Assignment 5: Using Fluorescence Anisotropy to Measure Protein-Protein (CheY-CheZ) interactions

1.2 The single-set-of-sites model, which is corresponds to the hyperbolic function in Origin (P1**x*/(P2 + *x*)) is inadequate to describe the results, because the anisotropy of labeled CheZ is not zero when the CheY concentration (*x*) is zero.

Inadequate: $An(Y) = A1*[Y]/(K_D + [Y])$ (1)

1.4. The creation of a function that has a finite (nonzero) anisotropy at zero concentration (A0) and a *change* in anisotropy (A1) that is proportional to the extent of ligand binding shows improved behavior, but still suffers from systematic deviations between the data and the fit.

Improved but not sufficient: $An(Y) = A0 + A1*K_A*[Y]/(1 + K_A*[Y])$ (2)

1.5. The assumption that the free concentration of CheY, [Y], is well approximated by the total CheY concentration, $[Y]_{tot}$, is not justified. Therefore, [Y] ($[Y]_{free}$) must be expressed as a function of $[Y]_{tot}$, $[Z]_{tot}$ and K_A. This equation is based on the single-set-of-sites model, and is derived using the expression for the association constant of Y:Z complex formation (Eq. 3) and the mass conservation equations for Z (Eq. 4) and Y (Eq. 5).

$$[YZ] = K_A[Y][Z] \quad (3)$$

$$[Z]_{tot} = [Z] + [YZ] = [Z] + K_A[Y][Z] \qquad [Z] = [Z]_{tot}/(1 + K_A[Y]) \quad (4)$$

$$[Y]_{tot} = [Y] + [YZ] = [Y] + K_A[Y][Z] \qquad [Y]_{tot} = [Y] + K_A[Y][Z]_{tot}/(1 + K_A[Y]) \quad (5)$$

Equation 5, rearranged as a quadratic in [Y] (Eq. 6), yields two roots (Eq. 7).

$$[Y]_{tot} + K_A[Y]_{tot}[Y] = [Y] + K_A[Y]^2 + K_A[Z]_{tot}[Y] \quad (6a)$$

$$K_A[Y]^2 + [Y] + K_A[Z]_{tot}[Y] - K_A[Y]_{tot}[Y] - [Y]_{tot} = 0 \quad (6b)$$

$$[Y]^2 + (1/K_A + [Z]_{tot} - [Y]_{tot})[Y] - [Y]_{tot}/K_A = 0 \quad (6c)$$

$$[Y] = -(1/K_A + [Z]_{tot} - [Y]_{tot})/2 \pm \mathrm{sqrt}((1/K_A + [Z]_{tot} - [Y]_{tot})^2 + 4[Y]_{tot}/K_A)/2 \quad (7)$$

That the "+" root corresponds to [Y], which follows from consideration of the values obtained for the "+" and "-" forms of Equation 7: when $[Y]_{tot} = 0$, and when $[Z]_{tot}$ (0.2 μM) and K_A are positive and finite. With this constraint [Y] must equal 0; only the "+" form yields a satisfactory result.

One is led to write an Origin script in the following form:

$$f_B = [YZ]/([Z] + [YZ]) = K_A[Y]/(1 + K_A[Y]) \quad (8)$$

where Eq. 7 is used in the formula for f_B, which is part of the equation for the anisotropy

$$An(Y) = A0 + A1*f_B([Y]_{tot}, [Z]_{tot}, K_A) \quad (9)$$

This model-dependent expression is used to retrieve estimates of K_A from the anisotropy data collected as a function of the total Y concentration.

Origin script

Independent variable
x = $[Y]_{tot}$
Dependent variable
y = anisotropies
Parameters
a0 = anisotropy at $[Y]_{tot}$ = 0; a1 = change in anisotropy; Ka = association constant;
Mt = total CheZ concentration (fixed at 0.2 μM)
Temporary variables
bq = first order coefficient in quadratic equation
yfree = estimate of free Y concentration
fb = fraction of CheZ in YZ complex
y = a0 + a1*fb;

Y-Script
```
bq = Mt - x + 1/Ka;
yfree = (-bq + sqrt(bq^2 + 4*x/Ka))/(2);
fb = Ka*yfree/(1 + Ka*yfree);
y = a0 + a1*fb;
```

Table 1 – Results of Individual & 'Global' Fits to the Single Site Model*

Labeled	Global			Individual		
Residue	**A0**	**A1**	**K_A (μM^{-1})**	**A0**	**A1**	**K_A (μM^{-1})**
16	0.118 ± 0.001	0.012 ± 0.001	42 ± 8	0.118 ± 0.001	0.013 ± 0.001	19 ± 11
180	0.075 ± 0.001	0.053 ± 0.001		0.075 ± 0.001	0.051 ± 0.001	69 ± 13
194	0.097 ± 0.001	0.024 ± 0.001		0.097 ± 0.001	0.024 ± 0.001	30 ± 10
210	0.100 ± 0.001	0.024 ± 0.001		0.101 ± 0.001	0.028 ± 0.001	12 ± 3

*Mt was fixed at 0.2 μM throughout. 'Global' means that K_A was minimized as a shared parameter that was common to all the datasets. χ^2 for individual and global fits were 7.2 x 10^{-7} and 1.5 x 10^{-6}, respectively; not a large difference.

1.6.

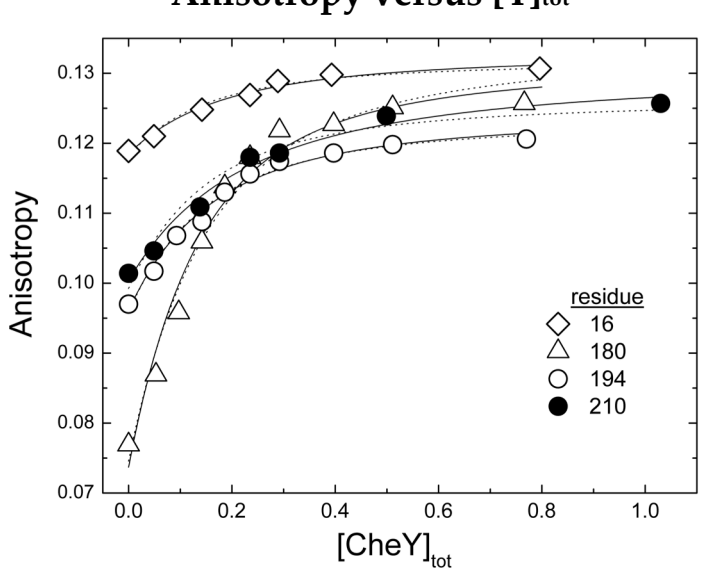

Figure 1 - Anisotropy of fluorescein-labeled CheZ as a function of the CheY concentration (μM). Curves represent individual (solid) and global (dotted) fits of the data to a single site model.

Comments: The initial anisotropies, A0, in the absence of Y show larger variations than the anisotropy in the presence of saturating amounts of CheY, A0 + A1, where the values tend to converge. This implies that the dynamics in the absence of Y reflect local conditions in the protein, *e.g.* in the tether *versus* not in the tether. Y binding, which reduces dynamics to a similar value, result in anisotropies that may then reflect 'end over end' rigid-body motions of the CheZ protein. With respect to global *versus* individual fits, it is observed that:

(*i*) Sharing Mt among all the data sets is of no consequence, because this parameter is fixed at 0.2 μM throughout.

(*ii*) The initial anisotropies, and the extent to which the anisotropies change, are different for each labeled form of CheZ, therefore it is *not* appropriate to share these parameters.

(*iii*) K_A, on the other hand, is expected to be the same in all these cases, so it is reasonable to expect that this value can be constrained to be the same in all the fits. Table 1 permits the comparison of the global estimate for K_A to the estimates obtained for each data set individually, with the following observations:

- *(a)* The average K_A calculated from the individual estimates (32.6 μM^{-1}) is similar to the global value (42.5 μM^{-1}).
- *(b)* The relative error in K_A generated with the global fit, ~20%, is comparable to the smallest relative error generated by the individual fits, which range from ~20 to 60%. The precision in the global fit comes without a significant increase in χ^2. That the relative errors and χ^2 are not significantly *worse* in the global fit is construed to mean that the sharing of the K_A parameter is justified.

Is the global fit essential for this data analysis? *No*. Each individual set stands on it's own, although the binding of CheY to each of the CheZs could have been determined more than once to establish the repeatability of each measurement.

When is global fitting necessary/justified? Global fitting is justified when aggregate data sets extend the measurement range, and the fitting equation is used to find estimates of the same parameter(s). Two examples:

(i) For example, if the binding experiments were conducted at several different values of $[Z]_{tot}$ for each of the labeled forms of CheZ, then these data would cover a greater range of conditions that together could be used to find one estimate of K_A. *When there are multiple overlapping sets of data that share one or more variables.*

(ii) In an equilibrium sedimentation experiment, the sedimentation data are often collected over a greater range of parameter space (C(*r*) vs. *r*), by using two or three values of the initial (uniform) protein concentration (C_{intial}) and/or two or three values of the angular velocity (ω^2). With this larger, overlapping set of data, more accurate estimates of MW (e.g. in the case of a single species), or MW and K_A (e.g. in the case of a dimerizing species) can be obtained.

In these situations, where multiple overlapping data sets extend the range of the data, the global fit provides a convenient way to appropriately weight the relative contributions of the different data sets in the overall fit.