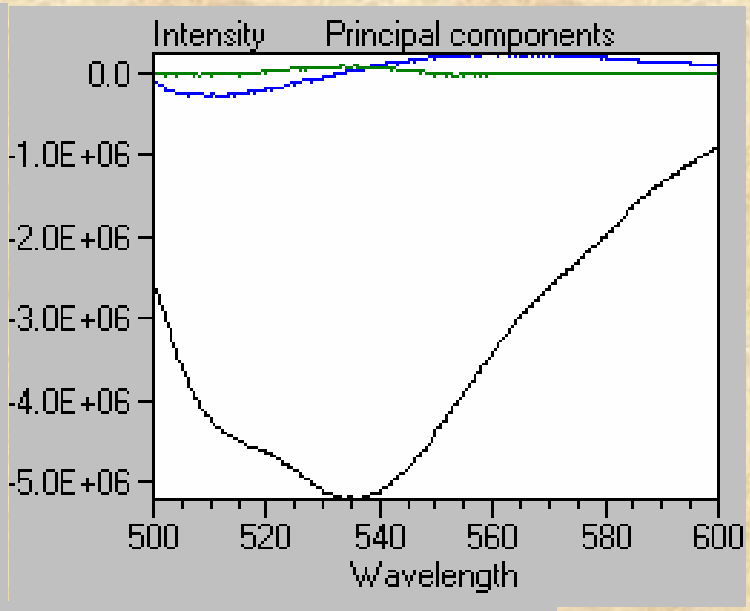
Where A is the data matrix and t_i and p_i are called the target and the projection vectors. These are mathematically defined and do not necessarily represent the response of specific chemical components.

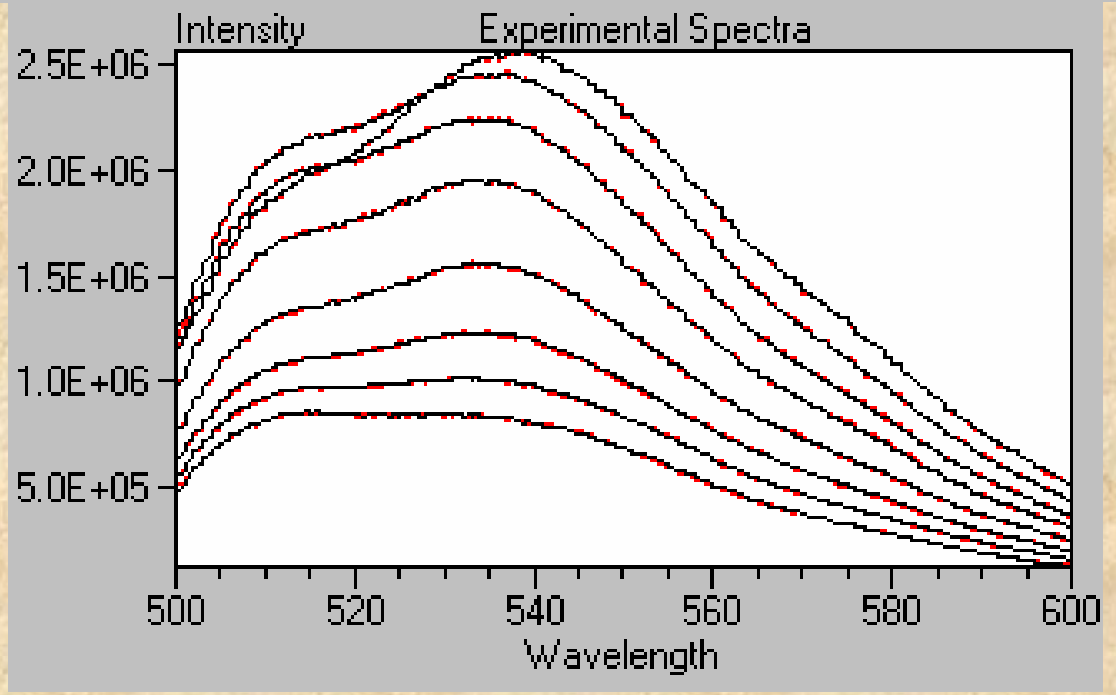
Q is the number of projection vectors used to represent the data. The large the value of q the better the agreement with the data. If q is at least equal to the smallest of the dimensions of the data matrix, the agreement is perfect, including the noise.

However, this is not interesting. What is interesting is to find a smaller number of components that still agrees with the data. Generally these components are identified with different species of the mixture.

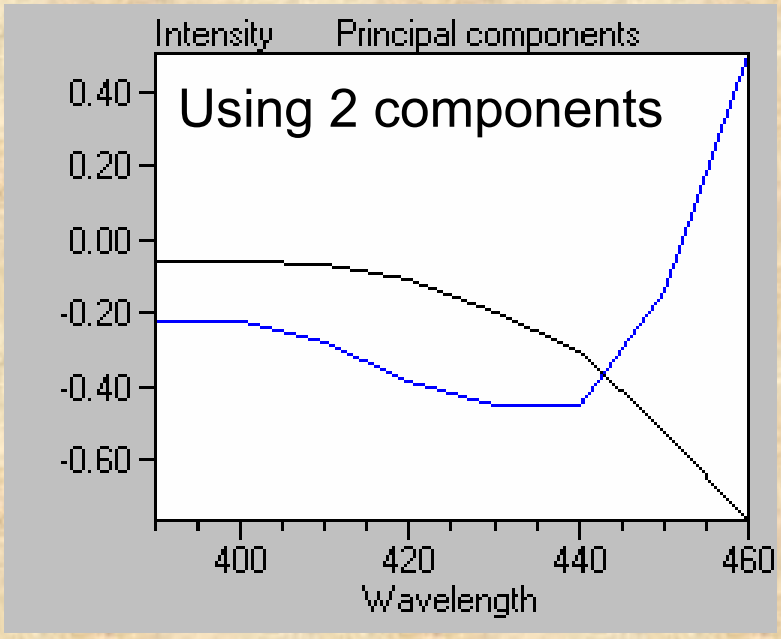
Principal component analysis of the 3-dye mixture

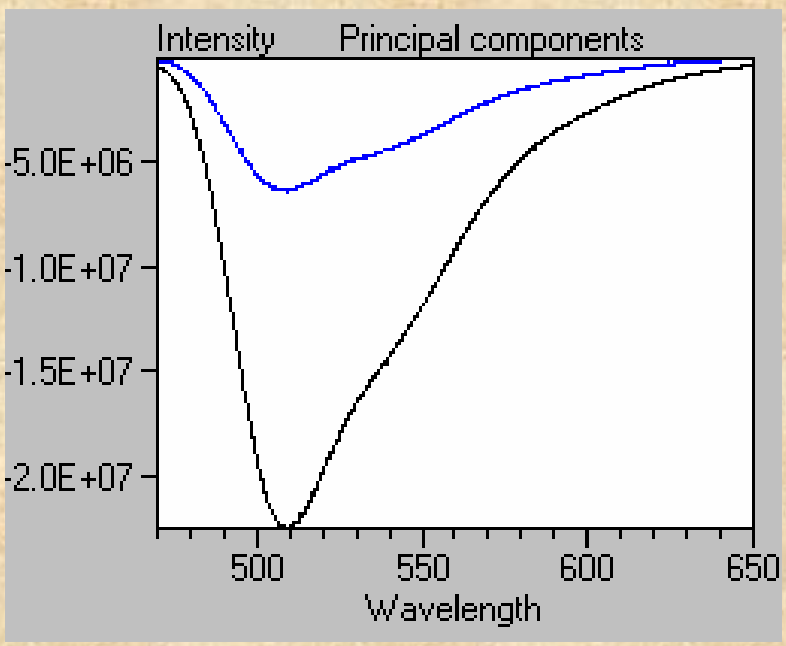


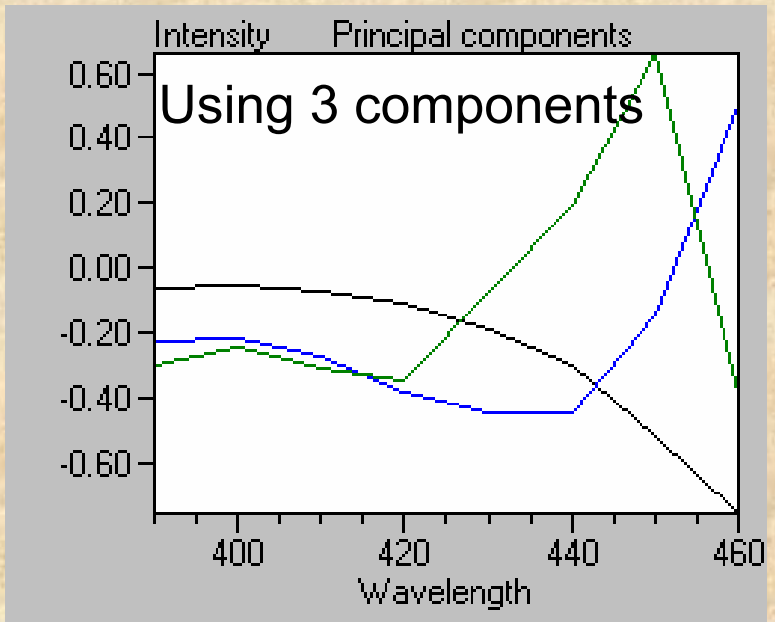


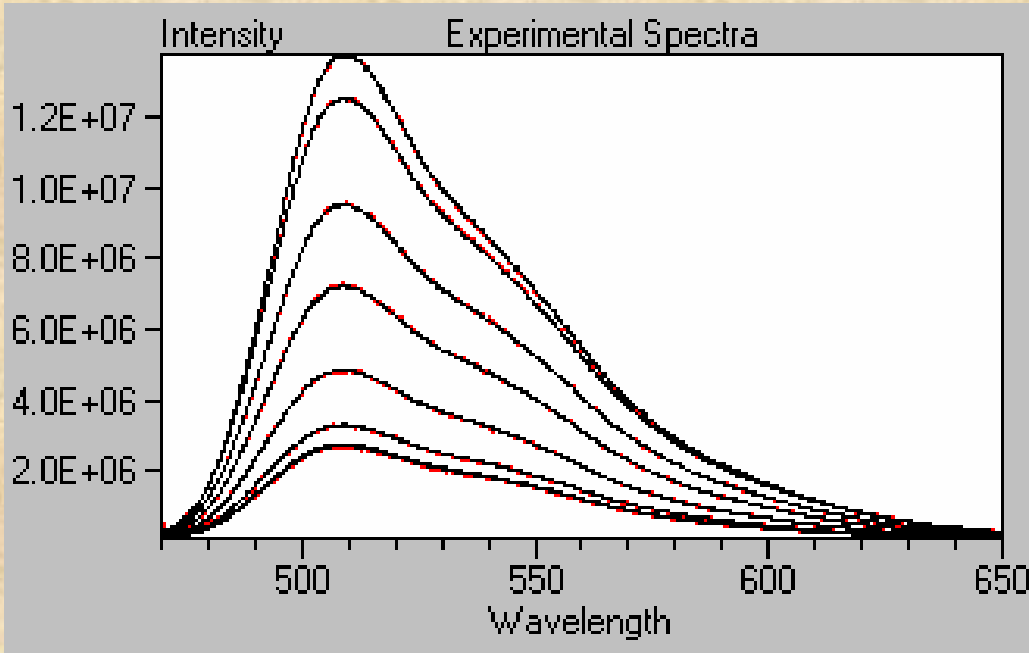


Principal component analysis of fluorescein titration experiments









Least-squares analysis

- Resolution of multi-exponential component decays
- Discrete exponentials
- Lifetime distributions
- Global analysis methods

Purpose of least-squares analysis

When is it applicable?
What does it provides?
What is the rigorous error analysis?

For the frequency-domain data, G. Weber proposed an exact solution (J. Phys. Chem. **83**, 1333 (1979)). However, the exact solution requires "infinite" precision of the data. It works well for 2-3 frequencies. Still used for fast evaluation and for microscopy where lifetime is measured in a large number of pixels.

When more data points are available and for time-domain data the least-squares method provides a statistical basis for the evaluation of the parameters

Definitions:

$$X^2 = (1-w)\sum_r [(\phi_r^c - \phi_r^m)/\sigma_r^\varphi)]^2 + w\sum_r [(M_r^c - M_r^m)/\sigma_r^M)]^2$$

$$X^2_{\text{Reduced}} = X^2/\nu$$

w is the relative weight of phase and modulation
σ represents the standard deviation of the phase and modulation
v is the number of degree of freedom

It is important to have the correct evaluation of the uncertainties of the data points

2-3 exponential analysis is done in less than one second in modern computers

Understanding the correlation between parameters!

The chi-square value must be close to 1 for a good fit

In the time-domain data fit analysis, the chi-square is typically close to 1 for a good fit

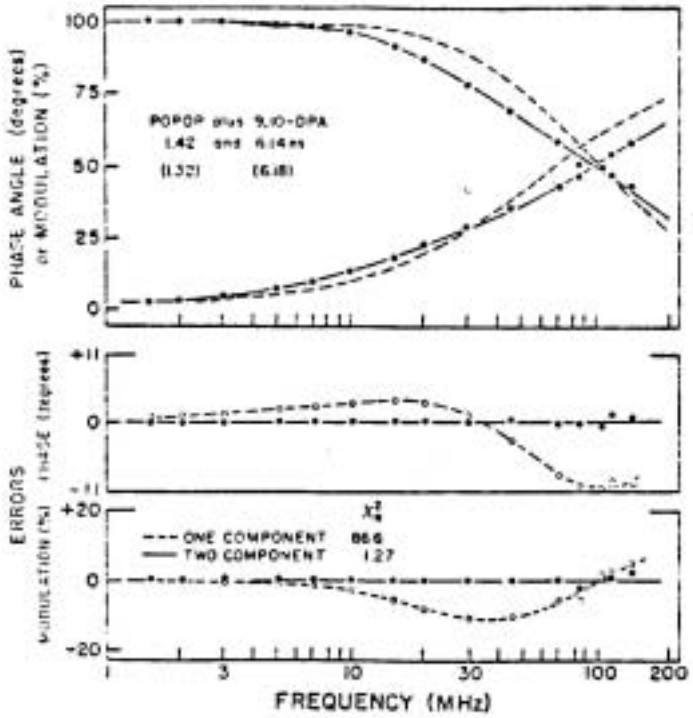
In the frequency-domain, frequently the chi-square value is below 1. This value simply implies that the evaluation of the errors is done incorrectly.

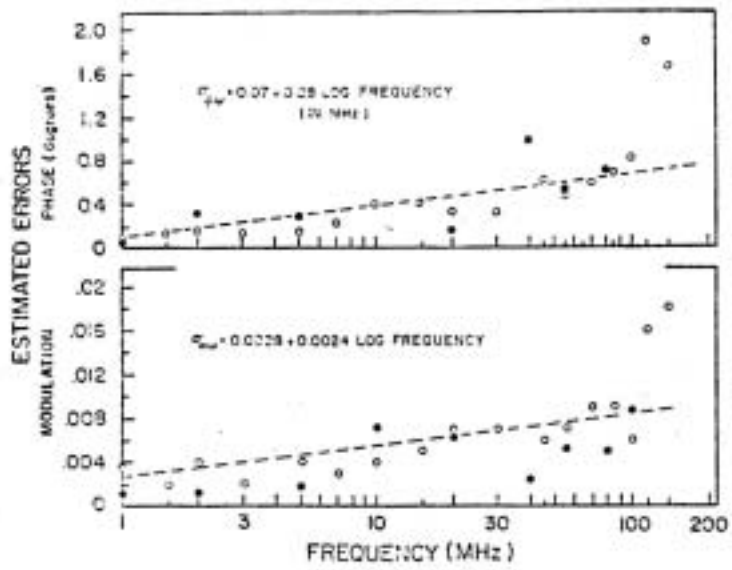
In the frequency-domain technique, it is assumed that the phase uncertainty is about 0.2 degrees and the modulation uncertainty is 0.004. These uncertainties correspond to a relative error on the order of 0.2-0.4%.

In many modern instruments and for bright samples, the errors could be below these values, and the chi-square will be below 1.

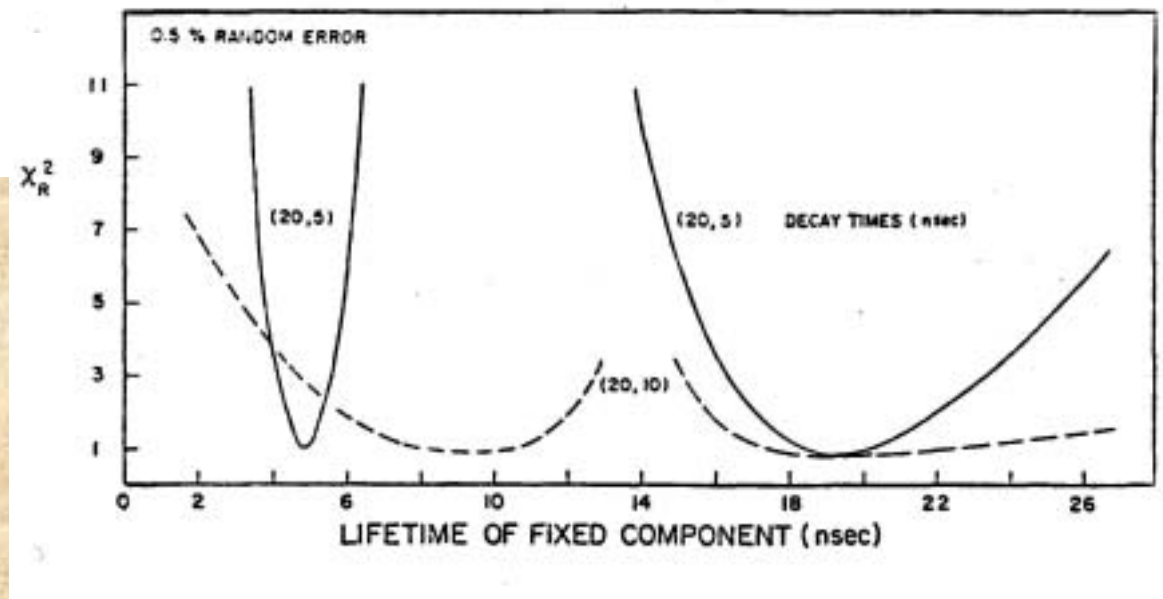
In the time domain, the uncertainties of each data point are calculated assuming a Poisson statistics of the counts in each data bin.

Errors at each point must be evaluated

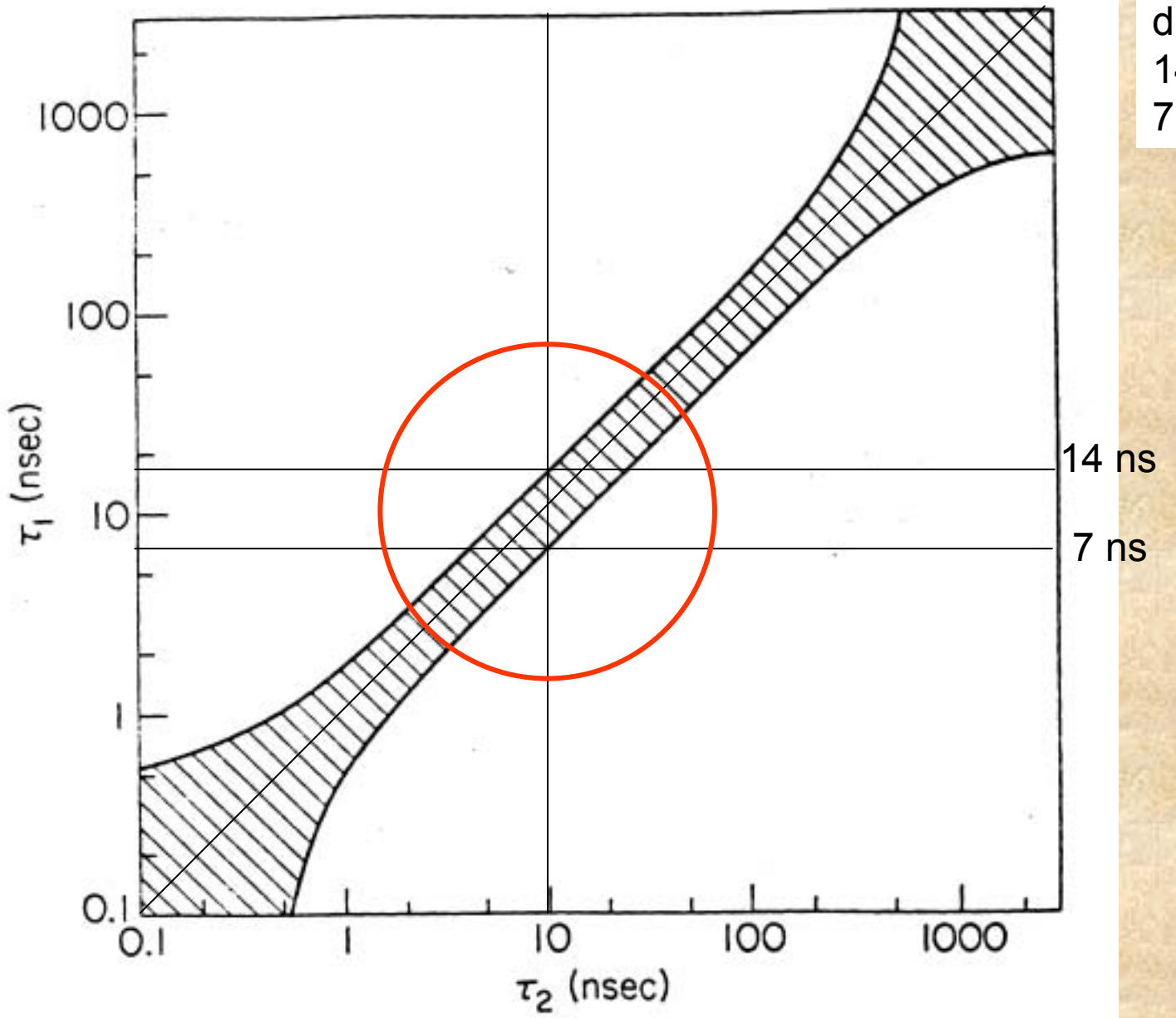




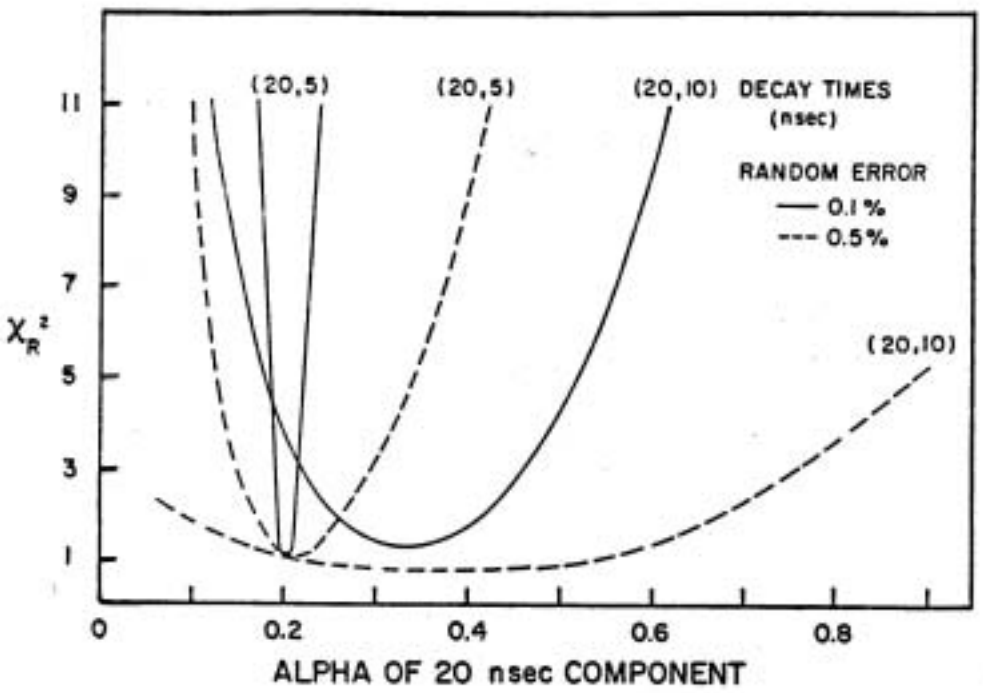
Analysis of the chi-square surface

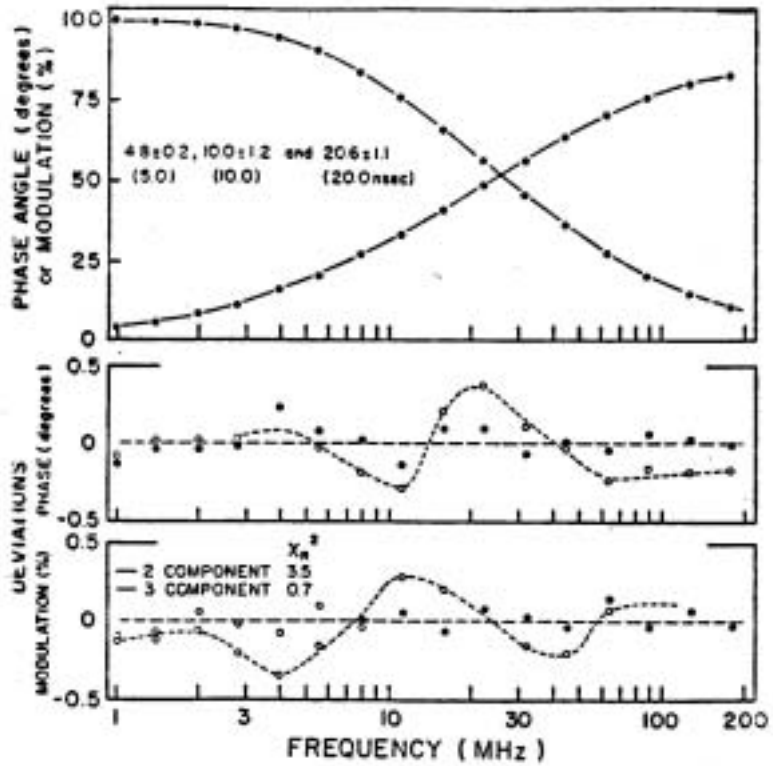


2-components lifetime resolvability plot



10 ns lifetime can be distinguished from 14ns and longer or 7ns and shorter


(20,5)
(20,5)
(20,10)
DECAY TIMES
(nsec)
RANDOM ERROR
0.1 %
0.5%
(20,10)
χ_R^2
11
9
7
5
3
1
0
0.2
0.4
0.6
0.8
ALPHA OF 20 nsec COMPONENT


PHASE ANGLE (degrees)
or MODULATION (%)
100
75
50
25
0
4.8±0.2, 10.0±1.2 and 20.6±1.1
(5.0) (10.0) (20.0nsec)
0.5
0
-0.5
χ_R^2
2 COMPONENT 3.5
3 COMPONENT 0.7
1
3
10
30
100
200
FREQUENCY (MHz)

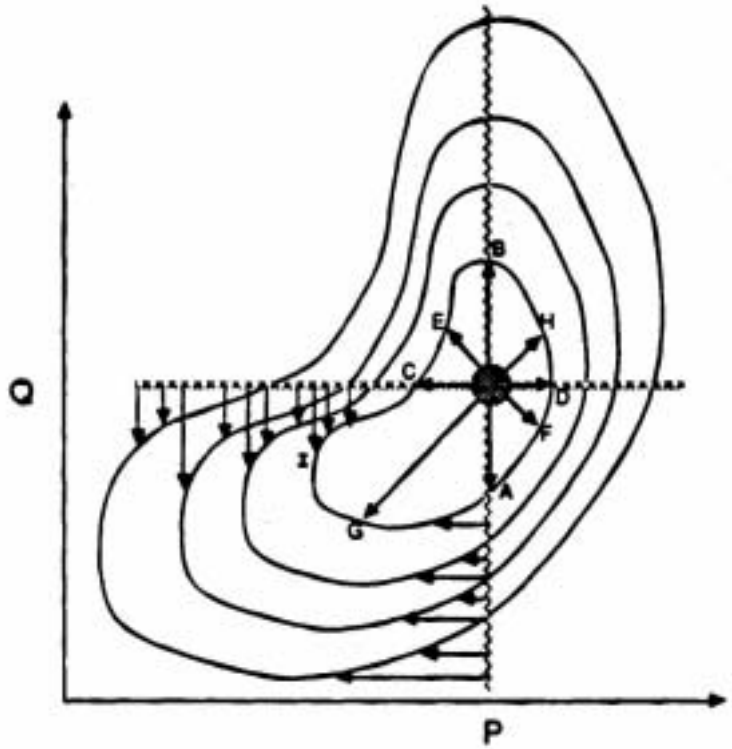


Figure 10: CHISQUARE CONTOUR PLOT OF THE TWO FITTING PARAMETERS "P" AND "Q". THE SHADED REGION SCHEMATICALLY DEPICTS THE ZERO'TH ORDER ERROR ANALYSIS USING THE *linear approximation*. THE LINES "A-B" AND "C-D" REPRESENT THE REGION OF THE ERROR SURFACE EXAMINED BY *uni-dimensional* SEARCH ALONG EACH PARTICULAR PARAMETER AXIS, HOLDING ALL OTHER FITTING PARAMETERS FIXED. THE LINES "E-F" AND "G-H" REPRESENT THE REGION OF THE ERROR SURFACE EXAMINED BY THE *directed-search* ALONG THE EIGENVECTORS OF THE CURVATURE MATRIX AT THE MINIMUM CHISQUARE. THE *exhaustive search* ALGORITHM, PERFORMS A WHOLE SERIES OF MINIMIZATIONS ALONG EACH PARAMETER AXIS (DENOTED AS DIRECTED ARROWS ORIGINATING FROM THE PARAMETER AXIS TO THE LOWEST CHISQUARE CONTOUR AVAILABLE) ALLOWING THE ALL OTHER FITTING PARAMETERS TO COMPENSATE.

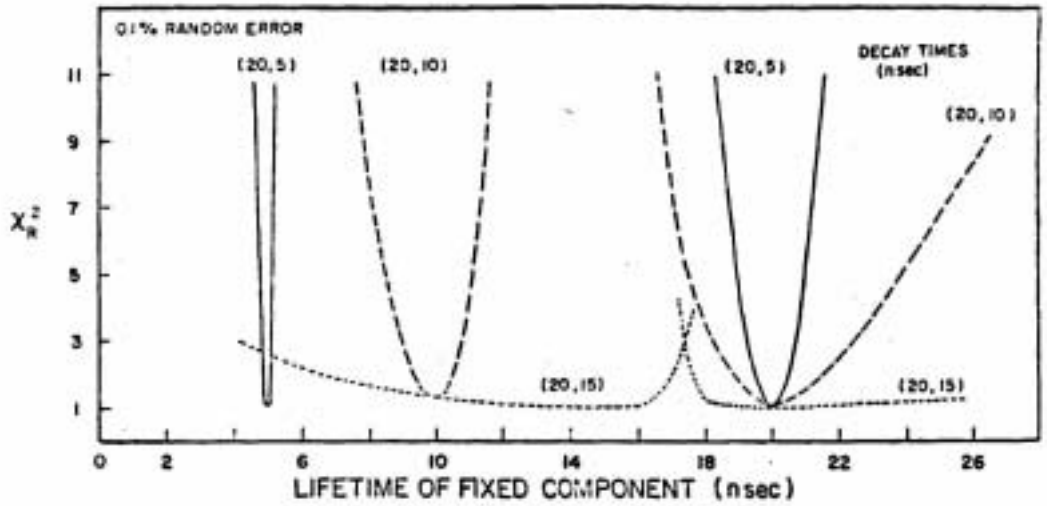
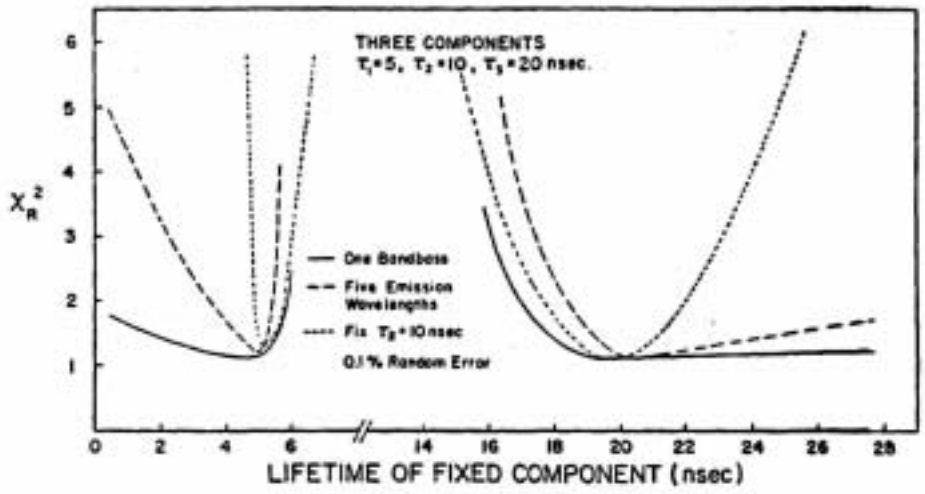


Figure 6. Sensitivity of χ^2_R to the Values of the Decay times for a Two-Component Decay and Data with 0.1% Random Error.

The values of χ^2_R were obtained by keeping one decay time fixed at the value indicated on the x-axis. Then, the other parameters were varied to yield a minimum value for χ^2_R. If χ^2_R increases rapidly as a lifetime is varied, then this value may be estimated with little uncertainty. If χ^2_R is not sensitive to variation of a lifetime, then the uncertainty in this estimated value is greater. In general, the lifetimes are known to be within the range where $\chi^2 = \nu \chi^2_R$ increases from the minimum value (χ^2_{min}) to $\chi^2_{min} + 1$.



Dependence of the chi-square for the lifetime in a 3-exponential decay. Chi-square value shown for measurements at 1 emission wavelength and 5 emission wavelengths and when the central lifetime (10 ns) is know. The random error level was 0.1%

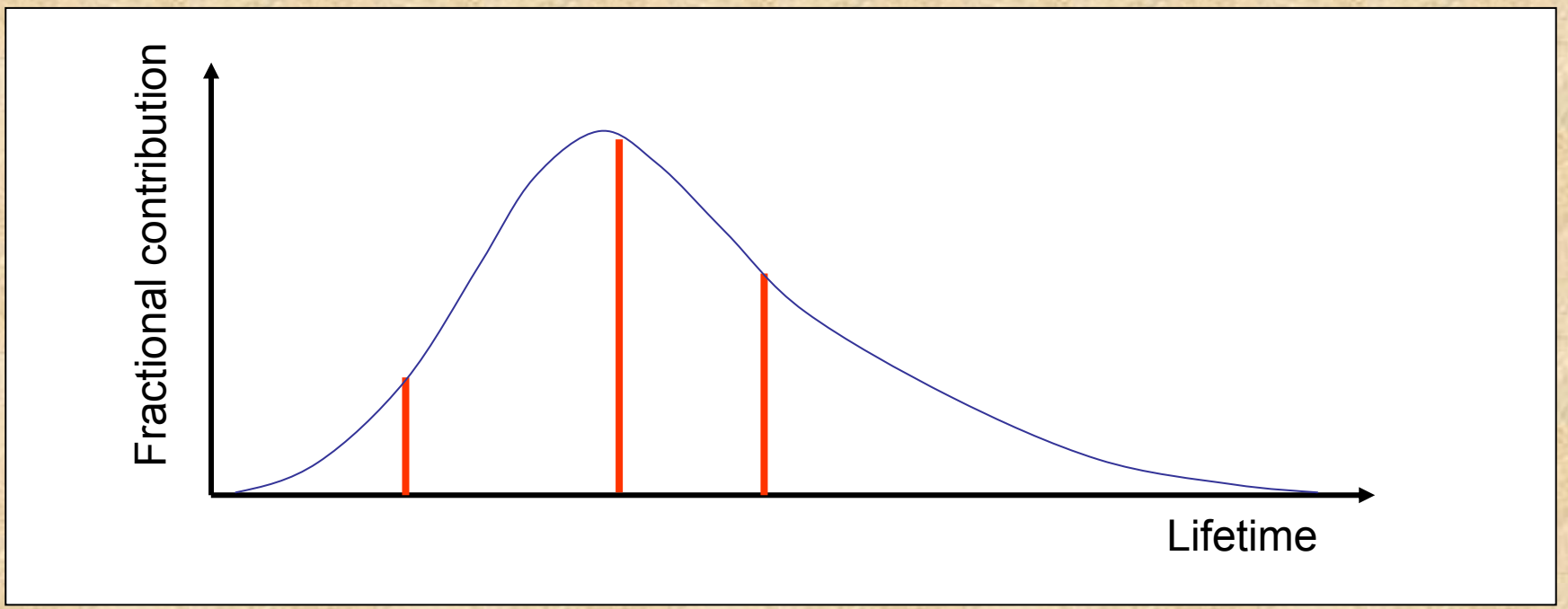
3-lifetimes are difficult to resolve. Multiple emission wavelengths, if the components have different spectra, helps to resolve the components (Global methods)

Continuous distribution of lifetime components

What is a distribution?
When should it be used?

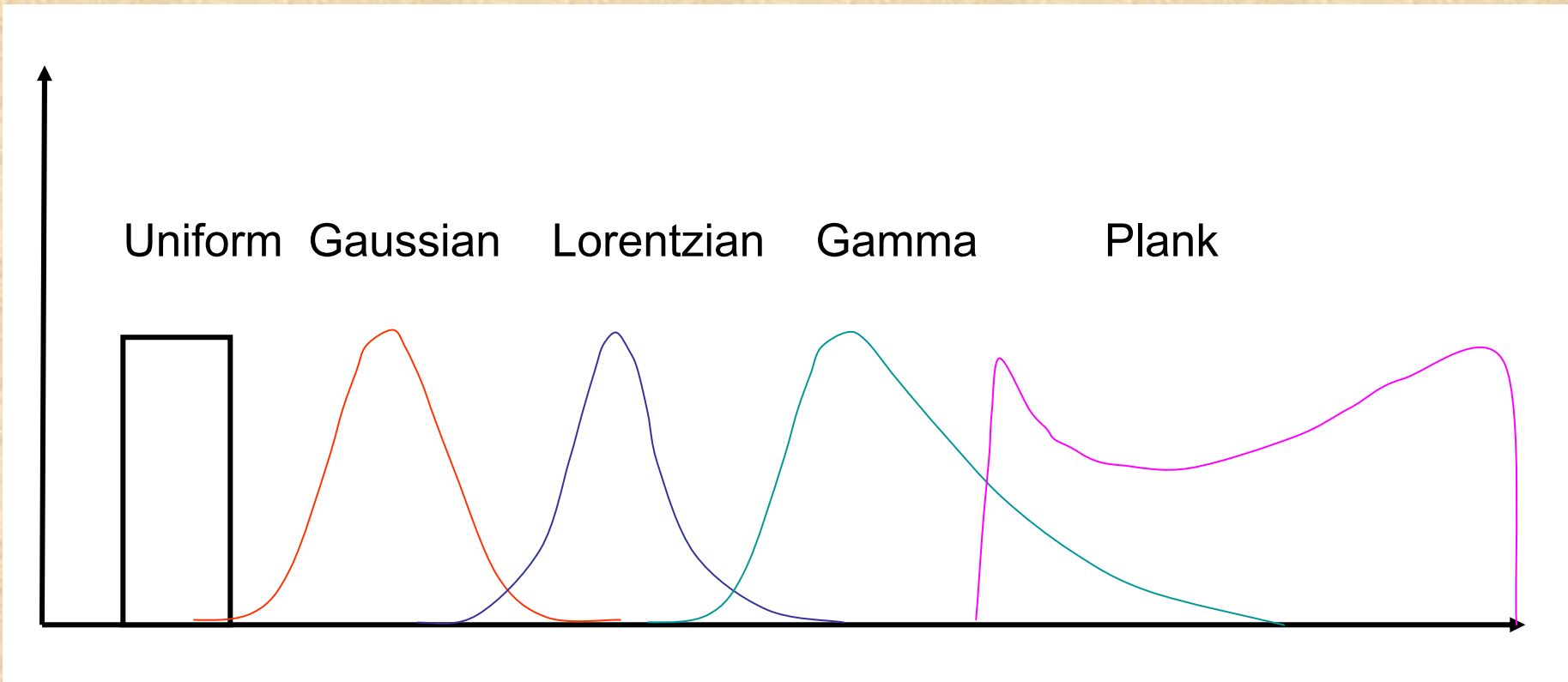
What parameters provide?

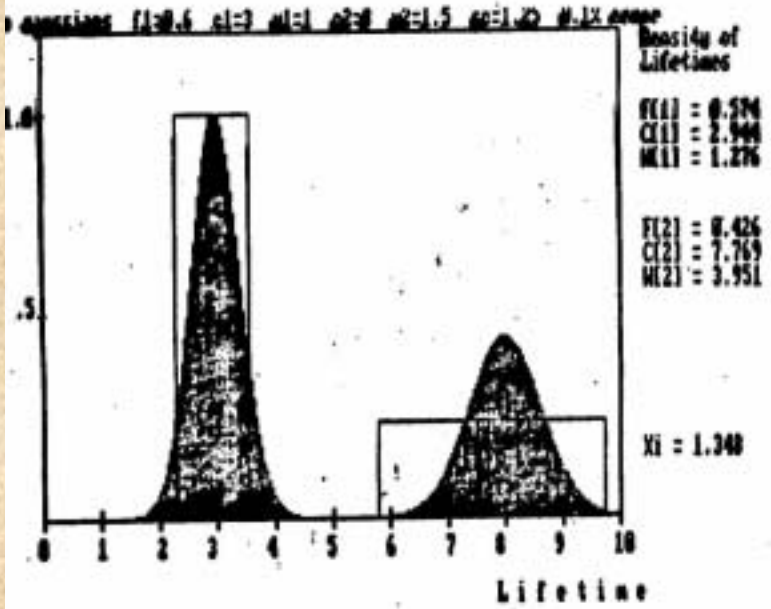
Use and abuse

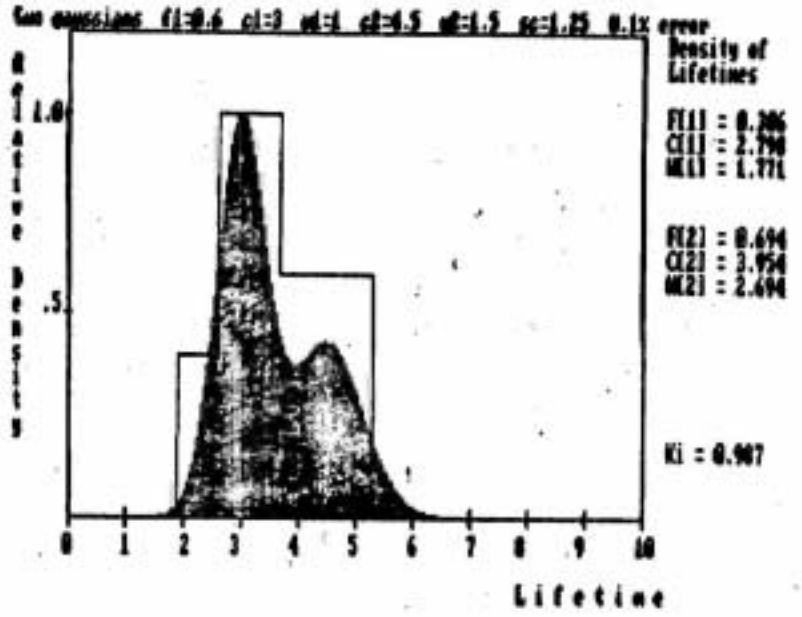


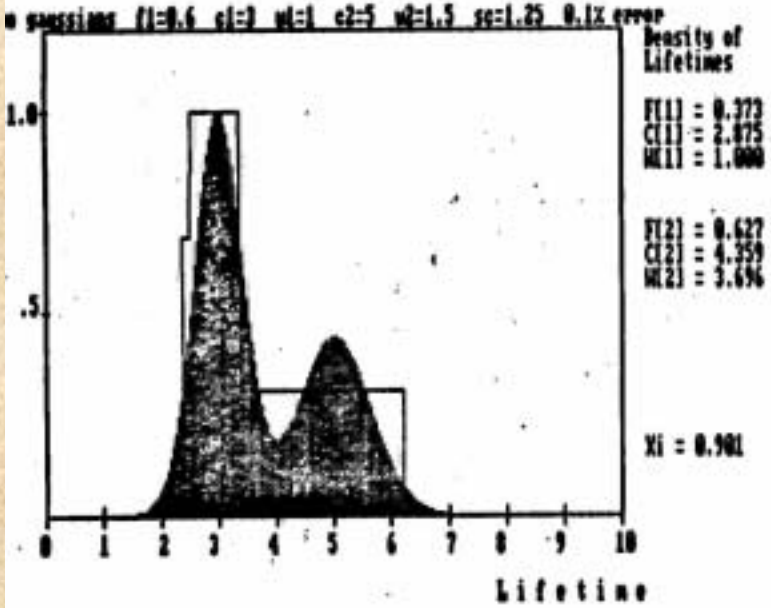
Type of distributions

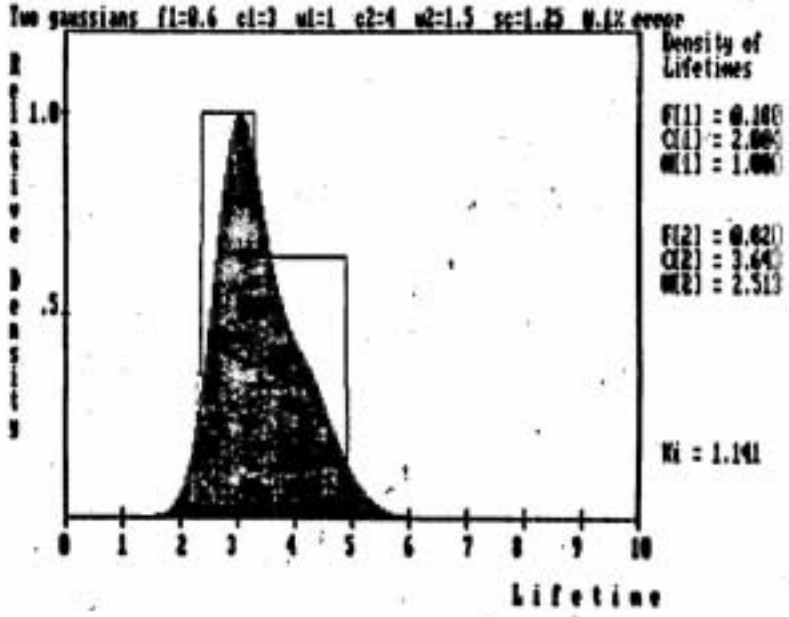
Uniform
Gaussian
Lorentzian
Gamma
Plank
Arbitrary

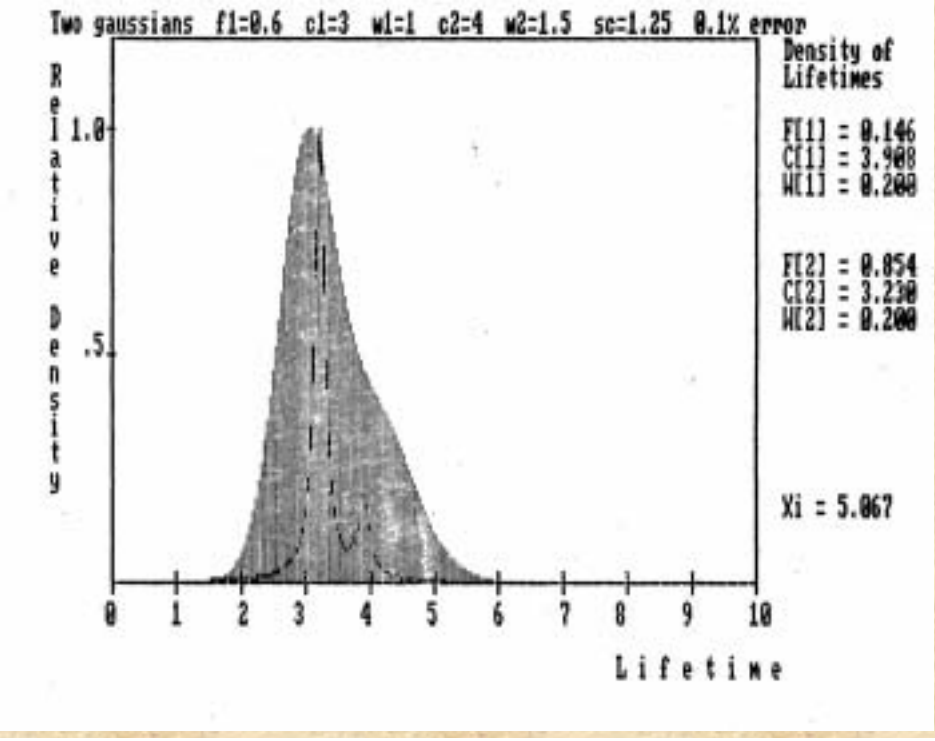


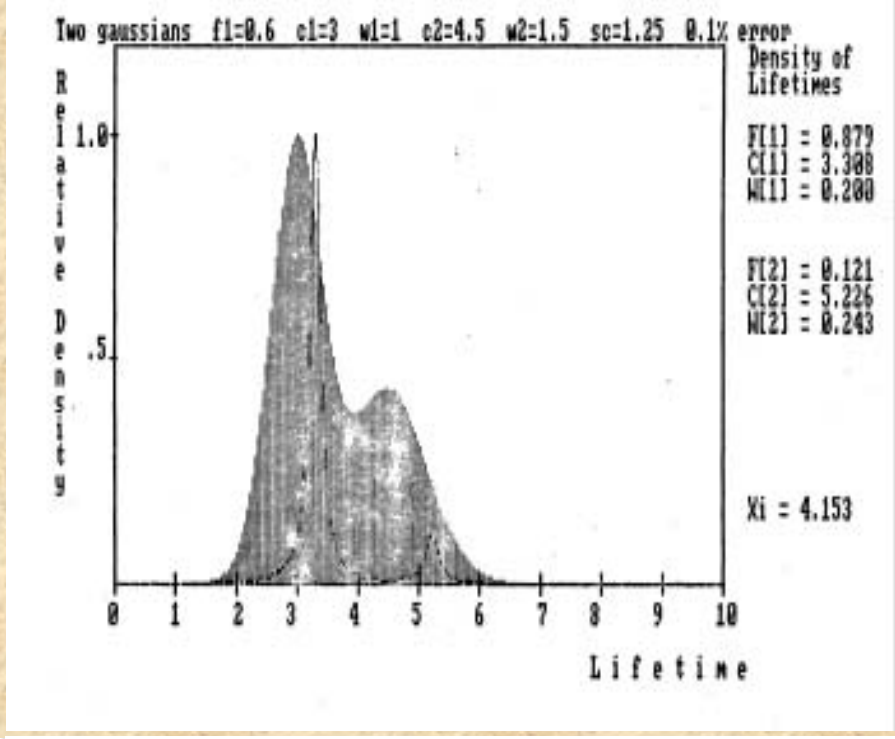

Density of Lifetimes
Xi = 1.340
Lifetime

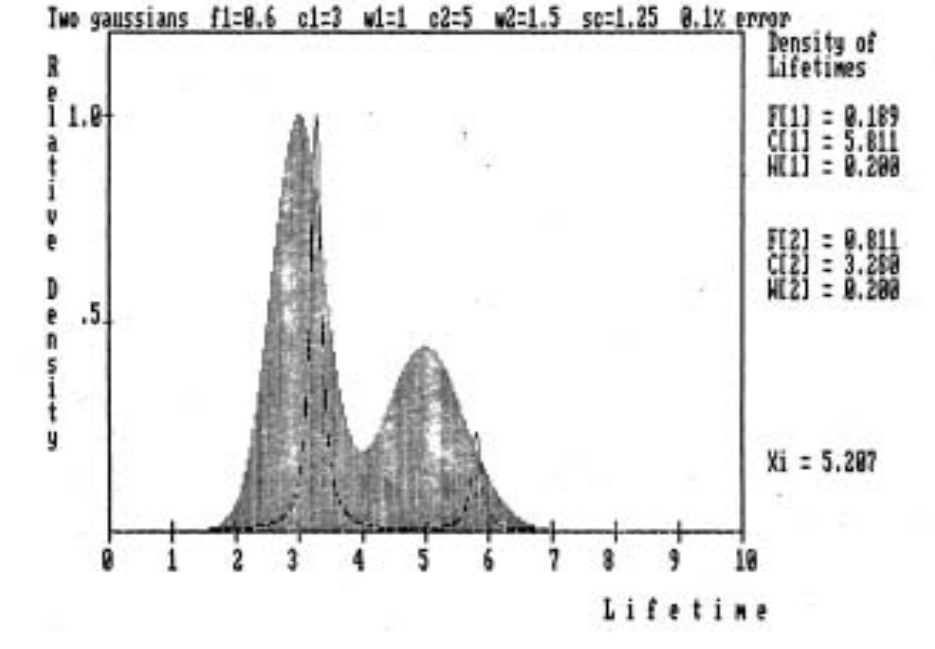

Density of Lifetimes
Xi = 0.907
Lifetime

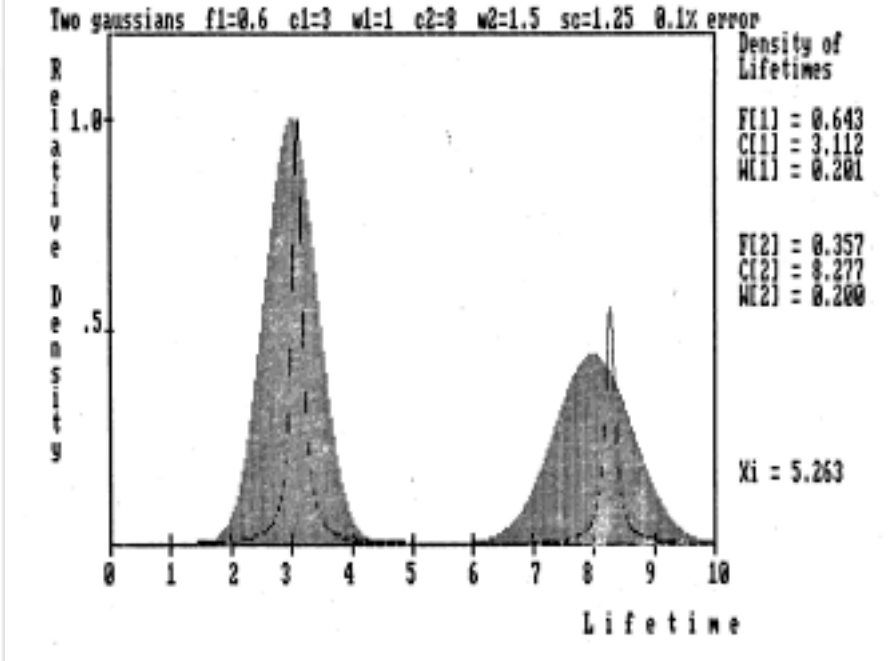

Density of Lifetimes
Xi = 0.901
Lifetime


Density of Lifetimes
Xi = 1.141
Lifetime


Two gaussians f1=0.6 c1=3 w1=1 c2=4 w2=1.5 sc=1.25 0.1% error
Density of Lifetimes
F[1] = 0.146
C[1] = 3.900
W[1] = 0.200
F[2] = 0.854
C[2] = 3.230
W[2] = 0.200
Xi = 5.067
Relative Density
Lifetime


Two gaussians f1=0.6 c1=3 w1=1 c2=4.5 w2=1.5 sc=1.25 0.1% error
Density of Lifetimes
F[1] = 0.879
C[1] = 3.300
W[1] = 0.200
F[2] = 0.121
C[2] = 5.226
W[2] = 0.243
Xi = 4.153
Relative Density
Lifetime


Two gaussians f1=0.6 c1=3 w1=1 c2=5 w2=1.5 sc=1.25 0.1% error
Density of Lifetimes
F[1] = 0.189
C[1] = 5.811
W[1] = 0.200
F[2] = 0.811
C[2] = 3.200
W[2] = 0.200
Xi = 5.207
Relative Density
Lifetime


Two gaussians f1=0.6 c1=3 w1=1 c2=8 w2=1.5 sc=1.25 0.1% error
Density of Lifetimes
F[1] = 0.643
C[1] = 3.112
W[1] = 0.201
F[2] = 0.357
C[2] = 8.277
W[2] = 0.200
Xi = 5.263
Relative Density
Lifetime

Maximum entropy method

The shape of the distribution is not assigned. A number of exponential components are used to obtain the best distribution that maximizes the "entropy". The first guess is a uniform distribution. Then the computer algorithm modifies the amplitude of the components to obtain a best fit

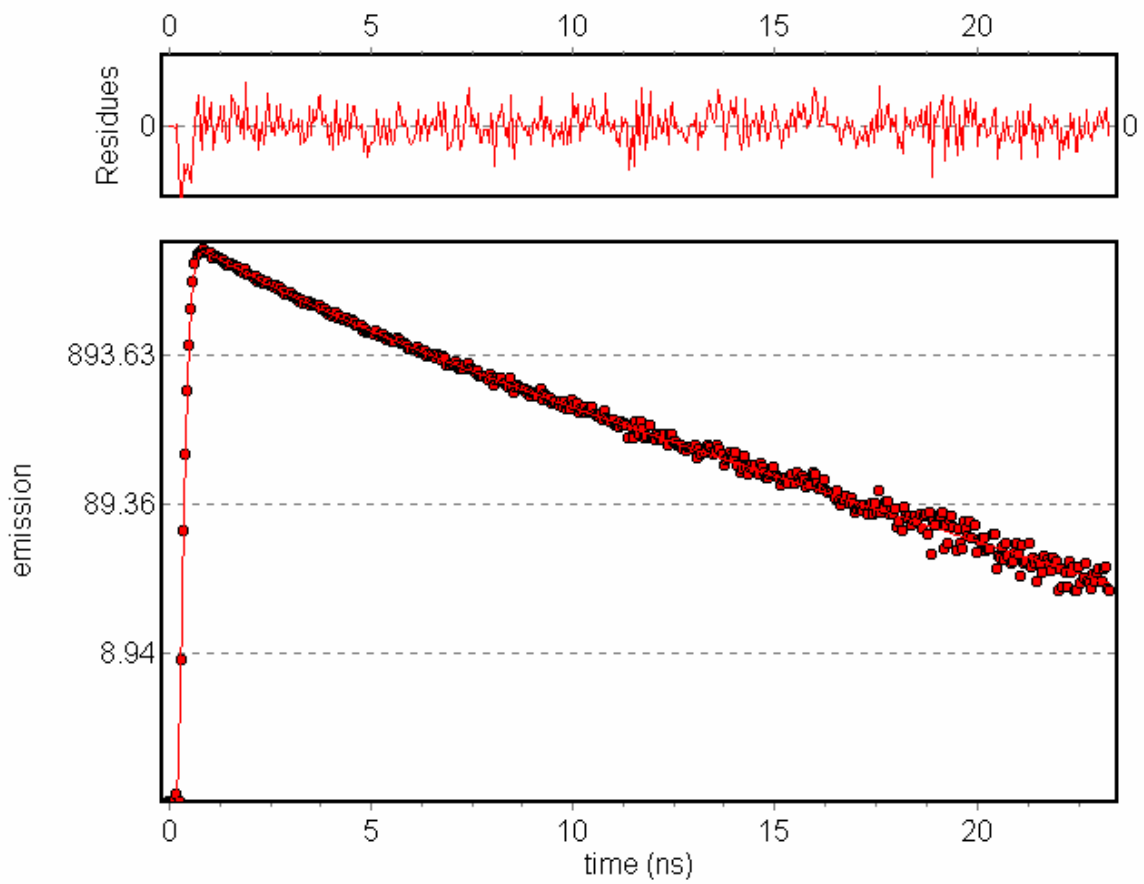
Definition of entropy: In the information theory sense. It is an inversion method:
Given a set of data ***c***, find the best function ***u*** (sum of exponentials) that maximizes the probability to represent the data with some conditional information *I*.

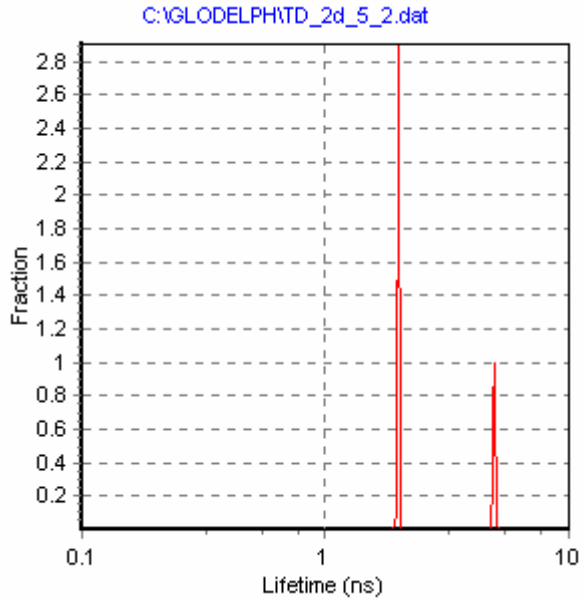
$$\text{Prob}(\boldsymbol{c}|\boldsymbol{u}I)=\exp(-X^2)\Delta u_1\,\Delta u_2\ldots\,\Delta u_m$$

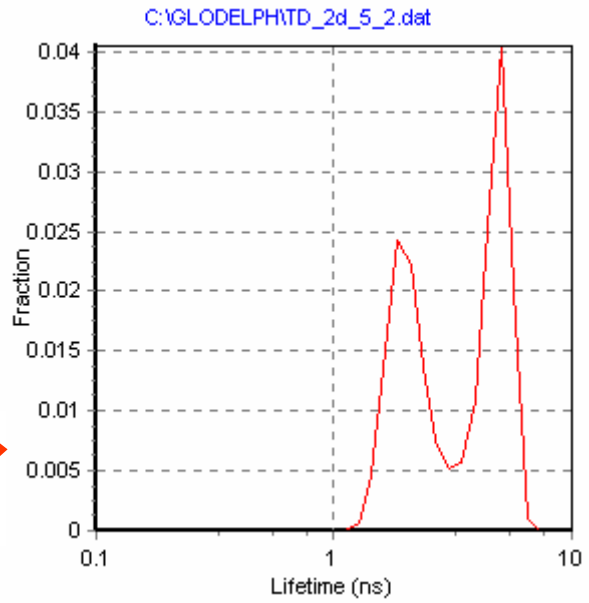
X^2 is the chi-square as defined previously and Δu_m are (constant) increments of the parameters.

[Quote from Numerical methods: "The maximum entropy property has caused MEM to acquire a certain cult popularity; one sometimes hears that it gives a better estimate than it is given by other methods. Don't' believe it"]

Two-discrete exponentials (2 ns and 5 ns) fitted with two gaussians



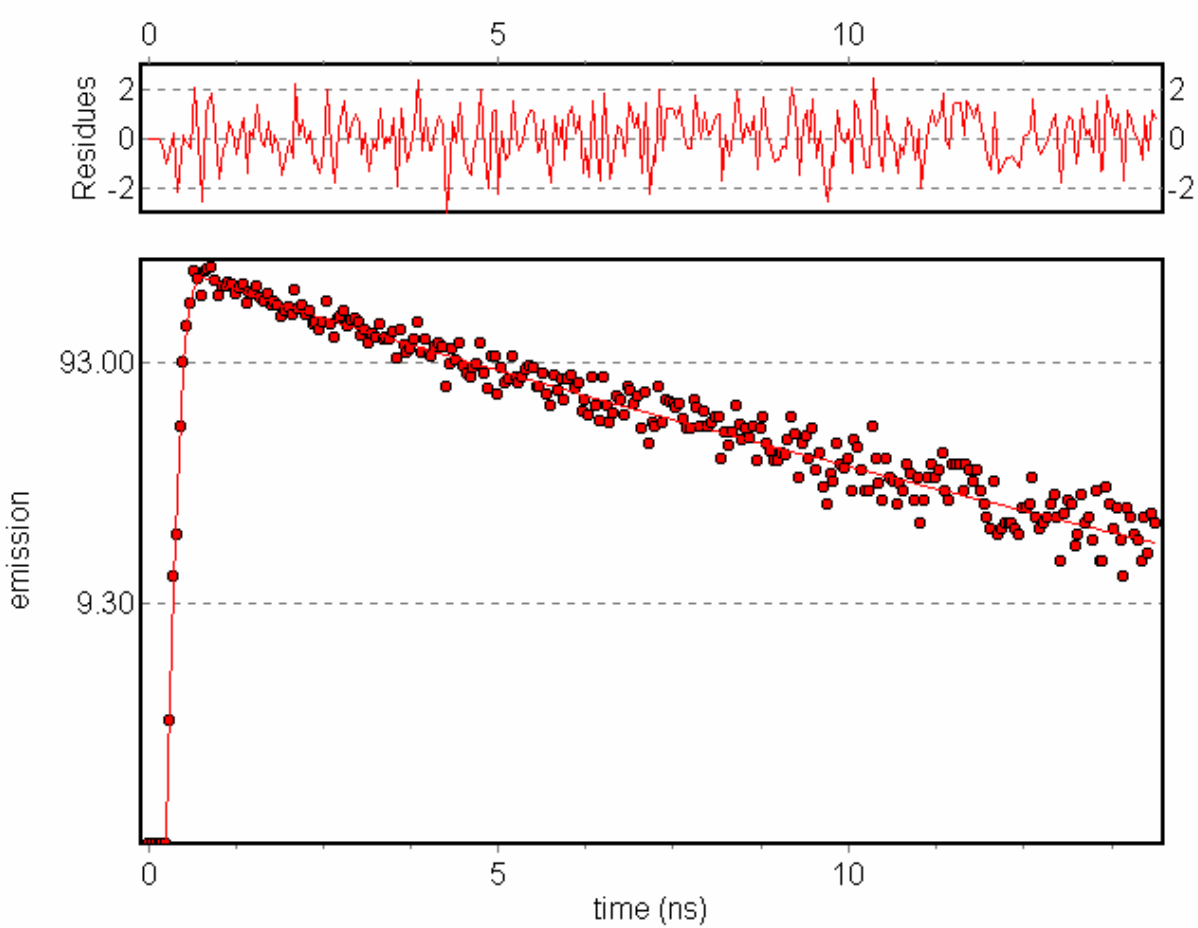


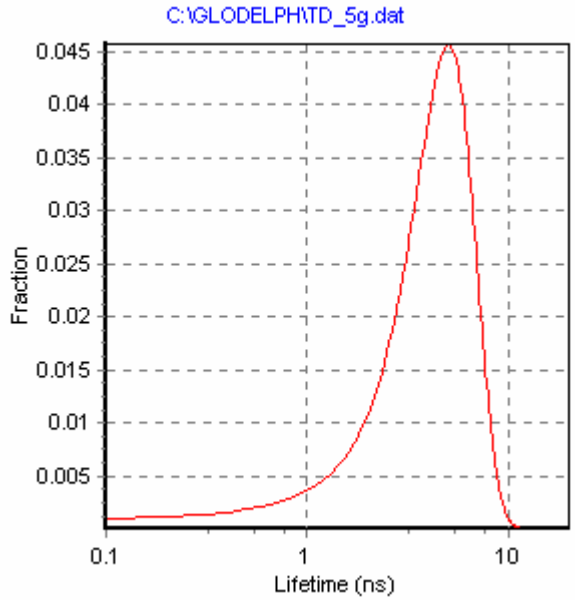



```
Local chisquare =          0.99445
        Sas 1->0 = F             1.0000000
distr gauss 1->0 = V             4.9478493
      width 1->0 = V             0.0584067
        sas 2->0 = V             4.7309626
distr gauss 2->0 = V             1.9901782
      width 2->0 = V             0.0116156
          qshift = V            -0.0042814
      Background = F             0.0000000
```

Maximum entropy result

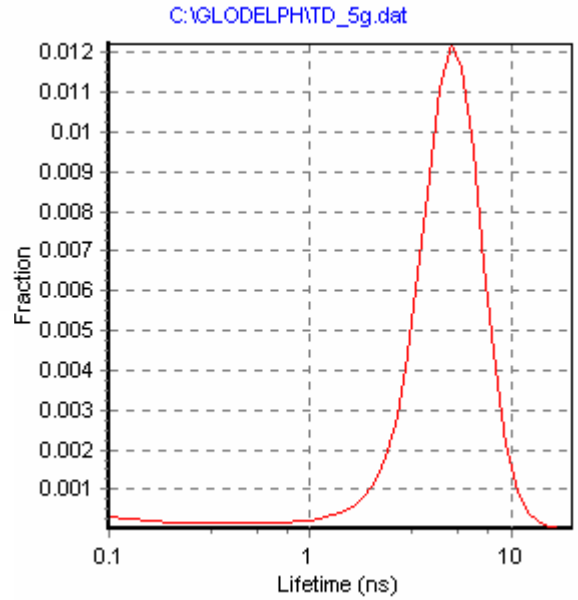
Gaussian distribution, center 5 ns, width 2 ns






```
Local chisquare =          0.99747
        sas 1->0 = V       0.0457739
distr gauss 1->0 = V       5.0295207
      width 1->0 = V       1.7859436
          qshift = V       0.0010388
      Background = F       0.0000000
```

Maximum entropy result



Global analysis methods

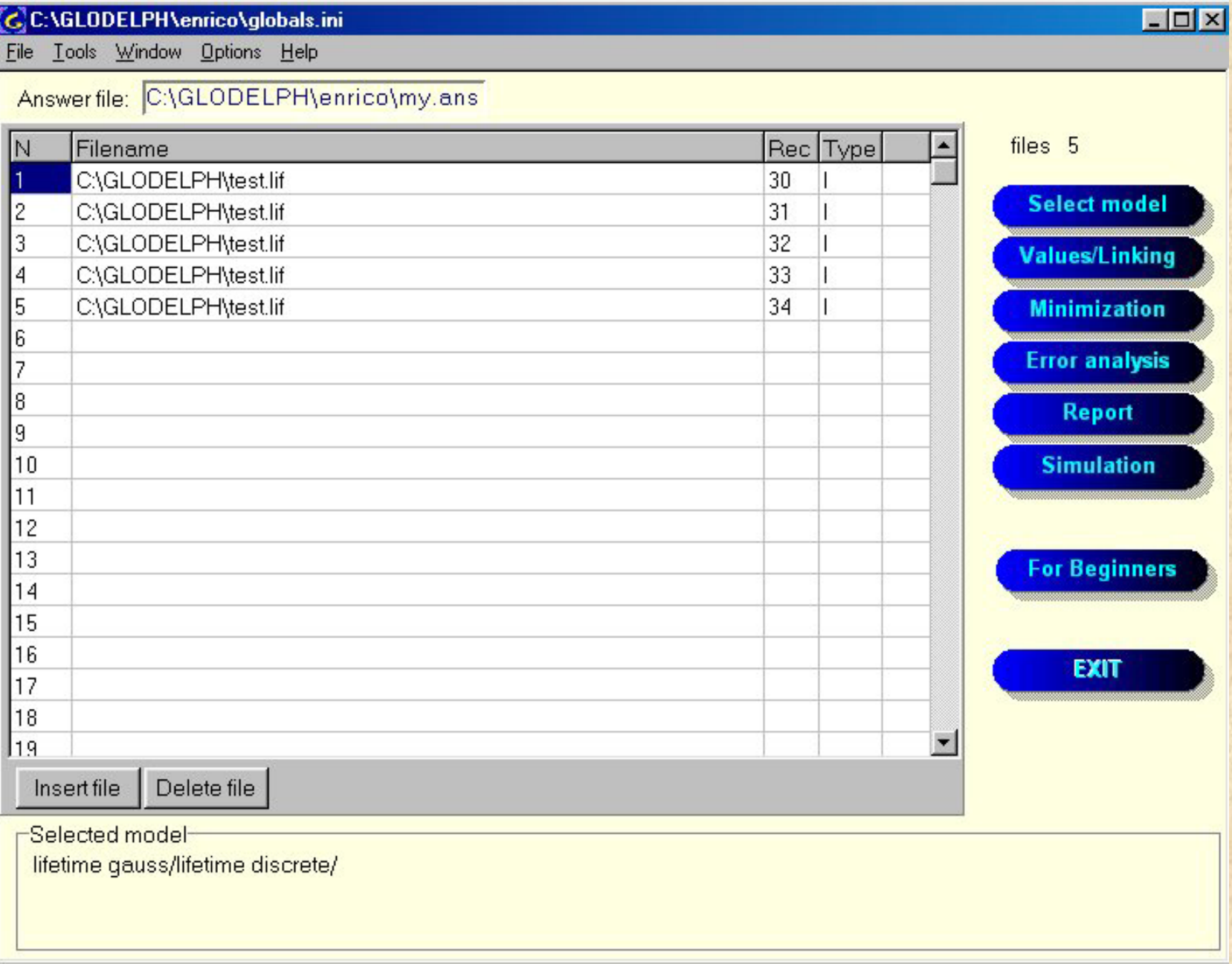
(originally developed by J. Beechem and J. Knutson)

It applies to multiple decay curves

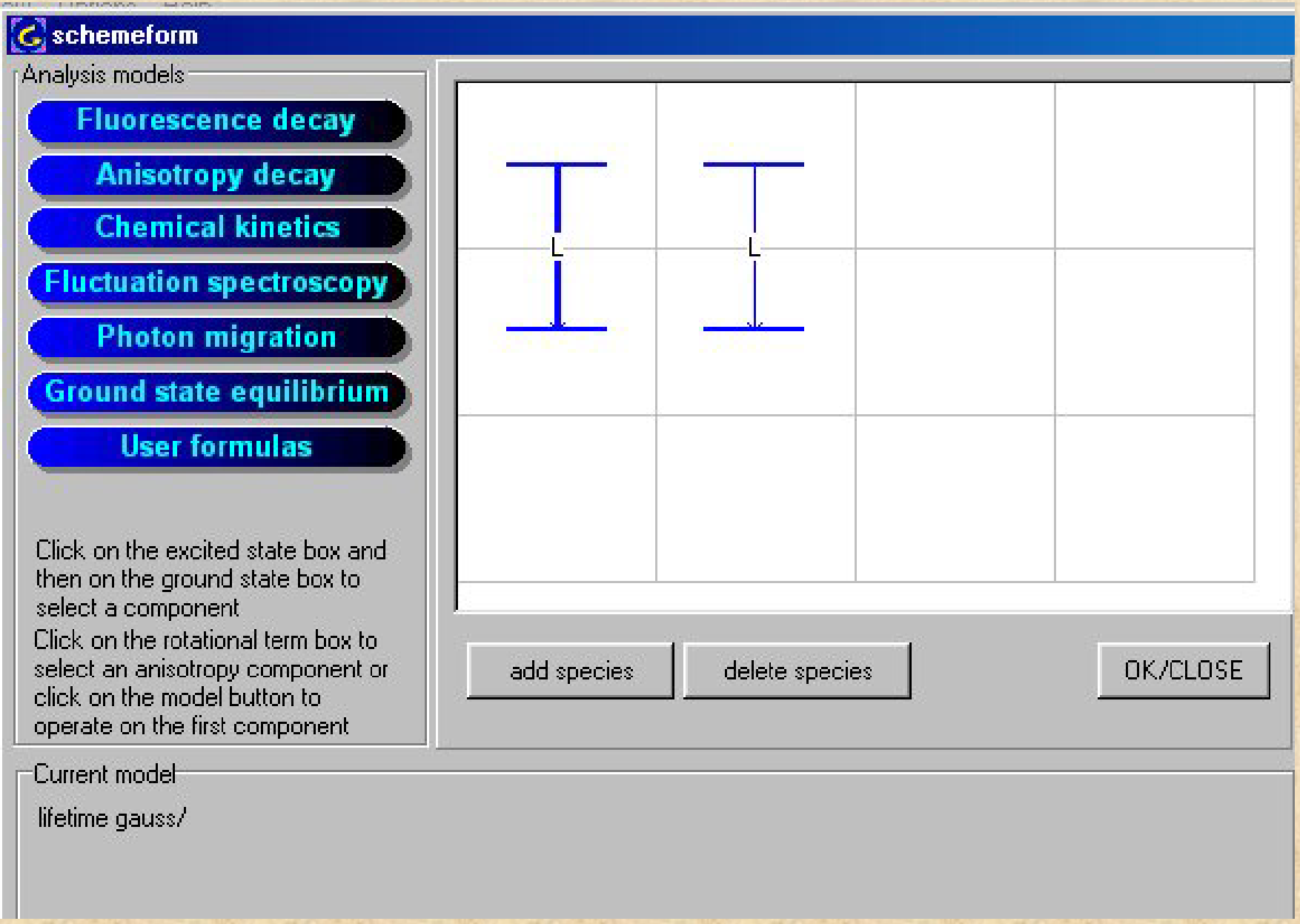
Strategy:

- Establish a decay model
- Establish links between parameters in different files
- Judge the “global” fit based on the global chi-square

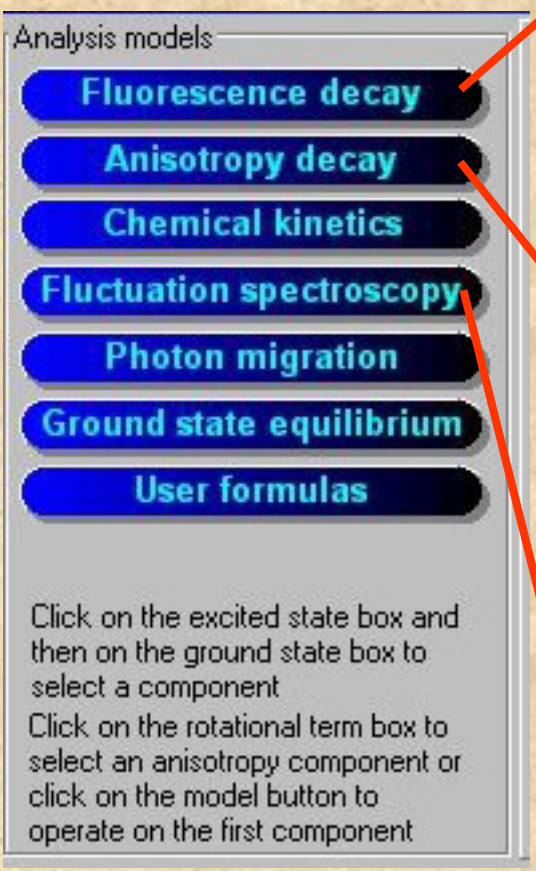
Globals Unlimited entry screen

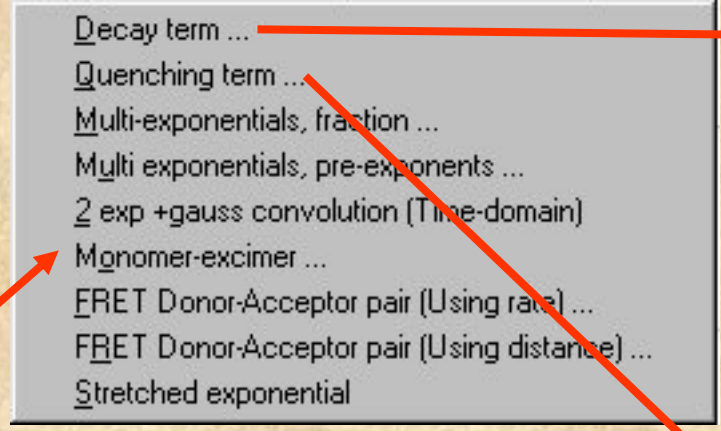


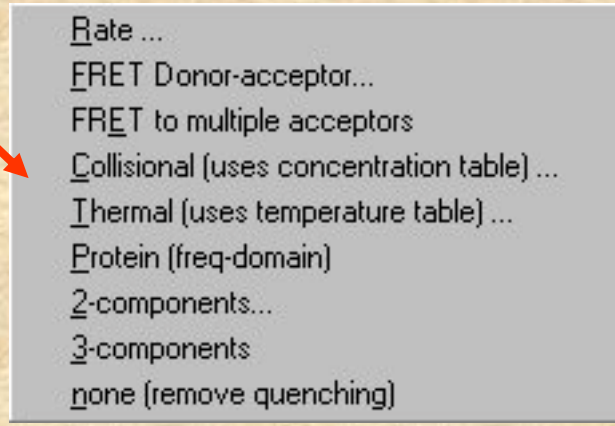
Select model screen

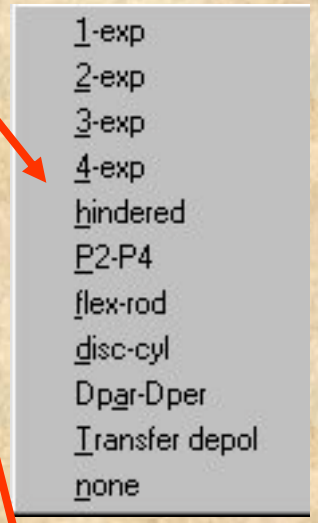


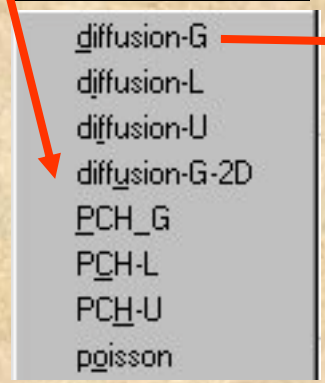
Pre-selected models

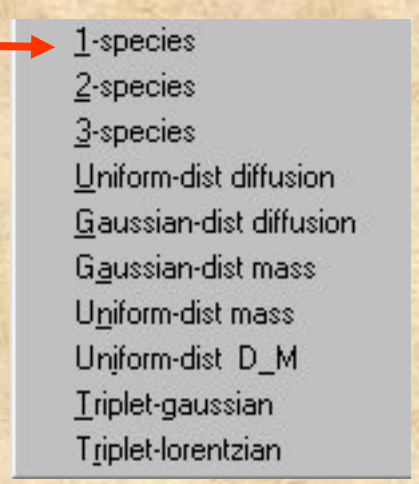




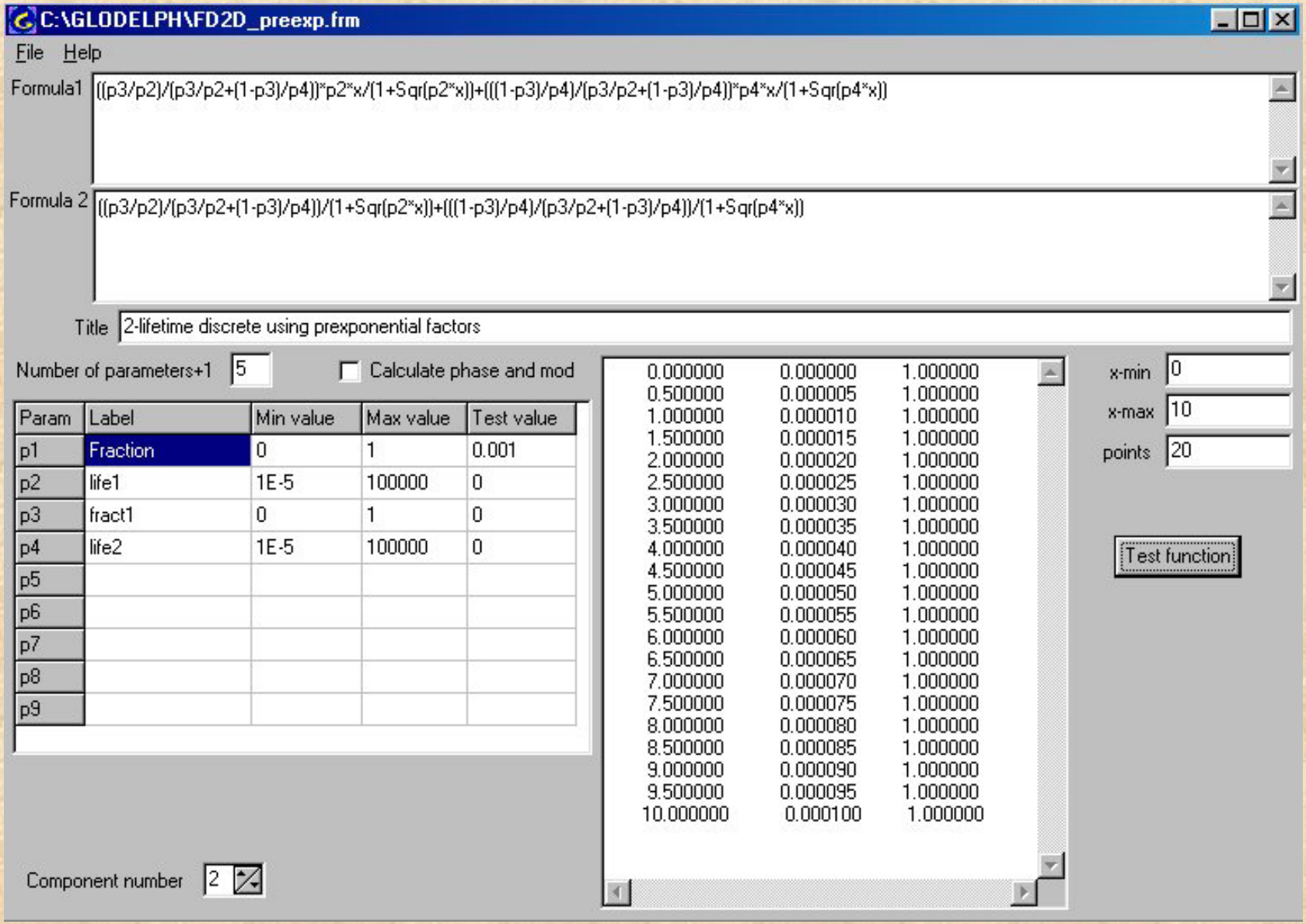




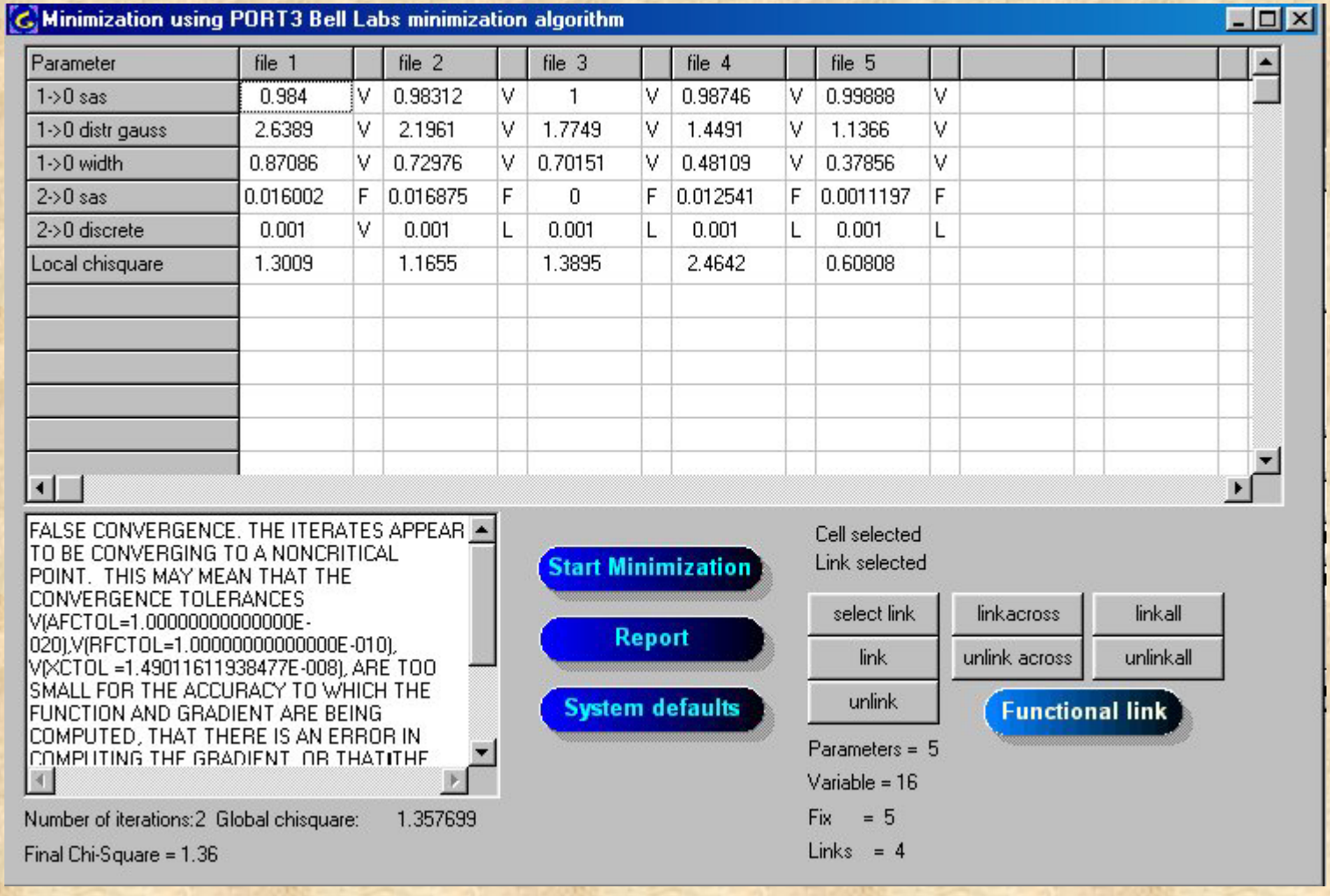




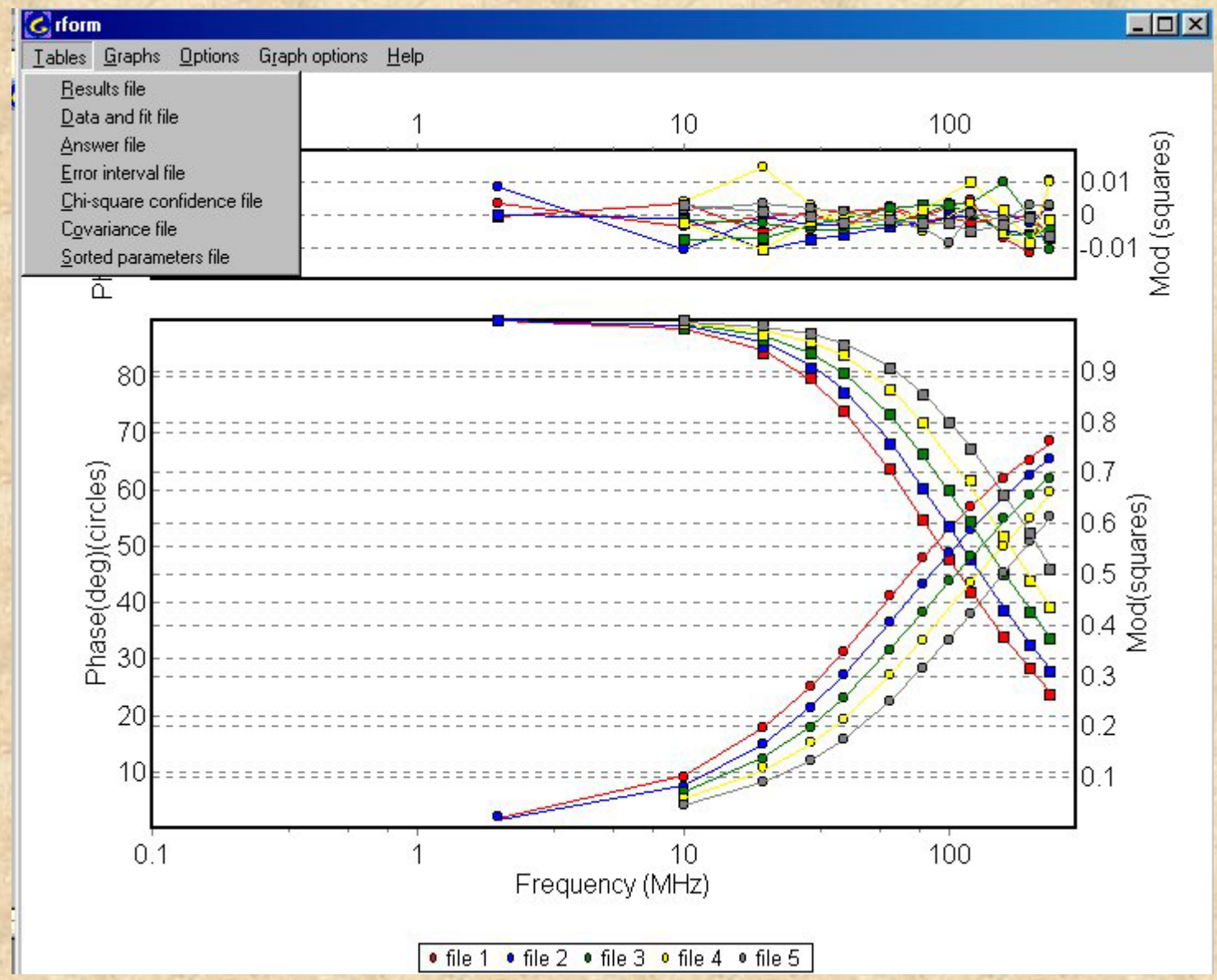
Enter your own model



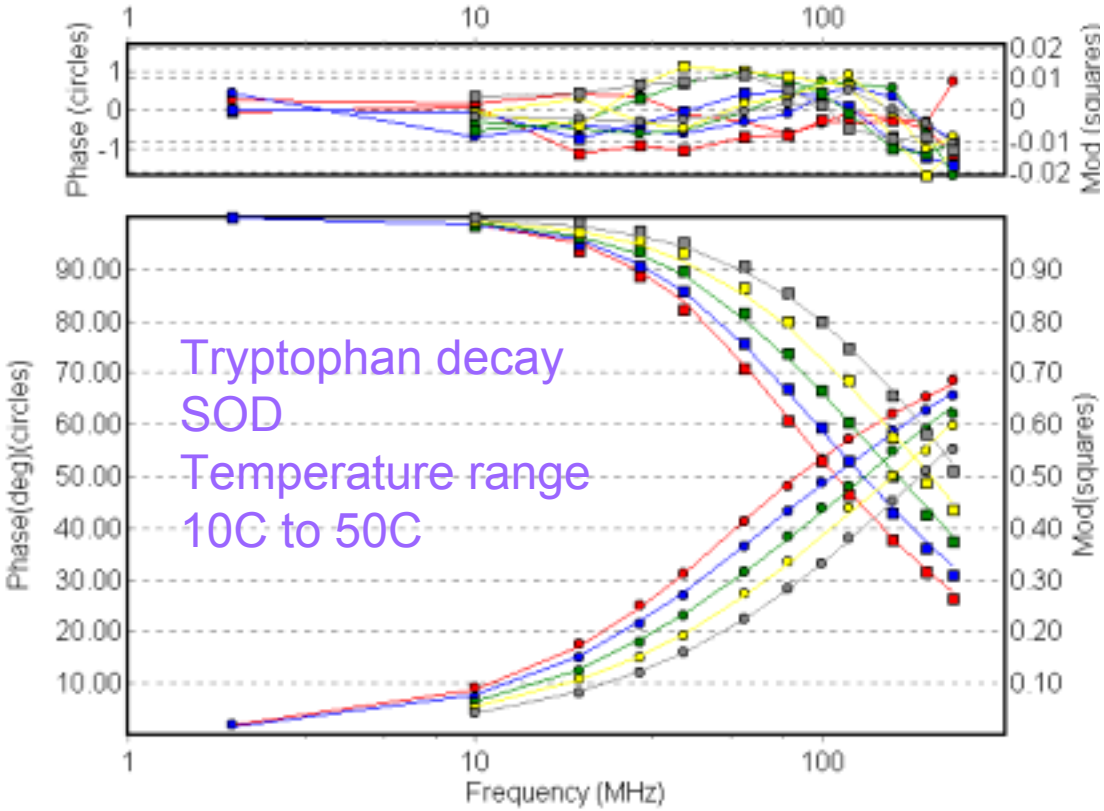
View results/Linking screen



Report screen



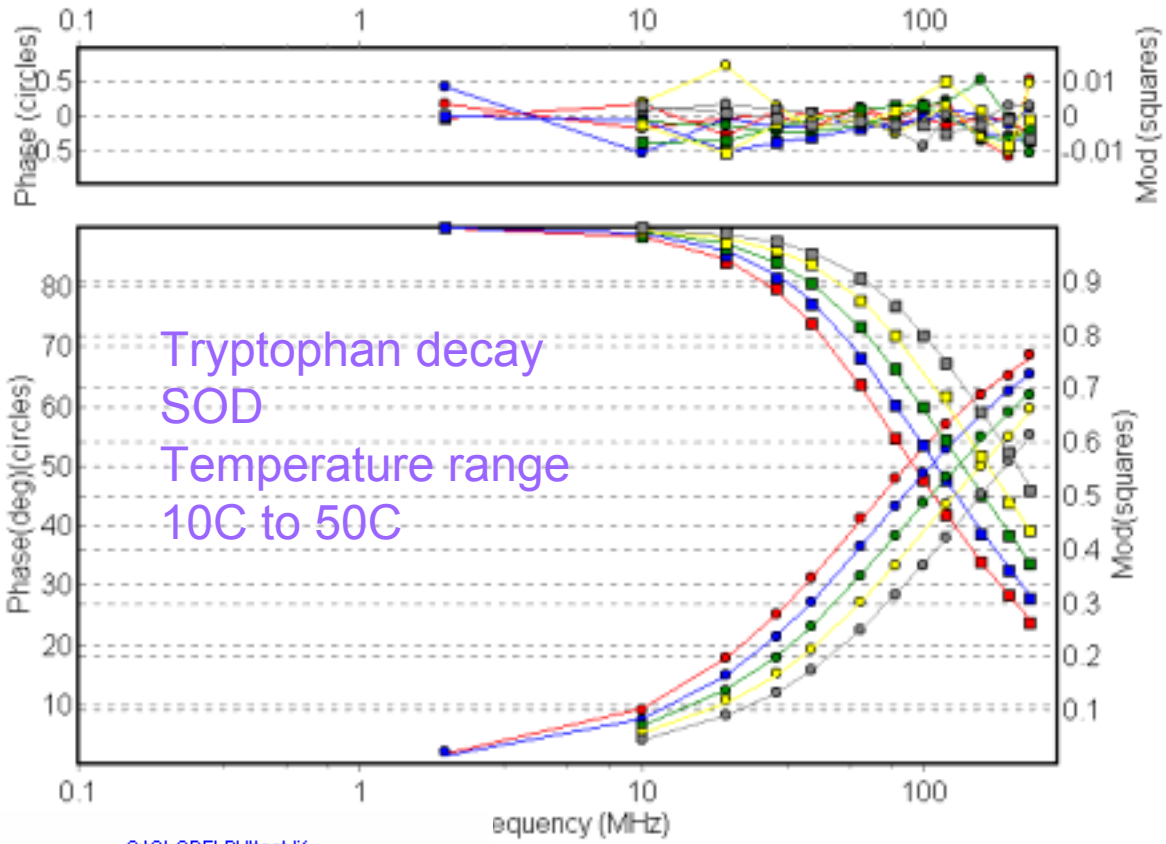
Fit using 2 exponential decay, lifetime linked Global chi-square 6.37

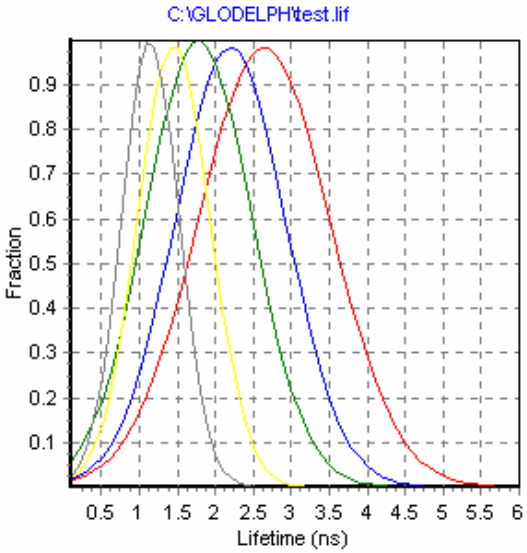


Relatively bad fit, the model is not correct

Everything unlinked
chi-square=2.77
Unlink long component
Chi-square=2.77

```
Final chisquare =       6.36699
Total degrees of freedom =       112
Total number parameters =        11
Final fitting parameters are:
Experiment #                    1 results:
Local chisquare =         4.65379
     sas 1->0 = V          0.9130451
discrete 1->0 = V          2.6736159
     sas 2->0 = F          0.0869549
discrete 2->0 = V          0.5527723
Experiment #                    2 results:
Local chisquare =         5.56336
     sas 1->0 = V          0.7257106
discrete 1->0 = L          2.6736159
     sas 2->0 = F          0.2742894
discrete 2->0 = V          0.9613297
Experiment #                    3 results:
Local chisquare =         9.64283
     sas 1->0 = V          0.4820184
discrete 1->0 = L          2.6736159
     sas 2->0 = F          0.5179816
discrete 2->0 = V          1.0744863
Experiment #                    4 results:
Local chisquare =         8.34
     sas 1->0 = V          0.2939740
discrete 1->0 = L          2.6736159
     sas 2->0 = F          0.7060260
discrete 2->0 = V          1.0354601
Experiment #                    5 results:
Local chisquare =         3.99731
     sas 1->0 = V          0.1590942
discrete 1->0 = L          2.6736159
     sas 2->0 = F          0.8409058
discrete 2->0 = V          0.9299485
```





Fit using 1 gaussian distribution+scattering

Global chi-square=1.36

The fit improves dramatically

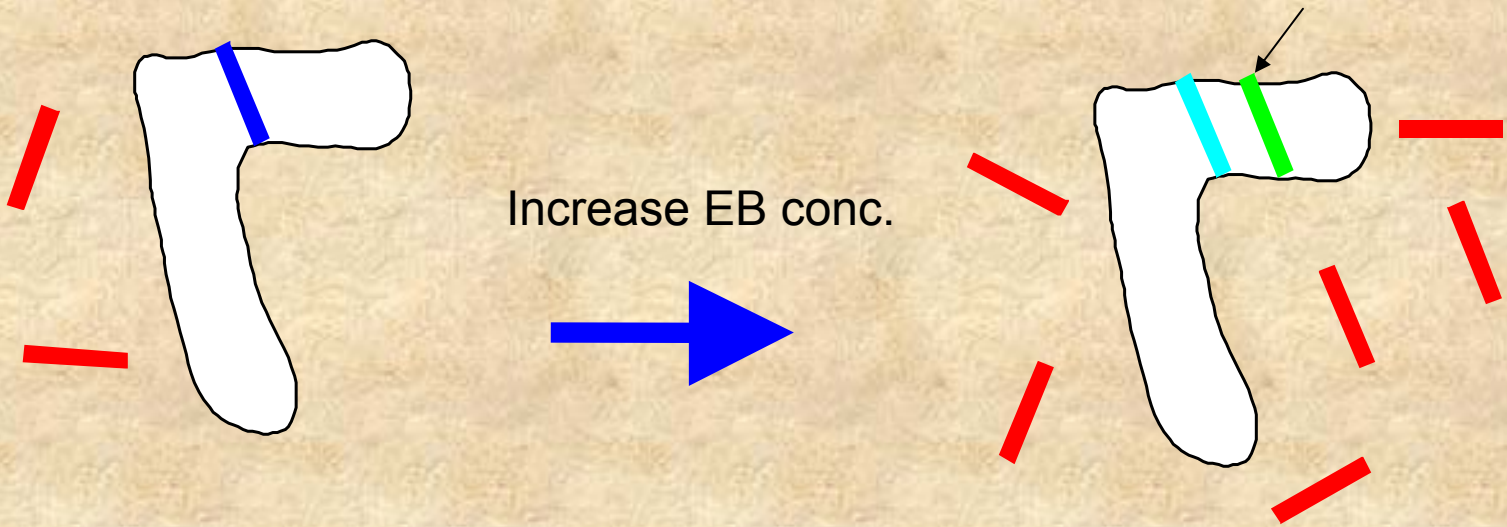
```
Final chisquare =       1.36036
Experiment #          1 results:
Local chisquare =         1.30053
        sas 1->0 = V    0.9839553
distr gauss 1->0 = V    2.6389380
      width 1->0 = V    0.8706837
        sas 2->0 = F    0.0160447
   discrete 2->0 = V    0.0010000
Experiment #          2 results:
Local chisquare =         1.16505
        sas 1->0 = V    0.9831309
distr gauss 1->0 = V    2.1961937
      width 1->0 = V    0.7302908
        sas 2->0 = F    0.0168691
   discrete 2->0 = L    0.0010000
Experiment #          3 results:
Local chisquare =         1.39069
        sas 1->0 = V    1.0000000
distr gauss 1->0 = V    1.7746593
      width 1->0 = V    0.7019813
        sas 2->0 = F    0.0000000
   discrete 2->0 = L    0.0010000
Experiment #          4 results:
Local chisquare =         2.47840
        sas 1->0 = V    0.9873509
distr gauss 1->0 = V    1.4496568
      width 1->0 = V    0.4798635
        sas 2->0 = F    0.0126491
   discrete 2->0 = L    0.0010000
Experiment #          5 results:
Local chisquare =         0.60852
        sas 1->0 = V    0.9988288
distr gauss 1->0 = V    1.1366306
      width 1->0 = V    0.3788386
        sas 2->0 = F    0.0011712
   discrete 2->0 = L    0.0010000
```

Example of application of Global Methods to the analysis of "real" data. From D. M. Jameson

T_RNA Et-Br binding experiments

"Strong" binding site

"Weak" binding site

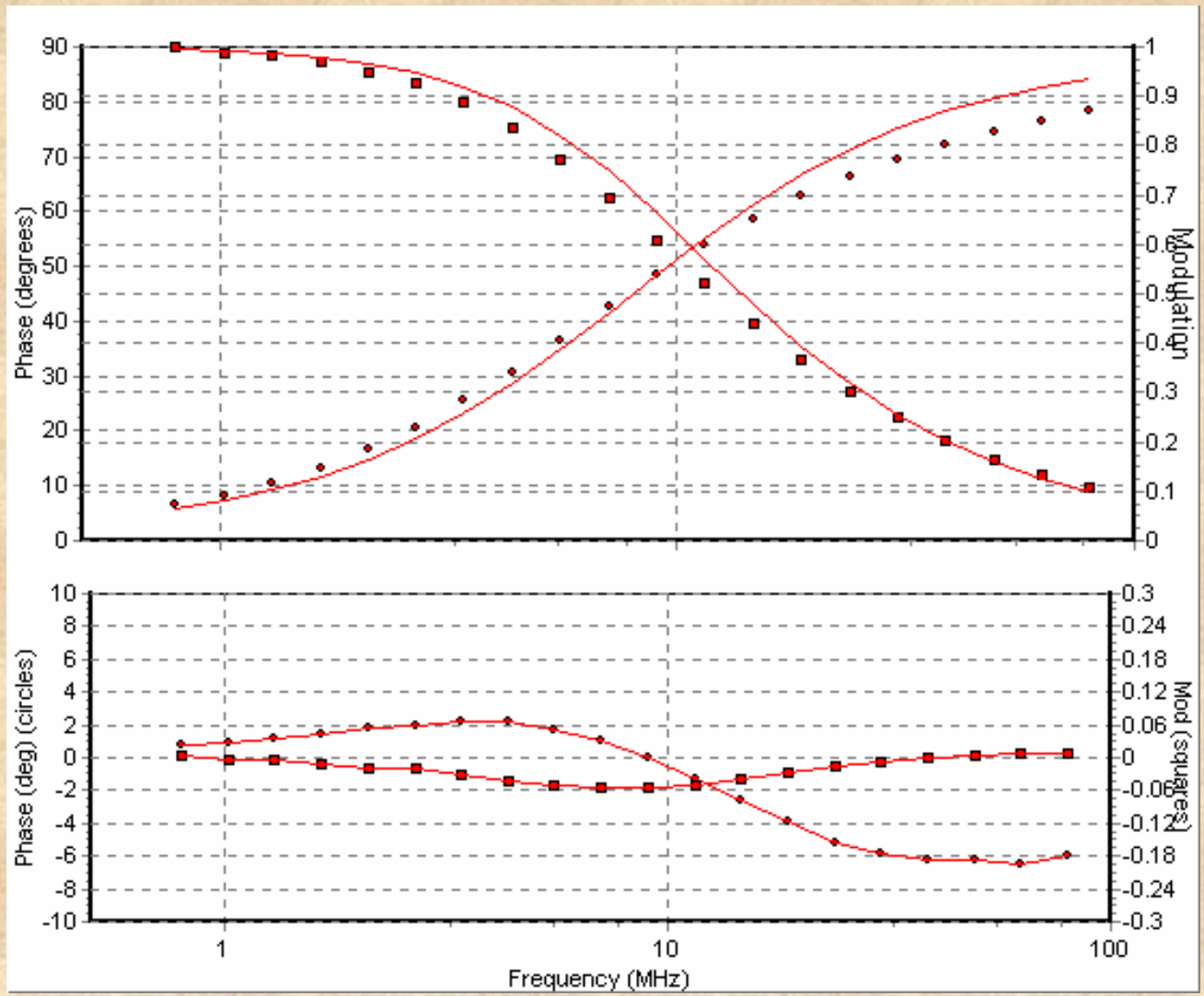


Question: What is the lifetime of the strong and the weak binding site???

D. M. Jameson

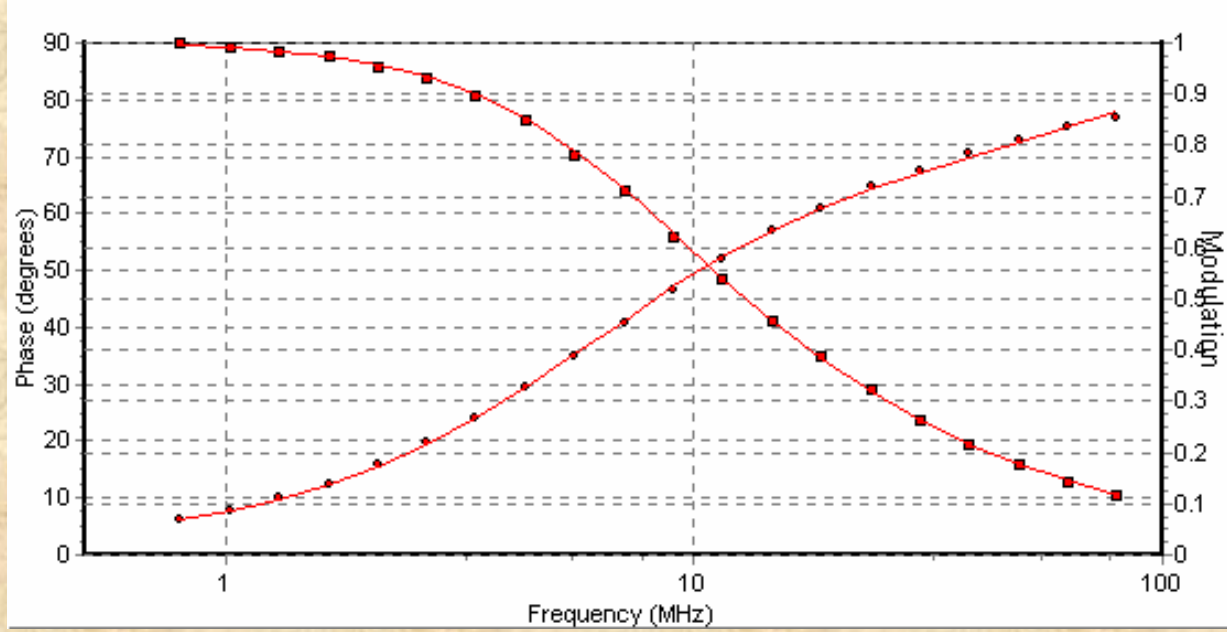
If the tRNA is in excess only one EB will bind to the “strong” binding site which has a K_d of around 1 micromolar. If the EB/tRNA ratio is increased, one or more additional EB’s will bind and the question is: What are the lifetimes of EB bound to different sites on tRNA?” Show below are phase and modulation data for a solution containing 124 μM yeast $tRNA^{phe}$ and 480 μM EB

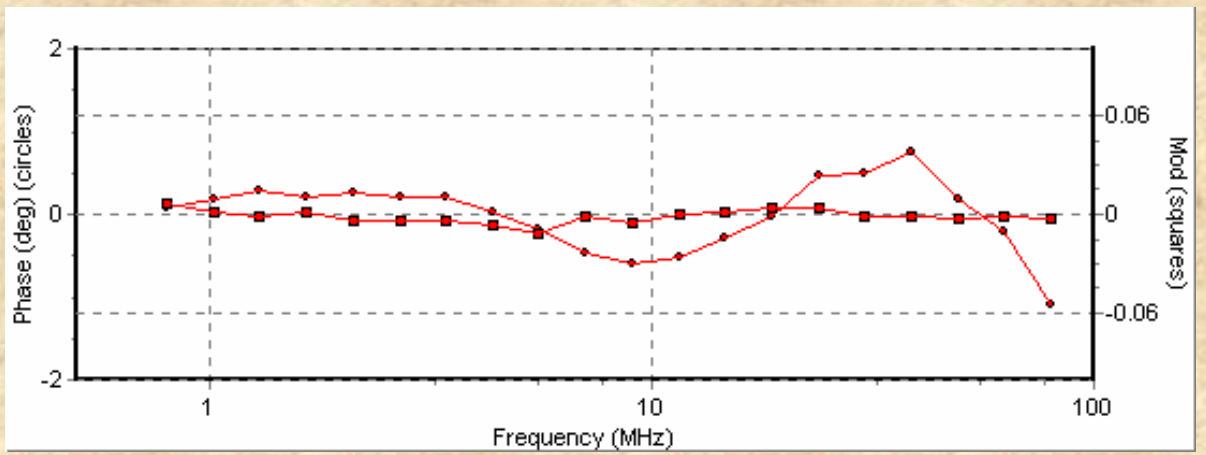
The phase and modulation data were first fit to a single exponential component shown as the solid lines in the top plot. The residuals for this fit are shown in the bottom plot.
In this case $\tau = 18.49$ ns and the χ^2 value was 250.



The data were then fit to a 2-component model shown here In this case the two lifetime components were 22.71 ns with a fractional intensity of 0.911 and 3.99 ns with a fractional intensity of 0.089.

The χ^2 for this fit was 3.06 (note the change in scale for the residual plot compared to the first case shown).





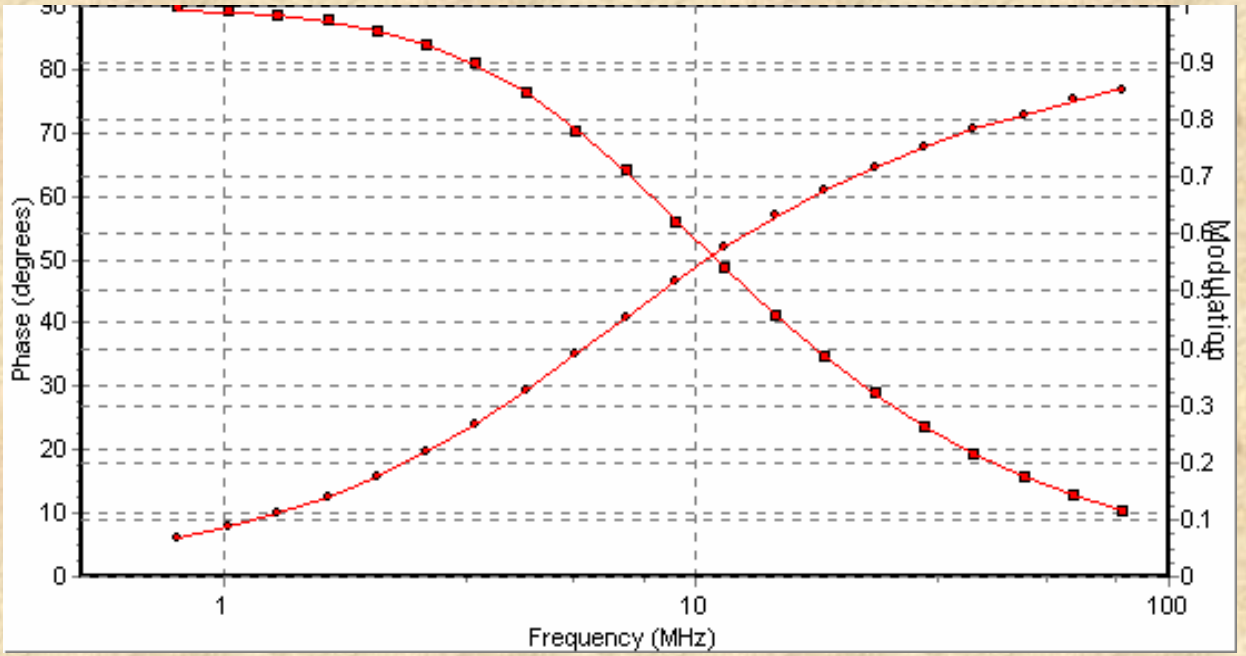
A 3-component model improves the fit still more. In this case

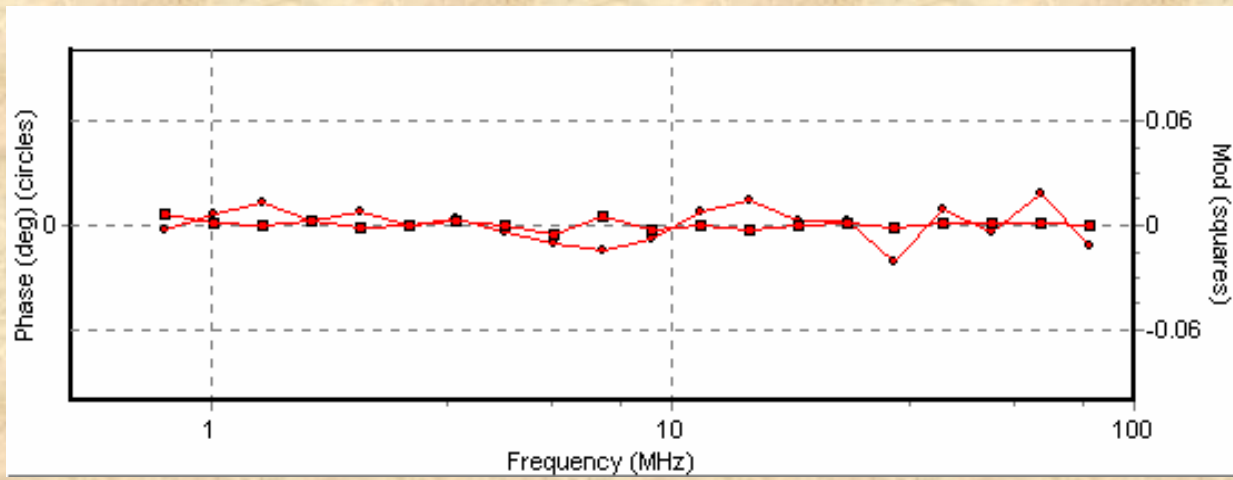
τ_1 = 24.25 ns, f_1 = 0.83

τ_2 = 8.79 ns, f_2 = 0.14

τ_3 = 2.09 ns, f_3 = 0.03

χ^2 = 0.39.



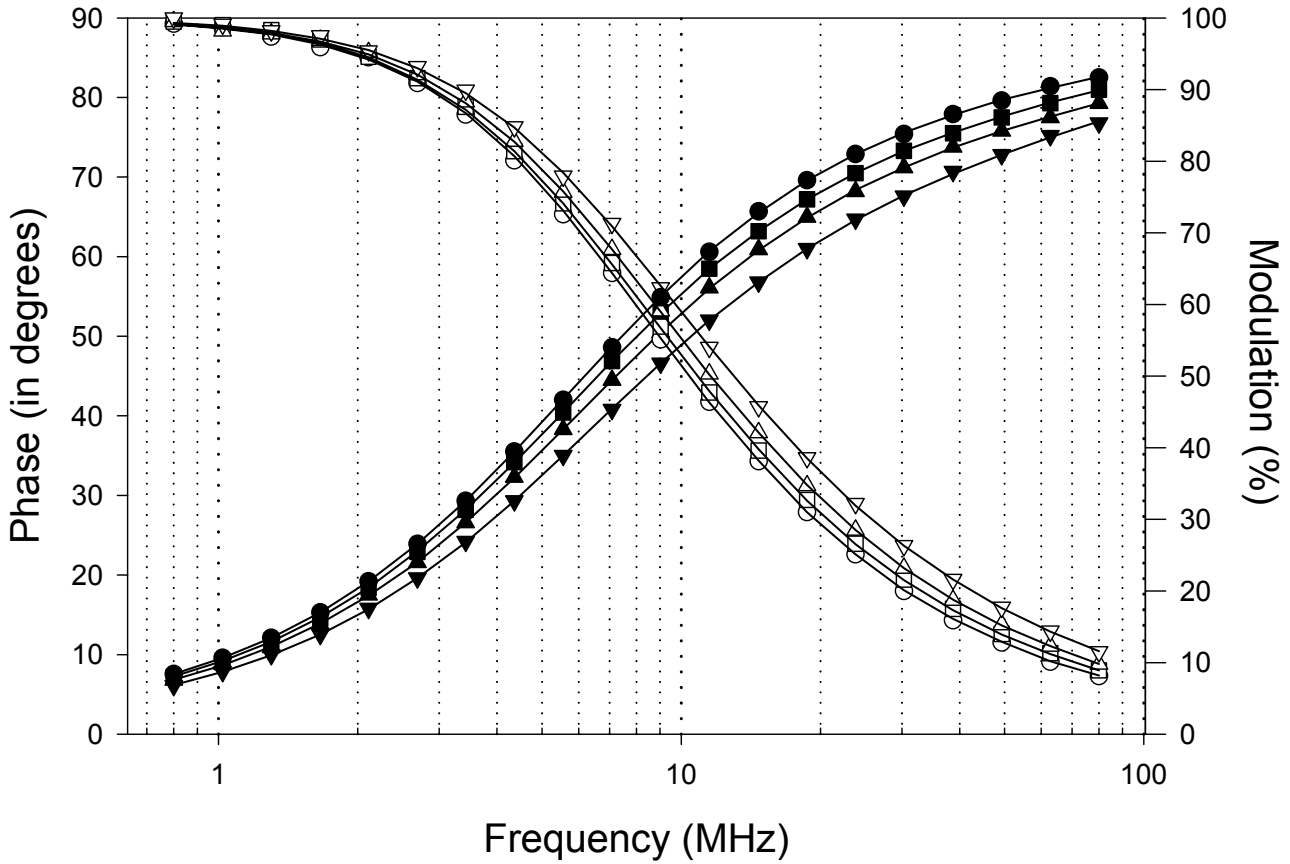


Adding a fourth component – with all parameters free to vary - does not lead to a significant improvement in the χ^2. In this case one finds 4 components of 24.80 ns (0.776), 12.13ns (0.163), 4.17 ns (0.53) and 0.88 ns (0.008).

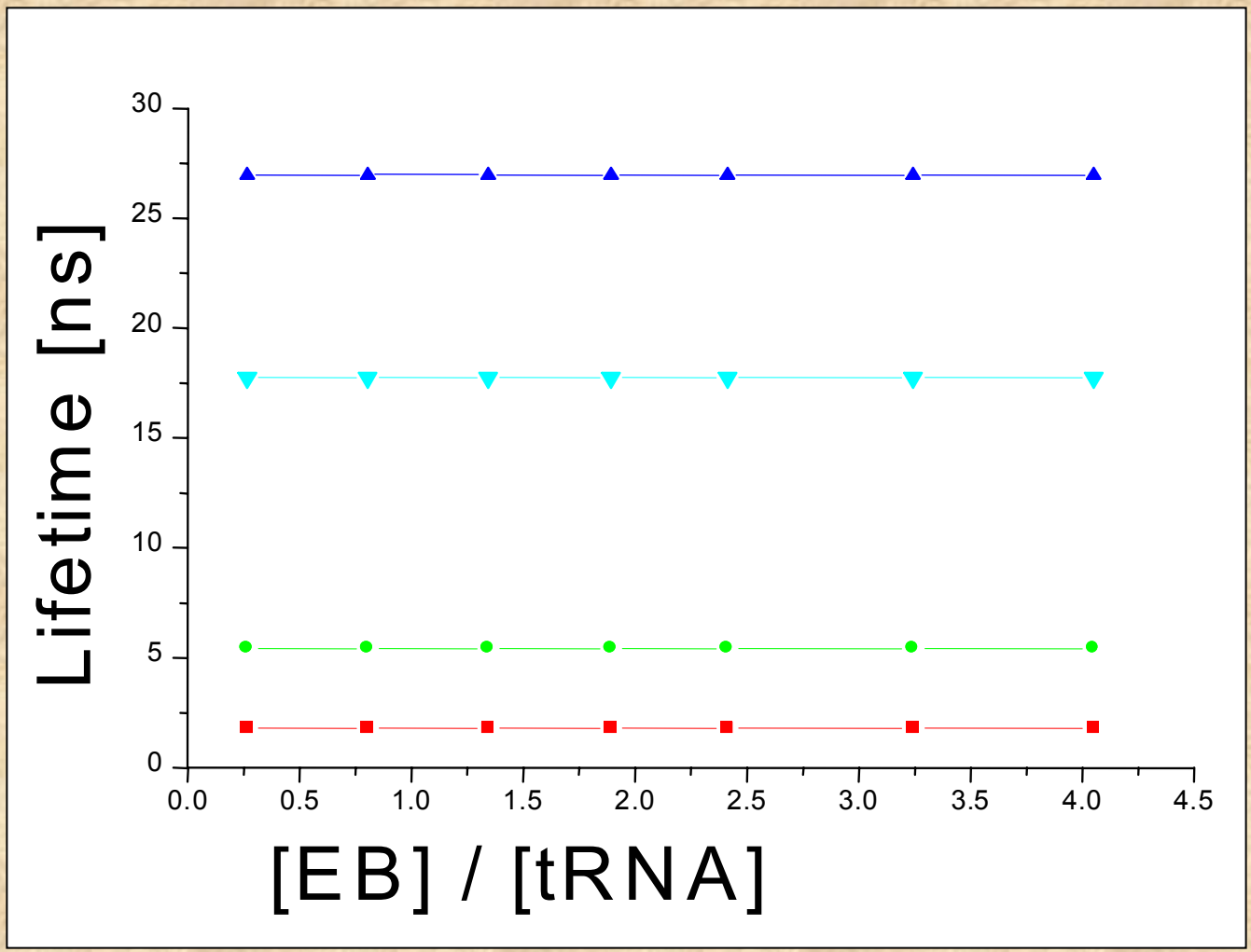
But we can actually fix some of the components in this case. We know that free EB has a lifetime of 1.84 ns and we also know that the lifetime of EB bound to the "strong" tRNA binding site is 27 ns. So we can fix these in the analysis. The results are four lifetime components of 27 ns (0.612), 18.33 ns (0.311), 5.85 ns (0.061) and 1.84 ns (0.016). The χ^2 improves to 0.16.

One can also carry out "Global Analysis" on such data. In Global Analysis multiple data sets are analyzed simultaneously and different parameters (such as lifetimes) can be "linked" across the data sets.

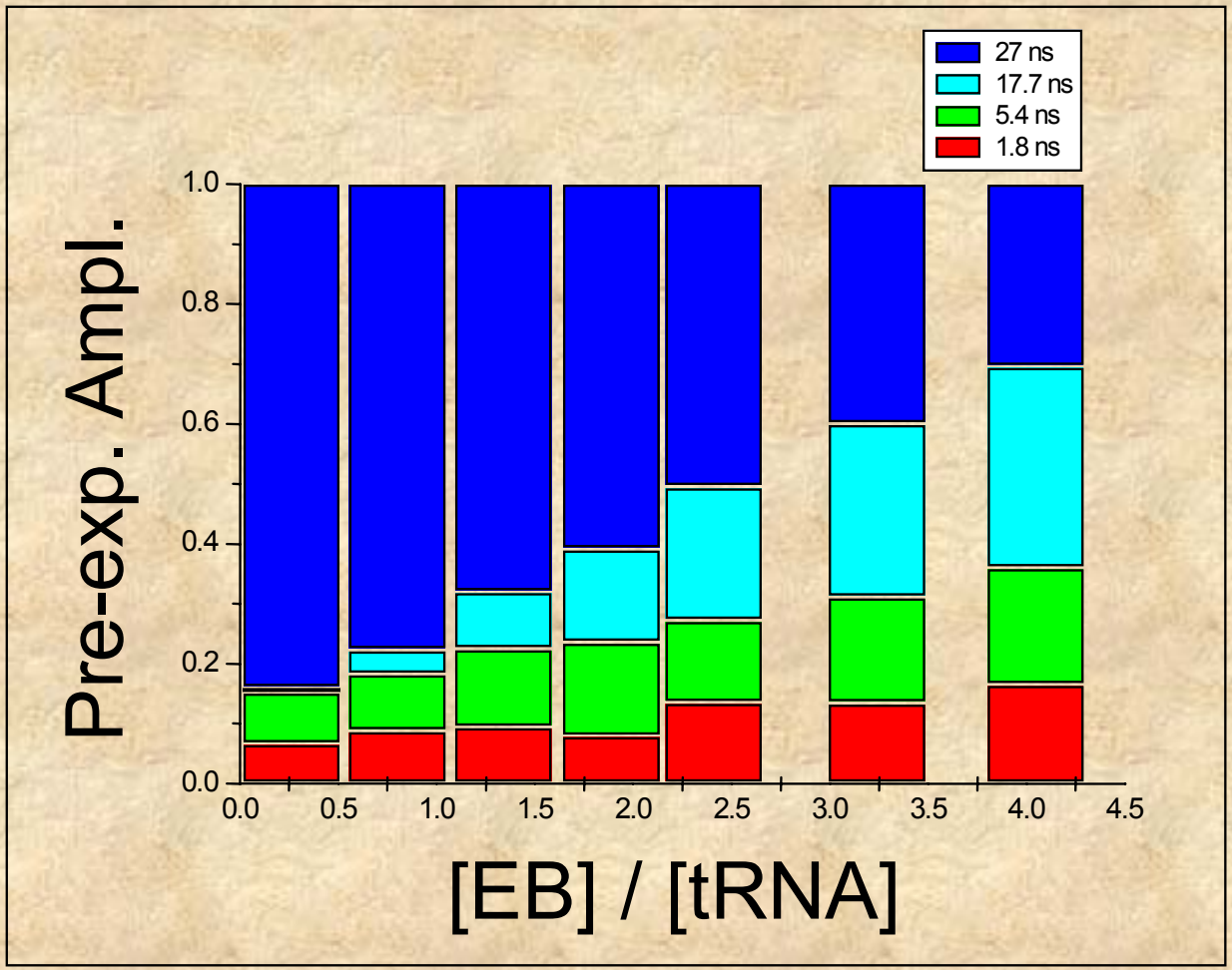
In this system, 8 data sets, with increasing EB/tRNA ratios, were analyzed. Some of the data are shown below for EB/tRNA ratios of 0.27 (circles), 1.34 (squares), 2.41 (triangles) and 4.05 (inverted triangles).



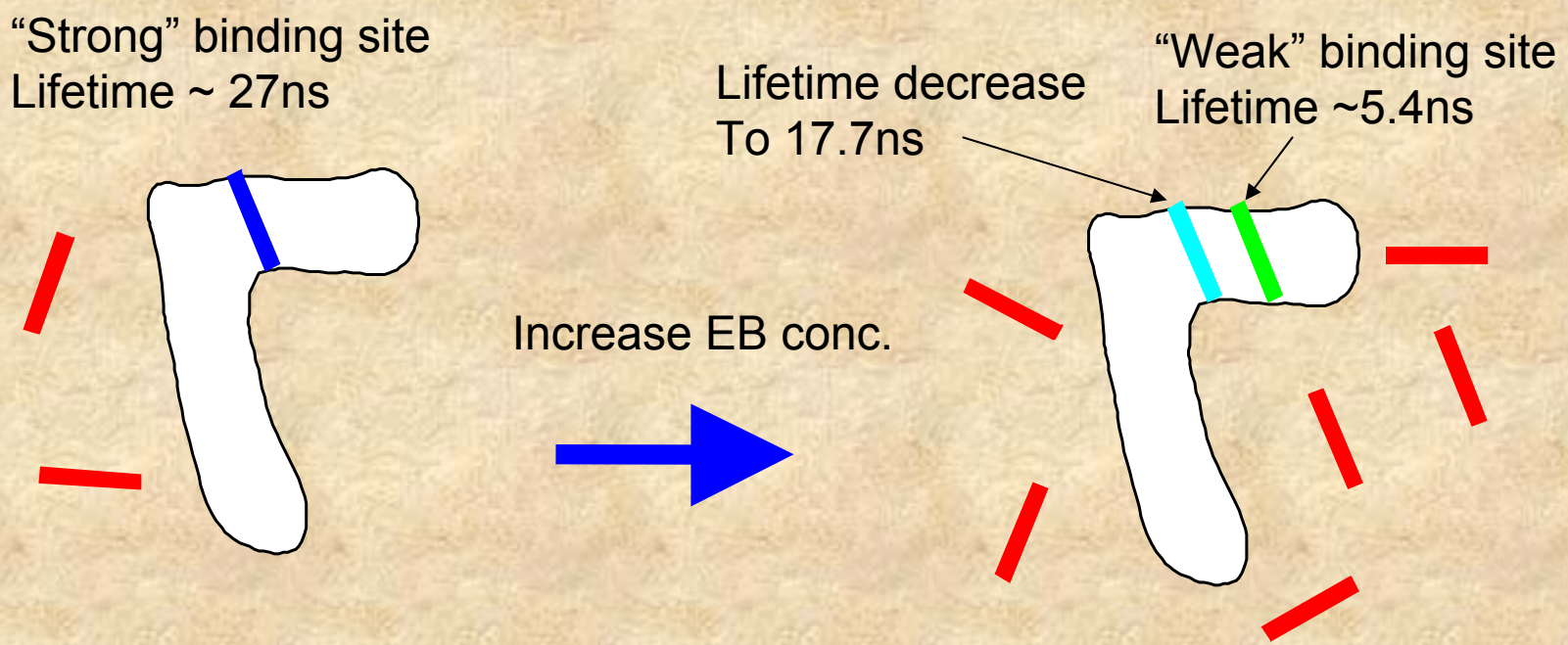
Global Analysis on seven data sets fit best to the 4 component model with two fixed components of 27ns and 1.84ns and two other components of 17.7ns and 5.4ns.



As shown in the plot below, as the EB/tRNA ratio increases the fractional contribution of the 27ns component decreases while the fractional contributions of the 17.7ns and 5.4ns components increase.



The Model



Question:
Is the drop in the lifetime of the "strong" binding site due to a change in tRNA conformation or energy transfer???

Globals demo